What is the economic impact of hazardous prescribing of NSAIDs with anticoagulants and does a pharmacist-led information technology intervention reduce this impact?

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2021

Leonie S Penner

School of Health Sciences

Manchester Centre for Health Economics

Contents

Words (excluding references): 88814

Tables: 35

Figures: 16

Appendices: 14

References: 630

List of tables

List of figures

List of appendices

List of abbreviations

ACS	Acute Coronary Syndrome
ADE	Adverse Drug Event
ADR	Adverse Drug Reaction
AdViSHE	Assessment of the Validation Status of Health-Economic decision models
AF	Atrial Fibrillation
ARISTOTLE	Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation
ASPIRE	Action to Support Practices Implementing Research Evidence
BMI	Body Mass Index
CBA	Cost-Benefit Analysis
CCG	Clinical Commissioning Group
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CHEERS	Consolidated Health Economic Evaluation Report Standard
CPRD	Clinical Practice Research Datalink (Gold)
CUA	Cost-Utility analysis
DAG	Direct Acyclic Graph
DALY	Disability-Adjusted Life-Years
DCE	Discrete Choice Experiment
DHI	Digital Health Intervention
DID	Difference In Difference
DOAC	Direct Oral Anticoagulant
DQIP	Data-driven Quality Improvement in Primary Care
EDQM	European Directorate for the Quality of Medicines & Healthcare
EEPRU	Policy Research Unit in Economic Evaluation of Health and Care Interventions
EQ-5D-3L	EuroQol 5-Dimension 3-Level
EQ-5D-5L	EuroQol 5-Dimension 5-Level
GI	Gastro-Intestinal
GP	General Practitioner (or family practitioner)
HES	Hospital Episode Statistics
HPE	Hazardous Prescribing Event
HR	Hazard Ratio
HTA	Health Technology Appraisal
ICER	Incremental Cost-Effectiveness Ratio
IMD	Index of Multiple Deprivation
INR	International Normalized Ratio
IPTW	Inverse Probability of Treatment Weighting
ISAC	Independent Scientific Advisory Committee

ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITSA	Interrupted Time Series Analysis
NCC MERP	National Coordinating Council for Medication Error Reporting and Prevention
NHS	The National Health Service
NHS/PSS	NHS and Personal Social Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NIPPS	Neighbourhood Integrated Practice Pharmacists in Salford
NMB	Net Monetary Benefit
NSAID	Non-Steroidal Anti-Inflammatory Drug
OAC	Oral Anticoagulants
ONS	Office for National Statistics
OR	Odds Ratio
OSA	One-way Sensitivity Analysis
OTC	Over The Counter
PDRM	Preventable Drug Related Morbidity
PIM	Potentially Inappropriate Prescribing
PINCER	Pharmacist-led Information Technology Intervention for Medication Errors
PROTECT	Programme grant: Avoiding patient harm through the application of prescribing safety indicators in English general practices
PSA	Probabilistic Sensitivity Analysis
PSM	Propensity Score Matching
QALY	Quality-Adjusted Life-Year
RCT	Randomised Controlled Trial
RE-LY	Randomized Evaluation of Long-Term Anticoagulation Therapy
SMASH	Safety Medication dASHboard
STOPP	Screening Tool of Older Persons Potentially Inappropriate Prescriptions
THIN	The Health Improvement Network
VKA	Vitamin-K Antagonist
VTE	Venous Thromboembolism
WTP	Willingness To Pay

Glossary of key terms

Adverse drug event	Harmful and unintended consequences of medication use
Cost-benefit analysis	A comparison of the cost and benefits of a set of alternative strategies with the benefits measured in monetary units
Cost-effectiveness analysis	A comparison of the cost and benefits of a set of alternative strategies with the benefits measured in non-monetary units
Decision-analytic model	A mathematical approach to estimating the expected costs and benefits of alternative strategies
Decision tree	A type of decision-analytic model which outlines the probabilities of a series of events occurring and the costs and outcomes of those events
Electronic audit and feedback interventions	Interventions that use computer interfaces to provide clinical performance summaries to healthcare professionals that target behaviour change as part of clinical practice improvement
Economic evaluation	A framework for the comparative analysis of alternative courses of action in terms of both their costs and consequences
EQ-5D-3L	The five-question generic measure of health related quality of life which measures health in five dimensions (mobility, self-care, usual activities, pain & discomfort, anxiety & depression) with each domain having three levels (typically: no problems, some problems, extreme problems).
Hazardous prescribing event (HPE)	A clinically meaningful hazardous prescribing event occurs when, as a result of a prescribing decision or prescription writing process, there is an unintentional significant (1) reduction in the probability of treatment being timely and effective or (2) increase in the risk of harm when compared with generally accepted practice. Synonym to prescribing errors.
Health related quality of life	The utility value associated with a particular health state or the way that health state affects a person's quality of life.
Incremental cost-effectiveness ratio	The ratio of the incremental costs and benefits produced by the potential introduction of a healthcare intervention
Interrupted time series analysis (ITSA)	A quasi-experimental method of statistical analysis that uses data before and after initiation of an intervention from a single unit. The intervention is expected to 'interrupt' the outcome measure after initiation.
Net monetary benefit	A measure of cost-effectiveness that represents the value of an intervention in monetary terms when a willingness to pay threshold for a unit of benefit (for example a measure of health outcome or QALY) is known
Non-steroidal anti-inflammatory drugs	Drug group used to reduce pain, decrease fever and inflammation in diseases, such as arthritis, headache, or chronic back pain. Short form: NSAID
Routinely collected health data	Data collected for purposes other than research, such as electronic health records, administrative data or disease registries.

Oral anticoagulants	Drug group used to treat and prevent embolic events, such as stroke that includes vitamin-K antagonists, such as warfarin, and the direct oral anticoagulants (DOAC), such as rivaroxaban and apixaban.
Probabilistic sensitivity analysis	An approach for accounting for uncertainty in economic evaluations where each parameter is assigned a probability distribution and the model is repeatedly run over many iterations with new parameter values drawn for each iteration
Quasi-experimental study designs	An empirical study design that aims to estimate the causal impact of interventions without random assignment.
State-transition model	A type of decision-analytic model which conceptualises a health care decision problem into a set of mutually exclusive and exhaustive health states, such as Markov cohort models.
Quality-adjusted life-year	A generic outcome measure that combines health related quality of life with length of life.
Utility or health state utility value	The measure of value attached to a specific health state in a decision-analytic model.

Abstract

Introduction: Digital health interventions, such as electronic audit and feedback (e-A&F), are increasingly used to improve the quality and safety of healthcare. Where there is limited opportunity to implement randomised controlled trials, evidence of effectiveness has to be based on quasi-experimental methods. Estimating the health and economic impact of these interventions often relies on the use of process indicators rather than patient outcomes. This thesis assesses the economic impact of the Safety MedicAtion daSHboard (SMASH), an e-A&F intervention that combines digital case finding with pharmacist-led medication review. In particular, it explores the feasibility of conducting an economic analysis as part of a quasi-experimental study that relies on electronic health records to measure exposure and outcomes. Effectiveness was measured as the reduction in the number of hazardous prescribing events (HPEs). One specific HPE targeted by SMASH, prescription of an NSAID to patients with oral anticoagulation, was used as a case study to demonstrate how this can be done. The aim of this thesis was to assess the cost-effectiveness of SMASH in reducing this HPE.

Methods: This thesis comprised a series of studies that: (i) estimated the cost of the complex SMASH intervention using a micro-costing approach; (ii) quantified the increased likelihood of adverse drug events (ADEs) associated with NSAID use in anticoagulated patients in an observational cohort study using routinely collected linked primary and secondary care data (Clinical Practice Research Datalink, CPRD GOLD; Hospital Episodes Statistics, HES; and Office for National Statistics, ONS); (iii) estimated the incremental costs and quality-adjusted life-years (QALYs) associated with the HPE from the perspective of the NHS and Personal Social Services by conceptualising a probabilistic state-transition model of potential treatment pathways related to ADEs associated with the HPE (cost year: 2019, discount rate: 3.5%, cycle length: 3 months, time horizon: lifetime); (iv) estimated the relative cost-effectiveness of SMASH compared with standard care for the HPE under investigation based on a probabilistic two-stage decision analytic model that combined the results of the earlier studies.

Results: The expected cost of SMASH at 12 months was £2149(2.5% to 97.5% credible interval £487 to £5790), and £205 (2.5% to 97.5% credible interval £46 to £559) per HPE avoided. The observational study projected an increased risk of serious GI events (HR 2.96, 95% CI 1.60 to 5.46) and stroke (HR 2.48, 95% CI 1.36 to 4.53) in the presence of the HPE in patients at risk. An anticoagulated patient with a concomitant NSAID prescription was estimated to have £244 (2.5% to 97.5% credible interval -£149 to £1073) higher costs and 0.04 (2.5% to 97.5% credible interval -0.17 to 0.05) fewer QALYs than a patient without an NSAID based on the results of the state-transition model. The decision-analytic model projected the costs and QALYs related to NSAID use in anticoagulated patients from the process indicator of HPE reductions. At the threshold used to determine cost-effectiveness by the National Institute for Health and Care Excellence (NICE) (£20000 per QALY gained), the incremental net monetary benefit of SMASH in reducing the HPE was estimated to be -£311 (2.5% to 97.5% credible interval -£542 to £73).

Conclusion: This thesis found that it was feasible to conduct an economic analysis as part of a quasi-experimental study that relies on electronic health records to measure exposure and outcomes. NSAID use in anticoagulated patients increased the risk of ADEs and was associated with a higher cost and lower quality of life. SMASH was not cost-effective in reducing this HPE type at the defined WTP threshold. The study provides a 'real world' estimate of the cost-effectiveness of SMASH based on projected harm from routinely collected healthcare data and the methodology is generalisable to other HPEs targeted by SMASH, where no data on patient outcomes associated with the HPE are available. Future research is needed to estimate the cost-effectiveness of SMASH in reducing other HPEs.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning

Copyright

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the 'Copyright') and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the 'Intellectual Property') and any reproductions of copyright works in the thesis, for example graphs and tables ('Reproductions'), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

In undertaking this PhD, I have received support, guidance and encouragement from a number of different people.

Firstly, I would like to sincerely thank my supervisors Professor Rachel Elliott, Professor Darren Ashcroft, Professor Niels Peek and Dr Sean Gavan. Their supervision, guidance and erudition has expertly guided me through the research and the writing of this thesis. I am especially grateful to Professor Rachel Elliott for being both an academic role model and dedicated mentor on my health economics journey. From our first encounter at the University of Nottingham, she encouraged me to reach beyond my training as a German pharmacist, and immerse myself in the world of economics, statistics and epidemiology. Without her, this PhD in health economics would not have been undertaken, and Rachel's continuous support and her detailed feedback to this day pushes me to improve myself and to go the extra mile. A special thanks also to Dr Sean Gavan who was always there to discuss my work or answer my questions even on short notice.

Thanks to my dear colleagues from the Manchester Centre for Health Economics and my fellow PhD girls from the Centre of Health Informatics for making this long journey an enjoyable one. A special thanks to Dr Georgios Gkountouras for his support. He was my saviour throughout the PhD and helped me find solutions to the most mysterious problems with my statistical analysis or the economic modelling.

I also want to thank Anna Wilding, Helen Hayes and Dr Luke Munford. Your academic and personal friendship was a pivotal element of this PhD journey. It always made me look forward to coming back to Manchester and enabled a great insight into British culture as well as its interesting humour.

I would also like to express my gratitude to my family, to those that are still among us and those that cannot be here anymore. Your support, kindness and love are what grounded me through out these restless times commuting between countries and airports. I am especially grateful to my mother who not only fostered my academic curiosity from a young age but also made sure that there is always a place to come home to.

Finally, I want to thank my 'adorable' husband David. He encouraged me to go abroad and to pursue my career even though this meant living apart for another three years. His invigorating comments on my thesis gave me the strength and patience I needed to finalise this work. With you, I feel there is nothing we cannot achieve.

Chapter 1 - Introduction

Healthcare systems around the world have to make decisions on how to allocate their resources every day. New treatments and procedures, diagnostic technologies, interventions or programmes are developed and decisions on which can be reimbursed by the healthcare system are required. Decisions to reimburse a new technology will prevent the resources required for the implementation of this new technology to be used for something else. Many countries, such as Australia (1), Canada (2) and the United Kingdom (UK) (3), use economic evaluations to aid their decision making. Economic evaluations compare alternative options in terms of their healthcare costs and consequences (4). Evidence from economic evaluations allows decision-makers to understand whether the health gained from a new intervention is greater than the health forgone (opportunity cost) per additional measure of effect gained. The National Institute for Health and Care Excellence (NICE) provides guidance on health technologies to inform decision making by the National Health Service (NHS) in England. Cost-effectiveness analysis (CEA) incorporates the basic principles of an economic analysis as required by NICE in a healthcare setting due to a focus on outcomes, such as health improvement or reduction of risk.

Extensive methodology exists for undertaking economic evaluations as part of randomised studies that are conducted under tightly controlled conditions and incorporate prospective data collection (5, 6). In many cases, conducting randomised studies is not ethical or feasible (or neither) (7-9), and then evaluations have to rely on evidence from pragmatic 'real world' studies. This is a relatively little explored terrain for health economics and no comprehensive guidance exists on how economic evaluations can be conducted alongside quasi-experimental studies (10, 11). Only recently, NICE acknowledged quasi-experimental studies as a minimum evidence standard for digital health technologies (12). However, there is an abundance of quasi-experimental designs and statistical methods for analysis (13, 14) and it is not clear from the proposed evidence framework how this evidence should be used. Each of these methods relies on different assumptions (7, 10, 13) that could have implications for economic evaluations. This thesis explores how a health economic analysis

can be done in the context of a pragmatic study that uses a quasi-experimental design and relies on routinely collected health data to measure exposure and outcomes. The intervention chosen for this dissertation is in the setting of medication safety in primary care. The European Union (EU) Council and Commission have highlighted the need for cost-effectiveness evidence specifically for interventions in the area of patient safety (15, 16).

Medicines are widely used to treat, manage or prevent diseases. In 2019, the overall number of prescribed medicines in England reached 1.12 billion prescription items (17). The number of prescription items is increasing. Compared with 2018 the number of prescription items increased by 2.8% in 2019 (17). Most of the prescribed medicines do not cause harm, but there is a residual risk of adverse drug events (ADEs) that can lead to hospital admissions (18-21). ADEs can be preventable if they are a result of a medication error (22). The World Health Organization (WHO) (23) refers to a statement of the United States National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP), to define medication errors: 'A medication error is any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the healthcare professional, patient, or consumer. Such events may be related to professional practice, healthcare products, procedures, and systems, including prescribing, order communication, product labelling, packaging, and nomenclature, compounding, dispensing, distribution, administration, education, monitoring, and use' (24).

Medication errors are responsible for 59% of drug related hospital admissions (25) and 3.7% of all hospital admissions worldwide (26). Awareness of the burden of medication errors was highlighted in 1999 with the landmark report 'To Err is Human' by the United States (US) Institute of Medicine (27). The report stimulated discourse and research in patient safety and led to an increasing number of publications and research awards in this area (28). The UK responded with the publication of 'An Organisation with a Memory', which looked at how the NHS should change to learn from failures with regards to preventable ADEs (29). In the EU, the additional cost to healthcare providers of preventable ADEs was estimated to be between 17-38 billion Euros in 2015 (30). For NHS England the direct healthcare cost related to medication errors was estimated to be about £89.1 million

per year in 2016 (31). In addition to this financial burden, preventable ADEs can introduce physical and psychological harm (32, 33). For the EU harm from preventable ADEs was estimated to cause a total loss of 1.5 million disability-adjusted life-years (DALYs) (30). In accordance with the WHO, the sum of DALYs across a population are used as a measure of the disease burden representing the gap between the actual health status and an ideal health situation (34).

The WHO's third global patient safety challenge (2017) calls for commitment of healthcare providers and stakeholders to reduce preventable ADEs by 50% over five years (33). The United Kingdom (UK) government recently made the reduction of medication errors a policy objective (35) following recommendations from the WHO (33) and the European Directorate for the Quality of Medicines & Healthcare (EDQM) (36). The policy encourages the implementation of interventions that target the reduction of medication errors. According to recent findings from a literature review, medication errors that were most serious or likely to cause harm were in primary care (31) where prescribing and monitoring errors account for more than half of preventable admissions to hospital (37).

So far, the majority of patient safety research has been conducted in secondary care settings and most interventions have been developed to reduce medication errors in hospitals (38, 39). This thesis focuses on an electronic audit and feedback (e-A&F) intervention aiming to reduce potentially hazardous prescribing in primary care, the Safety Medication dASHboard (SMASH) (40). Hazardous prescribing, in contrast to prescribing errors, describes medication combinations that are not generally accepted practice and likely to cause harm, such as the prescription of a non-steroidal anti-inflammatory drug (NSAID) to a patient with an oral anticoagulant (OAC) (41). The terms 'hazardous' or 'potentially inappropriate' can be used to imply the increased risk of ADEs, such as bleeding, without stating that the action was preventable and therefore an error. In the example of NSAID use that can contribute to the bleeding risk associated with the OAC, the patient might need the NSAID because no alternative treatment is available for the patient to manage pain to an appropriate level. Dependent on the patient's symptoms, different hazards are tolerable and the increased risk of not receiving the 'erroneous`

prescription might outweigh the risk of ADEs. That's why in this thesis the term 'hazardous prescribing event' (HPE) is used preferably compared with 'medication error'.

The WHO proposes three different intervention formats suitable to reduce error rates: medication reviews, automated information systems and education interventions (23). SMASH combines aspects of pharmacist-led medication reviews and reconciliation with an automated digital component searching electronic health records to identify patients potentially at risk due to hazardous prescribing events (40). This type of multifaceted intervention has been found to be the most promising approach to reduce hazardous prescribing (23).

SMASH focuses on a specific set of hazardous medication use events, referred to as HPEs and hazardous monitoring events (HMEs). In order to identify the HPEs and HMEs targeted by SMASH, prescribing safety indicators are used. These indicators provide specific definitions for each type of hazardous prescribing or monitoring that are associated with an increased risk of harm for the patient (42, 43). As described above, NSAID use in anticoagulated patients is one type of HPE targeted by SMASH. SMASH is a local development of a pharmacist-led information technology intervention (PINCER) that was tested in a randomised controlled trial (RCT) compared with simple feedback (44, 45). Simple feedback included a report of patients with HPEs but did not include a pharmacist service to review the report. PINCER and SMASH aim to reduce a similar set of HPE types. The PINCER RCT in Nottinghamshire, Staffordshire and Central and Eastern Cheshire showed a significant reduction of the occurrences of HPEs (Odds ratio (OR) 0.71, 95% confidence interval (CI) 0.59 to 0.86) and HMEs (OR 0.56, 95% CI 0.44 to 0.70) at six months (44). After 12 months this effect waned. While the intervention was still cost-effective, no further reduction in HPE rates was evident. Compared with the PINCER intervention, SMASH provides continuous surveillance by pharmacists, which was hoped to target the challenge of diminishing intervention effects. Providing feedback at multiple instances was found to be a key design feature required for effective e-A&F interventions, such as SMASH and PINCER (46, 47).

A study in Salford, Greater Manchester, assessed the effectiveness of SMASH in reducing HPEs (48) using a quasi-experimental study design. HPE rates were reduced by 27.9% (95% CI 20.3% to 36.8%) at six months and the effect persisted with a reduction of 40.7% (95% CI 29.1% to 54.2%) at 12 months. With funding from the Health Foundation and East Midlands Academic Health Science Network (AHSN) a large-scale rollout of PINCER to 370 practices in the East Midlands between 2015 and 2017 was evaluated and demonstrated a significant reduction of HPE rates. As a result of these positive findings, PINCER was included in the national AHSN innovative network programme and is rolled-out nationally momentarily. In Greater Manchester, SMASH is implemented as part of this national rollout.

Promising results of interventions aiming to reduce HPE rates were also found in Scotland, where a similar set of prescribing and monitoring types was developed and their prevalence in 315 Scottish general practices was examined (49, 50). The set of HPEs was used in the Data-driven Quality Improvement in Primary Care (DQIP) intervention, and its effectiveness in reducing HPE occurrences tested in a stepped wedge RCT, which found a reduction of patients at high risk (OR 0.63, 95% CI 0.57 to 0.68) compared with standard care (51, 52).

While evidence on the effectiveness of these interventions in reducing HPE rates is available, the impact on patient outcomes and potential cost-savings is sparse. The PINCER RCT assessed the cost-effectiveness of the original PINCER study by projecting patient outcomes and costs of the different HPE types. However, the set of HPEs included in SMASH and PINCER has changed since then and the evidence is comparably old (the RCT was run between 2006 and 2008 (45)). A 4.5 year programme grant, 'Avoiding patient harm through the application of prescribing safety indicators in English general practices' (PROTECT), with over £2.4 million awarded by the NIHR Programme Grant for Applied Research is looking at building better estimates on harm and economic impact of the HPE types targeted by PINCER (53). A qualitative work stream of PROTECT generated recommendations for the recently started national rollout of PINCER. The quantitative work stream involves estimating the economic impact of each of the HPE types targeted by PINCER. This dissertation is imbedded in the PROTECT programme grant. While this programme of work investigates the cost-effectiveness of SMASH, not of PINCER, the work on the economic

impact of each HPE type can be used interchangeably between the two interventions because the set of HPE types targeted is overlapping to a large extent.

In this thesis, the cost of delivering SMASH was estimated and combined with data derived from the quasi-experimental study of the effectiveness of SMASH (48). To analyse the impact that SMASH has not only on the process indicator of HPE incidence but also on patient outcomes, the occurrence of ADEs associated with the HPE types needs to be quantified. According to a recent report by the Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) driving the new UK policy implications, evidence on quantifiable patient harm related to HPEs is limited (31). There are very few studies linking HPEs with patient harm directly. Only reviews of ADEs or hospitalisations exist which assessed the preventability of these events retrospectively (20, 37, 54). A recent initiative from the NHS developed a dashboard linking a set of the mentioned prescribing safety indicators, derived from those used in SMASH and PINCER, with hospitalisation in the UK based on secondary care health records (55). The dashboard, however, does not provide data on the difference between ADE incidence in the presence and absence of the HPEs. Because there are little precise data on patient harm and costs associated with HPEs, and because they carry a large potential burden, a key objective for this thesis and the PROTECT programme grant was to develop methods to estimate the health related and the economic impact of HPEs and how to incorporate this in the economic evaluation of SMASH and PINCER respectively. In order to project harm and costs from an HPE, the use of routinely collected health data was explored. Estimating harm for each HPE was beyond the scope of this dissertation. Therefore, this dissertation focuses on the cost-effectiveness of SMASH in reducing one type of HPE targeted by the intervention. As part of PROTECT the other HPE types are being investigated. The HPE type investigated in this thesis was used to illustrate the methods of how to project the long-term consequences of reducing the HPE incidence by using estimates from routinely collected health data.

The HPE type chosen to demonstrate this was the HPE of prescribing an NSAID to a patient with concomitant OAC treatment that is associated with an increased gastro-intestinal (GI) bleeding risk. OAC treatments include treatment with vitamin-K antagonists (VKAs) and direct oral anticoagulants (DOACs). Patients receiving OAC treatment are referred to as

anticoagulated patients. This HPE type included two drugs in the prescribing safety indicator definition, and it was neither the least nor the most frequent HPE. As a result, the HPE type was considered to be representative of the other types of HPEs targeted by SMASH.

1.1 Aims and objectives of the dissertation

The aim of this dissertation was to assess the cost-effectiveness of SMASH in reducing the number of patients receiving oral anticoagulation with a concomitant hazardous prescription of an NSAID. This HPE type was used as an example to demonstrate how observational data from quasi-experimental studies and routinely collected health data can be incorporated into an economic evaluation of a system-level digital health intervention (DHI). The methods used in this dissertation inform an approach that can be generalised to estimate the cost-effectiveness of interventions to reduce different type of HPEs.

The aim is achieved by meeting the following specific objectives:

1. identify the costs associated with the provision and implementation of the SMASH intervention;
2. incorporate the cost of the intervention with the effectiveness data derived from the quasi-experimental study into a decision-analytic model to assess cost per HPE avoided;
3. quantify patient harm associated with NSAID use in anticoagulated patients by measuring the risk difference in related ADEs of patients exposed to the HPE and those at risk of the HPE using routinely collected health data;
4. model potential treatment pathways related to the consequences of NSAID use in anticoagulated patients informed by findings on ADEs related to the HPE;
5. generate input parameters for estimated harm to populate the model on HPE consequences to estimate the economic impact of NSAID use in anticoagulated patients;
6. assess the cost-effectiveness of the SMASH intervention in reducing this specific HPE by combining results on costs per HPE avoided by the intervention with the modelled patient harm and costs associated with the HPE occurrence;
7. reflect on conducting an economic evaluation as part of a quasi-experimental study that relied on routinely collected health data to measure exposure and patient outcomes.

1.2 Organisation of the dissertation

This dissertation begins with an overview of HPEs in general and their prevalence in the UK in particular in Chapter Two followed by a summary of evidence on the link between HPEs and patient harm or ADEs. Chapter Two also describes the literature on interventions aiming to reduce HPEs in primary care and how their cost-effectiveness is assessed. Finally, SMASH is described in detail.

Each of the subsequent four chapters describes a separate study, and methods and results are reported within each chapter. As was mentioned above, the HPE under investigation is the hazardous prescription of an NSAID to a patient with concomitant oral anticoagulation.

Chapter Three reports an economic evaluation of SMASH to estimate the incremental cost per HPE avoided compared with standard care. This chapter demonstrates how micro-costing was used to estimate the cost of delivering SMASH and how the results of the quasi-experimental effectiveness study were incorporated into the decision-analytic model.

Chapter Four describes a cohort study using a linked dataset from primary care and secondary care health records to estimate the relative risk of specific ADEs associated with the presence of the HPE (NSAID use in anticoagulated patients).

Chapter Five utilises the results of Chapter Four to conceptualise and populate a state-transition model and estimate the long-term impact of the presence of the HPE on cost to the healthcare system and health related quality of life.

Chapter Six combines the results from the three previous chapters in a cost-utility analysis of SMASH in reducing NSAID use in anticoagulated patients. The findings are presented as the incremental cost per additional quality-adjusted life-year (QALY) generated by SMASH compared with current practice.

Chapter Seven concludes the dissertation by discussing the importance of the findings from the four studies and reflects upon the aim of this programme of work.

Chapter 2 - Background

This chapter aims to create an understanding of the burden of hazardous prescribing by describing how HPEs are measured followed by evidence on the prevalence of HPEs, harm associated with HPEs and the economic impact of HPEs. A brief summary of interventions that aim to reduce hazardous prescribing is provided leading to a systematic literature review of economic evaluations of interventions that were designed to reduce hazardous prescribing. Finally, the intervention that is evaluated in this dissertation (SMASH) is described in detail.

2.1 Identifying hazardous prescribing events

This section describes how HPEs are commonly identified by introducing different types of criteria used to identify HPEs. Criteria can be generic or describe specific types of HPEs. Dean et al. (2000) present a list of what to judge as HPEs, and what is not considered an HPE. Drug-drug interactions, for example, are considered HPEs but illegibility of a prescription is not. In contrast to these generic criteria, other studies used specific criteria defining specific drug combination to be inappropriate or hazardous. Commonly used methods are the identification of prescribing safety indicators that describe a specific type of an HPE (39, 43, 50, 56) or the use of set criteria to assess potentially inappropriate medications (PIMs) (57, 58).

The former, prescribing safety indicators define specific prescribing patterns with specific criteria that are associated with an increased risk of harm for the patient (43, 50, 56). The difference between the list of generic criteria by Dean et al. (2000) and specific prescribing safety indicators are explained by an example. The generic criteria by Dean et al. (2000) describe hazardous prescribing situations in general, for instance: ‘prescribing a drug for a patient for whom, as a result of a co-existing clinical condition, that drug is contraindicated’ (59). The list incorporates generic criteria to guide assessors on how to identify HPEs without specific examples of drugs or combinations. Stocks et al. (2015), on the other hand, define specific examples for such situations in the set of prescribing safety indicators, for

example, a patient with heart failure and a prescription of an NSAID (56). This is a specific example of a contraindication. The explicit definition of the prescribing safety indicators facilitates researchers to assess the prevalence of HPEs by screening electronic medical records. This can reduce confounding of prevalence estimates due to subjective evaluation of potentially hazardous situations. Nevertheless, the specific approach does not include all situations of hazardous prescribing detected with the situation list by Dean et al. (2000) and might underestimate the overall prevalence of HPEs.

The latter set of specific criteria, such as Beers criteria (60, 61) or the Screening Tool of Older Persons potentially inappropriate Prescriptions (STOPP criteria) (62) are used to identify PIMs. The criteria contain a list of medicines judged not to be appropriate in older patients as a result of a higher risk-to-benefit ratio and were developed in the United States (US). Beers criteria were not found to be predictive of ADEs, despite an update in 2015 (63). This has led to suggestions that the Beers criteria detect PIMs that are of lower clinical importance (43, 64). Hamilton et al. (2011) stated that PIMs identified according to Beers criteria do not increase the odds for ADEs significantly, in contrast to, PIMs identified by STOPP criteria (64). With every PIM identified by the STOPP criteria, the odds of an ADE increased by 87% (adjusted OR 1.87, CI 95% 1.51 to 2.26). Of the 235 potentially avoidable ADEs in the study, 159 (67.7%) involved a PIM from the STOPP criteria and only 67 (28.5%) from Beers criteria. In addition, the Irish STOPP criteria, updated in 2015 (65) fit the English prescribing routines better than the Beers criteria. Ble et al. (2015), for instance, found 33.8% of individual PIMs from the Beers criteria not to be licensed in the UK (66). Beers and STOPP criteria were developed to identify hazardous prescriptions in the elderly, not in a general population. Therefore, to cover the general UK population the former prescribing safety indicators are assumed to be more suitable because they represent HPEs that are associated with increased harm for all patients, not only the elderly.

2.2 Methods to measure prevalence of hazardous prescribing

This section focuses on different methods to measure HPE prevalence in primary care and how they might influence prevalence estimates. Four different methods to identify HPEs

are now described. The advantages and disadvantages of these four methods are summarised in Table 2.1.

Table 2.1: Comparison of methods to detect hazardous prescribing events

Detection method	Advantages	Source	Disadvantages	Source
Medical record review	Most appropriate to detect prescribing errors	(67)	Low yield for administration errors	(67)
	Identified more errors than with pharmacist screening	(68)	Expensive; time consuming	(67-70)
			Requires complete documentation	(67, 68, 70, 71)
			Dependent on qualification of reviewer	(67, 69, 70)
			Availability of records	(68)
Incident report	Can detect errors not documented in chart reviews	(71)	Under-reporting; low specificity	(67, 69-73)
	Integrated in daily routine; little additional resources	(68, 71)	Often no denominator, hence prevalence estimates often not comparable	(74)
	Low rate of false positives	(69)		
	Suitable for investigating error cause	(69)		
Pharmacy screening	Integrated in daily routine; little additional resources	(68)	Identifies fewer errors than chart review; under-reporting	(68)
	Less time consuming than chart review	(68)		
Direct observation	Most comprehensive for administration or dispensing errors	(69, 70, 75, 76)	Requires highly qualified observers	(69-71)
	Most specific	(69)	Fewer errors if staff knows they are being watched	(70, 71)
	Suitable for investigating error cause	(70)	More expensive than chart review	(69, 71)

The most common method to measure HPEs is the 'medical record review' (77). Medical records are reviewed retrospectively by researchers to identify HPEs. Medical records can be electronic health records or hospital drug charts. In a prospective method, referred to as 'pharmacy screening', community pharmacists record occasions within their daily routine where they needed to intervene due to uncertainties with the prescription. A special scenario of pharmacy screening is the 'incident report' method, where pharmacists use special report forms to report incidents. Such report forms are used, for instance, by the National Reporting and Learning System (NRLS), in the UK to record patient safety

incidents in all healthcare settings (78). In pharmacy screening, all HPEs identified by the pharmacists are recorded. Incident reports include interventions the pharmacist judged to be of importance to influence patient safety. In 'direct observation', an observer follows staff in their daily routine recording HPEs they recognise. For instance, an external researcher could join a prescriber during appointments with patients to check for HPEs during the prescribing process. Incident reports are a simple retrospective way to detect HPEs within the daily routine, but the sensitivity is extremely low due to under-reporting (67, 69-73). Franklin et al. (2007) investigated the difference between two measurement methods in secondary care: incident report and pharmacist screening (73). Only 4% of the errors identified in the pharmacist screening were reported as incidents (73). On the one hand, error reporting systems, such as the NRLS, can support root-cause analysis, due to comprehensive information required in report forms (69). On the other hand, they are not appropriate to be used to detect the actual prevalence of HPEs (68, 79).

Medical record review from charts is one of the most comprehensive measurement methods and the most suitable to identify HPEs (67, 68). Manual chart review, compared with other retrospective methods, is considerably more costly and time consuming (67-70). In a retrospective chart review pilot study by Barber et al. (2006), it took the investigator one day to analyse four patients on average (68). The ability to interpret the cause of an HPE using the chart review method can be difficult because all relevant information may not be available and it may not be possible to talk to the patient, carer or prescriber about the HPE retrospectively (68). Crucial for comprehensive prevalence estimates is the complete documentation of all relevant information within the charts and the accessibility of this information for the reviewer (67, 68, 70, 71).

A less expensive and less time consuming prospective design which can be integrated in the daily routine is prescription screening by a pharmacist as part of the normal dispensing process (68). Disadvantages for this method occur because of lack of time and limited accessibility of health records, which can lead to under-reporting of HPEs (68). On a busy day, pharmacists detect about 40% fewer HPEs than on less busy days (80). Some studies compared the type of HPEs detected with different measurement methods in the same set of patients (68, 81). In Barber et al. (2006), a total number of 134 HPEs were identified with

'pharmacist screening' (48 out of 134) and 'chart review' (93 out of 134) (68). Of the 48 HPEs detected in the pharmacist's screening, only seven of these HPEs were detected by chart review.

The studies by Barber et al. (2006) and Olsen et al. (2007) agree on the statement that different collection methods identify different HPEs, which results in lower prevalence estimates in studies using only one measurement method (68, 81). Of the studies identified in a systematic review of HPE prevalence in secondary care by Lewis et al. (2009), 27% of studies mix different methods to detect HPEs, which impeded the comparability of results (72).

The ability of the aforementioned methods to detect HPEs are among other things dependent on the qualification and the subjective view of the reviewer. The variation in number of detected errors per patient for different reviewers can be statistically significant as shown in a pilot study by Barber et al. (2006) (68). The reliability between researchers, the inter-rater reliability, varies according to the number of researchers and whether they work in teams or individually (82, 83). In order to increase reliability, some studies used panel discussions after the identification of HPEs, and only verified situations as HPEs where consensus within groups of healthcare professionals and researchers was reached (68, 84-87). Even though a high inter-rater reliability might lead to better reproducibility of results, this does not show if all important HPEs were detected (88). Inter-rater reliability can be high and can still assess only a small proportion of the actual occurrences of HPEs. A systematic review on inter-rater reliability of case note reviews by Lilford et al. (2007) identified that explicit criteria on reporting forms increased inter-observer agreement compared to implicit criteria (88). Consequently, the dependence on subjective judgement can potentially be reduced by defining explicit criteria for a specific set of HPEs and HMEs, such as the described PSIs.

Another vital part of the measurement method to identify HPEs, besides the influence of different reviewers, is the choice of denominator (72). The denominator is needed to express HPEs detected as a proportion of a relevant larger population. Various denominators have been used in studies to estimate medication errors: 'drug order, doses, opportunities for errors, patients, nurses, reports and triggers' (89), 'patient days, number

of admissions and time periods' (90), or no denominator at all in studies that simply report counts of HPEs (74, 89). The selection of younger patient populations with a lower prescription rate and fewer comorbidities, for example, might result in a reduced number of prescribed items per patient, hence less opportunity for HPEs and a potentially lower HPE prevalence. Depending on the objectives for the study it might be more relevant to detect the HPE prevalence per patient stay if the study focuses on risk for single patients in a hospital. In a study focussing on quality of prescribing of healthcare professionals, HPEs per prescribed item might be more appropriate (91). In incident report systems, the denominator is often unknown (74, 89). Results are normally presented in time frames, for instance errors reported within the last year because the denominator is not recorded (92).

In sum, the method of data collection to measure HPEs affects estimates of prevalence in various ways (75, 79). The screening source, the subjectivity of the reviewer and the choice of denominator can impact prevalence estimates. This contributes to a lack of comparability of prevalence estimates as described in the review by Lewis et al. (2009) (72).

2.3 Prevalence of hazardous prescribing

This section provides an overview of studies estimating the overall prevalence of HPEs in primary care in the UK. There are relatively fewer studies that have estimated the prevalence of HPEs in primary care settings compared with secondary care settings (38, 39). A recent systematic review by Assiri et al. (2018) summarised studies on medication errors and error related ADEs in primary care (93). No study in the systematic review estimated the overall prevalence of HPEs. Instead, the prevalence of specific types of HPE, such as PIMs for the elderly, drug-drug interactions or contraindications, was assessed. Out of the 46 studies around the world on HPEs included in the review, 37 reported PIMs in the elderly as described in section 2.1 (93). This restriction to elderly people did not represent the overall prevalence of HPEs. One of the limitations of the systematic review by Assiri et al. (2018) was that it excluded all studies where the numerator/denominator was not defined as errors per patient. To include studies using other denominators and to identify if any new studies were published since then for a UK population, an extended literature search in Medline and Embase was performed. Search terms included: 'error'; 'near miss';

‘preventable adverse event’; AND ‘prescription’ ‘prescribe’; ‘medication order’; AND ‘incidence’; ‘incident report’; ‘prevalence’; ’rate’; ‘epidemiology’; AND ‘UK’; ‘England’; AND ‘primary care’; ‘primary healthcare’; ‘general practitioner’. After screening of the abstracts, only three studies reported the prevalence of HPEs for a general population in the UK’s primary care setting (87, 94, 95). None of these studies were included in the systematic review by Assiri et al. (2018). The studies by Shah et al. (2001) (94) and by Quinlan et al. (2002) (95) were carried out before the study period of the systematic review (2006-2015). Avery et al. (2012) was probably excluded because of a different choice of denominator (prescription items) (87).

The study by Shah et al. (2001) analysed prescriptions from 23 physicians in three general practices (94). A research pharmacist reviewed 35145 prescription items retrospectively that were collected in a two month period. The prevalence of HPEs was estimated to be 7.46% (CI 95% 7.2% to 7.8%), with a wide variation of HPE rates per general practice. Quinlan et al. (2002) used the incident report method to identify HPEs during a two week period (95). In 38 pharmacies, 60525 items were dispensed, and 0.69% of these items were reported as incidents by the pharmacist. The latter PRACTICE study assessed the prevalence of HPEs in 15 general practices (87). 1777 patients were examined with 6048 prescription items. HPEs were identified by a pharmacist using medical records and validated by a research panel containing a GP, a pharmacologist and three pharmacists. A prevalence of 4.0% (CI 95% 3.5% to 4.5%) of prescription items was estimated for HPEs.

The estimated prevalence of HPEs by Quinlan et al. (2002) (0.69%) was low (95), compared to that estimated by Sha et al. (2001) and Avery et al. (2012). In Quinlan et al. (2002), pharmacists completed report forms prospectively to document interventions they made when screening the prescriptions. As this was done in addition to their daily routine, a lack of time might have led the pharmacists to under-report HPEs. This is in accordance with previous findings [2.2] on the effect of incident report, which due to under-reporting does not qualify for the assessment of prevalence. Avery et al. (2012) is the only study of these three where medical records were included in the HPE identification process. Quinlan et al. (2002) and Shah et al. (2001) screened prescriptions without taking the medical records into account. According to previous sections, medical record review is the most appropriate

measurement method to identify HPEs [2.2]. Additionally, the PRACTICE study is the only study that validates the identified errors to reduce subjectivity of the results. Consequently, the most recent multi-centre PRACTICE study was assumed to be the most appropriate to mirror HPE prevalence in GP practices.

Even though interpretation of prevalence estimates might have its limitations, multicentre evaluation of prescribing safety, such as by Avery et al. (2012), provide an overview of the level of HPEs in primary care (87). An HPE rate of 4% scaled up to the English population would result in almost 45 million HPEs per year based on the 1.12 billion prescriptions in England in 2019 (96). Even though 4% seems comparably low, the volume of prescription items in primary care makes HPEs a key safety challenge in England.

2.4 Harm from hazardous prescribing

This section describes consequences of hazardous prescribing and how these have been quantified in the published literature. A patient can experience health and non-health related consequences. These consequences can generate increased costs for the health system or decrease quality of life or patient satisfaction with the health service (30). Solely focusing on prevalence of HPEs has historically not been proven effective in increasing medication safety in healthcare settings (97-100). The majority of HPEs do not cause harm, and a focus on HPEs alone does not reflect the burden appropriately (100). A measurement of harm related to hazardous prescribing is therefore required.

The relationship between an HPE and harm to a patient, such as error related ADEs, can be difficult to estimate due to ethical and methodological challenges. With respect to methodological challenges, large sample sizes are required to estimate the relationship prospectively with sufficient power because the proportion of HPEs that lead to ADEs is low. In addition, depending on the pathology of the HPE, there can be a delay between the occurrence of an HPE and the observed health outcome, which can increase the required study follow-up period in prospective designs. With respect to ethical challenges, follow-up of patients, where potentially harmful HPEs are identified, with ethical obligations to intervene, are not an option. Therefore, attempts have been made to assess potential harm

from HPEs by more indirect measures. One way is to work back from ADEs. The assessment of causality and preventability of ADEs can estimate preventability of drug related morbidity (PDRM) [2.4.1]. Another way is to estimate the harm that could potentially occur from an HPE. The potential harm can be elicited by experts or observational data could be used to link HPEs with outcomes retrospectively [2.4.2].

2.4.1 Assessment of preventability of drug related harm

The first measure to estimate harm from HPEs was the retrospective assessment of PDRM. PDRM can be used as a proxy to identify the burden of HPEs because medication error related harm is considered preventable. ADEs detected are analysed if they could be caused by the HPE (causality) and if the harm outcome would have been preventable (preventability).

Causality assessment

Important factors in causality assessment are (i) temporal relationship of drug use and the ADE, (ii) physiological plausibility, (iii) de-challenge, and (iv) re-challenge (101). De-challenge is the effect of removing the drug exposure on the ADE presence, such as a rash that disappears after drug is stopped. Re-challenge is the effect of restarting the drug on the ADE presence, such as the restart of the drug causes the rash again.

Different tools or algorithms can be used to estimate the probability that the drug contributed to the ADE (102-106). A systematic review of causality assessments distinguished between expert judgement, algorithms and Bayesian approaches (107). This section briefly explained these assessment methods with an example and subsequently discussed strengths and limitations of these methods.

An example for expert judgement is the criteria set by the WHO, called the WHO-UMC System for Standardised Case Causality Assessment, that aims to classify the likelihood of causality between an ADE and a drug (108). The WHO-UMC system characterises the likelihood of causality on a five-category scale: probable, possible, unlikely, conditional or unclassifiable causality. The terms ‘conditional and unclassifiable’ describe cases where

more information is required to assess the probability of causality of the ADE. Another common causality assessment tool was the algorithm developed by Naranjo et al. (1981) that uses a weighted scoring system based on ten questions about the ADE and was found to be valid and reliable (106). The score characterises the likelihood of causality on a four-category scale: definite, probable, possible or doubtful causality. An example of a Bayesian assessment is the Bayesian Adverse Reaction Diagnostic Instrument (BARDI) (109). The tool calculates the probability of causation by combining clinical data (prior probability) and case-specific information. Expert judgement is considered the least reliable method to assess causality (107). Compared with the WHO-UMC assessment criteria, the inter-rater reliability with the Naranjo algorithm score was found to be the highest (110). Potential reasons could be the requirement to include personal experience when using the WHO scale that increases subjectivity of the assessment. The authors of the review by Agbabiaka et al. (2017) consider both algorithm-based and expert judgement-based tools as unreliable on the basis of subjectivity of the results. Inconsistent results can be a cause of different levels of clinical experience, profession or different interpretations of categories in e.g., the WHO-UMC system or of questions in the Naranjo algorithm (111). For example, nurses were found to identify fewer ADEs as 'certain' or 'definitely' associated with the drug, compared with physicians and pharmacists in the Naranjo and the WHO-UMC tool (110). Bayesian approaches were considered more reliable because they do not rely on expert input alone. However, they require a substantial time commitment, resources, complex calculations and depend on the availability and quality of available epidemiological data for the prior probability (101, 107).

In summary, causality assessment between an ADE and a drug can be difficult and there might still be other causes for the experienced harm (54, 112, 113). None of the causality assessment methods were considered to provide reliable, reproducible estimates on the probability of causality due to the 'unavoidable subjectivity of judgements' (107).

Preventability assessment

Different tools can be used to estimate preventability of ADEs. A detailed description of different type of tools were reported by Hakkarainen et al. (2012) (111). In this section, only the two most common tools are described. According to a systematic review on preventable drug related ADEs 59% of studies assessed preventability using the method proposed by Hallas et al. (1989) or by Schumock et al. (1992) (114, 115). Hallas et al. (1990) proposed specific criteria to group ADEs as definitely, possibly or not preventable. An ADE was definitely avoidable for example, if 'the drug event was due to a drug treatment procedure inconsistent with present-day knowledge of good medical practice or was clearly unrealistic, taking the known circumstances into account'. Schumock et al. (1992) propose and algorithm to determine whether the ADE was preventable or not. The algorithm included statements on contraindications, drug-drug interactions, inappropriate dosing or monitoring, not considering recorded allergies to the prescribed drug, toxic serum drug concentration and treatment compliance. An ADE was considered preventable if one of the statements was confirmed.

Limitations of tools previously used to assess preventability were similar to those identified for causality assessment (116) and were mainly related to reliability of the assessment methods (111).

Assessing preventable drug related medication harm

Overall, estimating causality and preventability is a complex process that requires detailed information on the ADE and the patient, and decisions have to be made case by case. For large numbers of ADEs, this is a time consuming and costly process that is only as good as the available data (101, 107). Comparability of the results is often not given due to the dependence on the type of assessor and the absence of a standardised method to establish preventability (111) or causality (110) that resulted in strong variations of reliability of studies. So far there is no gold standard to generate comparable estimates on causality and preventability of ADEs (101, 111).

Prevalence of preventable drug related harm

Several systematic literature reviews and meta-analyses on preventable harm were published recently. Panagioti et al. (2019) reported a systematic review and meta-analysis of preventability of any type of patient harm, not specifically PDRM (117, 118). Hodkinson et al. (2020) reported a systematic review and meta-analysis of PDRM (116). Alquena et al. (2020) reported a systematic review of patient harm from medication errors at the transition of care from secondary care to primary care (119). Overall, preventable harm occurred in 6% (95% CI 5% to 6%) of patients (118). Half of these were drug related, hence a consequence of a HPE (37, 116). The meta-analysis by Hodkinson et al. (2020) found a 3% (95% CI 2 to 4) prevalence of PDRM. Of these 26% (95% CI 15% to 37%) were severe or life threatening. The prevalence of PDRM increased with age and was higher in secondary care settings than in primary care.

Table 2.2: Overview of studies assessing the probability of drug related harm outcomes and the percentage of these that are considered definitely or possibly preventable in a UK setting

Study	Harm outcome measure	Sample size	Total harm (%)	Percentage of harm outcomes considered preventable (%)	
				Definitely preventable[f]	Possibly preventable[g]
Howard 2003 (37)	Drug related hospital admissions	4093	6.5	4.3[a]/67.0[c]	N/A
Morris 2004 (120)	Potentially drug related morbidity in general practice	49658	N/A	1.0[b]	N/A
Pirmohamed 2004 (54)	Hospital admissions due to ADR	18820	6.5	9.0[c]	63.0[c]
Davies 2009 (121)	ADRs in hospital inpatients	3322	15.8	6.4[c]	46.9[c]
Davies 2010 (112)	Re-admission to hospital due to ADR	290	20.8	14.3[d]	42.9[d]
Gallagher 2012 (113)	Admission to a children's hospital due to ADR	6821	2.9	7.6[e]	30.2[e]
Tangiisuran 2012 (122)	ADRs during patients stay at elderly care wards	560	13.2	69.0[h]	N/A

Denominator for: [a]all hospital admissions, [b]all primary care medical records, [c]total number of ADRs, [d]all ADR related readmissions, [e]all ADR related admissions; [f]'The drug event was due to a drug treatment procedure inconsistent with present day knowledge of good medical practice or was clearly unrealistic, taking the known circumstances into account.'(114), [g]'The prescription was not erroneous, but the drug event could have been avoided by an effort exceeding the obligatory demands.'(114), [h]all serious and life threatening ADRs; ADE: adverse drug event; ADR: adverse drug reaction

From the identified studies in the three systematic reviews, this literature review reports the studies from a UK setting (n=7) in detail. The studies assessed preventability of ADEs (37, 120) or adverse drug reactions (ADRs) (54, 112, 121-123) (Table 2.2). The harm outcomes are detected in a secondary or primary care setting, and both relate to medications or drugs prescribed in primary care. A UK study by Parekh et al. (2018) was not included because it included medication related harm as a result of non-adherence, and harm from medication errors was not reported transparently (124).

ADEs are 'any undesirable event experienced by a person while they are having a drug or any other treatment or intervention, regardless of whether or not the event is suspected to be related to or caused by the drug, treatment or intervention' (125). In contrast, ADRs imply a causal relation between drug and occurrence. The WHO has defined an ADR 'as a response to a drug that is noxious and unintended and occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or modification of physiological function' (126). As a result, all ADR are considered ADEs, but not every ADE is an ADR.

Howard et al. (2003) found 67% of ADEs to be preventable, resulting in 4.3% of all admissions to be preventable (37). Pirmohamed et al. (2004) found similar results for ADRs with 72% of ADRs being potentially or definitely preventable (54). Comparability with other estimates as seen in Table 2.2 is difficult due to the diversity of chosen denominators. For instance, Morris et al. (2004) (120) defined the denominator as all primary care medical records in contrast to Pirmohamed et al. (2004) and Howard et al. (2003) that used the total number of ADRs.

2.4.2 Assessment potential severity of hazardous prescribing

The second measure to estimate harm from HPEs was the assessment of potential future harm. In this method, the assumed severity of HPEs is estimated. A systematic review by Garfield et al. (2013) on severity assessment methods identified various different assessment tools (127). In 17 studies performed in the UK identified in the review by Garfield et al. (2013), 14 different tools were used (68, 128-143). Severity scales in these tools ranged from two to eight severity levels or were continuous scales. All UK studies,

except Webbe et al. (2007), included 'minor', 'insignificant' or 'low' harm as severity levels of HPEs. Webbe et al. (2007) excluded minor HPEs in the severity assessment (134). Two studies used a detailed lists of what kind of HPE belonged to which severity level (131, 133). Overall, methods were found to produce highly variable severity ratings, reliability of the tools was often low, and validity in other settings were not assessed (127, 144).

Potential severity of HPEs in the UK

The PRACTICE study is the only UK study rating severity of HPEs in primary care GP practices (87). The median severity, evaluated using the visual continuous scale (1-10) by Dean et al. (1999), was 3.0 for HPEs. Consequently, most errors are of minor (score of 0-2.9) to moderate severity (score of 3-7). The CHUMS study in nursing homes uses the same severity scale as the PRACTICE study, which allows comparability of severity results to some extent (145). Nevertheless, ratings still depend on the qualification, training and number of assessors and are therefore not directly comparable. In accordance with findings from the PRACTICE study, most HPEs in the CHUMS study are judged to be minor to moderate with a mean severity score of 2.6 compared to a mean severity score of 3.5 in the PRACTICE study. Whereas the highest score in the CHUMS study was 6.6, the highest rating in the PRACTICE study was 8.6. This may imply that, on average, HPEs in general practices are more severe than in care homes.

2.5 Economic impact of hazardous prescribing

This section describes how hazardous prescribing can affect patient outcomes and costs. A systematic review by Walsh et al. (2017) on the economic impact of medication errors, updated by Elliott et al. (2018), found evidence assessing patient outcomes and healthcare resource use associated with HPEs to be scarce (31, 146).

Health related patient outcomes

Patient health can be impacted by HPE related ADEs, lack of treatment efficacy and suboptimal patient adherences (23). These can directly affect patient outcomes, such as health status, health related quality of life and death. Changes in these outcomes can be

measured in life-years gained or in QALYs. QALYs describe not only the quantity of life but also the quality (147). Health outcomes could also be measured indirectly as an increase in the number of hospital admissions, or specific ADEs related to the HPE, such as strokes or GI bleeding events. Only few studies reported direct measures of patient outcomes associated with HPEs (31).

One study measured health related quality of life dependent on exposure to PIMs in Ireland (2017) (148). The exposure to PIMs as defined by the STOPP and Beers criteria and was assessed using pharmacy claims data. The questionnaire was sent by post and results were available for 663 (98.4%) patients. The study identified a significant reduction in the EQ-5D utility for patients with more than two PIMs compared with patients without PIMs (beta-coefficient -0.11, 95% CI -0.16 to -0.06). While this shows that the use of more than one PIM might be independently associated with decreased utility, this does not actually link HPEs and patient outcomes. The negative association between EQ-5D and PIM could be confounded by disease severity.

Studies linking HPEs with patient outcomes either relied on expert elicitation or focused on specific HPE types. Karnon et al. (2009) elicited QALY decrements associated with HPEs by severity level. For significant, serious, severe or life-threatening preventable ADEs, hypothetical QALY decrements and effect durations were estimated in discussions within the research team. A similar approach was taken by Yao et al. (2012). This form of using QALYs attached to HPEs introduces a high level of uncertainty around the estimates (31). Other studies focused on specific type of HPEs to estimate patient outcomes (149-152). If specific types of HPEs are used, the ADEs associated with these can be linked with QALY estimates from the literature. An ADE associated with a type of HPE known to increase the risk of GI bleeding events, can be linked directly to a utility. Evidence on quality of life decrements associated with GI bleeding event incidence can then be used instead of quality of life estimate of HPEs in general.

Costs associated with ADEs

This section starts with a description of what information an ideal study assessing cost of HPEs would contain, followed by a brief overview of the existing evidence that provides this information. The systematic review on the economic impact of medication errors by Walsh et al. (2017) provides a detailed overview of how resource use was estimated when assessing the economic impact of HPEs. This was further discussed in the EEPRU report by Elliott et al. (2018) (31). Details of studies reporting cost of HPEs are therefore not be repeated here. This section ends with the description of an example study that aimed to estimate the cost of medication errors in the UK to demonstrate how studies deal with the sparsity of data available on the economic impact of HPEs.

HPEs that cause ADEs are likely to have economic consequences, due to an increase of the use of health services as a result of, e.g., an increase in the number of hospital admissions (23). The resource use associated with HPEs is dependent on the incidence of HPEs and the related health services required to manage ADEs. Ideally, detailed information on (i) how many HPEs reach the patient that are not intercepted before the patient takes the medication, (ii) how many of these lead to ADEs that require additional health services, and (iii) what costs these generate are available. No studies were identified that provided the ideally available information. Cost had to be attached to intermediate outcomes, such as preventable ADEs or severity levels of ADEs. Costs were most often assessed for generic outcomes related to HPEs, e.g., the number of hospitalisations and/or medication related costs (146, 153). However, no standardised approach to the costs of HPEs was apparent. The EEPRU report evaluated all these studies as moderate quality evidence stating that there is a lack of good quality data. A review on the methods of calculating cost of HPEs came to a similar conclusion (153).

In a UK setting, cost estimates exist on some specific error types. Administration errors in anaesthesia (154), prescribing errors related to inhaler prescriptions (155), costs of dispensed PIMs in the elderly (57), or preventable ADEs at hospital admission state (156). However, these estimates usually represent a specific time frame or setting. They either do not report a cost per error (57, 154) or focus solely on medication cost not including the cost required to manage ADEs (155).

The EEPRU report aimed to estimate the costs associated with HPEs in the UK overall. Due to lack of data linking medication errors directly with ADEs, the resource use associated with medication errors in primary care had to rely on literature around preventable ADEs as described in section 2.4.1. Incidence of hospital admissions due to PDRM, for instance, were based on results from Pirmohamed et al. (2004) (54) and Howard et al. (2003) (37) [Table 2.2]. From Pirmohamed et al. (2004) the average number of days at hospital in patients with preventable ADEs was compared with a national average of length of hospital stay to estimate the additional cost of medication errors. Results on the healthcare cost to the NHS related to HPEs show a high level of uncertainty. The authors explain this uncertainty with the sparse data available. In order to estimate the healthcare costs associated with HPEs in more detail, a direct link between HPEs and ADE related resource use is required.

2.6 Non-health impact of hazardous prescribing

The evidence described in section 2.5 focused on health related consequences and associated costs of HPEs. As a result, the impact of HPEs is restricted to its impact on health related outcomes. Expressing the impact of reducing HPE rates on patient outcomes solely by its effect on health status, measured in QALYs, might underestimates the overall value of reducing HPE rates. Improving safety by reducing HPE rates is associated with benefits other than maximising health (157). Patients might be impacted by other aspects of unsafe prescribing than harm from ADEs. Decision makers not only consider impact on these health outcomes but also enhancements of patient experience, patient empowerment and improvement of public trust and confidence as outcomes (158). Trust issues as a consequence of HPEs in the health system have been reported (159-161). Patients need to trust healthcare professionals with personal and potentially intimate problems and trust the competencies of the physician (162). Experiencing an ADE due to a HPE can reduce trust in the decisions of the prescriber. Trust can be undermined towards the healthcare provider, such as the GP, or the healthcare system itself. Consequences of undermining trust could lead to patients not utilizing required healthcare services or not adhering to the proposed treatment (163-165). A meta-analysis of the effect of trust in the healthcare system on health outcomes reported that increased trust was associated with healthier

behaviours, an increased quality of life and satisfaction with the treatment with less severe symptoms (166). However, contributors to diminishing trust of patients have not been investigated widely (167) and evidence on regaining trust after disclosure of mistakes is sparse (168).

Estimations of consequences for the healthcare system caused by HPEs so far did not include decreased trust in the healthcare system or other non-health outcomes related to improved safety (169, 170). The effect of trust on health outcomes and its association needs to be investigated further because evidence so far is weak or of low quality (166, 167, 169, 171). Other potentially relevant aspects of safety improvements that require further research, were assessed by Steuten et al. (2010) (157). The authors identified specific attributes patients consider relevant in addition to the prevalence, health status and costs associated with prevention of HPEs and safety incidences in general. Preventability of harm, dread and controllability, as well as trust in the intervention were identified as additional attributes.

A literature review by Perry-Duxbury et al. (2019) of valuation methods in safety identified two methods used to estimate the value of safety in health (170): (i) discrete choice experiments (DCEs) and (ii) contingent evaluation. Stated preference methods, such as DCEs, are increasingly used in economic evaluations to estimate patient preferences (172, 173). DCEs ask patients to choose between alternative hypothetical scenarios that differ in specific attributes. A patient could be asked to choose between scenarios with different levels of attributes, e.g., health related consequences, preventability of incidences and trust in the safety system as identified by Steuten et al. (2010). In DCEs, preferences can be revealed without explicitly asking for the preferred levels for each attribute. Contingent valuation surveys also allow to identify the monetary valuation of effects in healthcare (174, 175). In this stated preference model, patients can express their preference for hypothetical scenarios in monetary values. Survey questions directly ask the patient to state the willingness to pay (WTP) for a non-monetary good, such as safety improvements. In comparison, DCEs indirectly assess patient's valuation by analysing the choices patients made between different scenarios.

Both methods require a clear definition of what is being valued, which can be difficult with regards to improvements in safety. The review by Perry-Duxbury (2019) found that studies valuating safety defined safety as a reduction in event risks or reduction in event cases. The studies did not include the feeling of safety or trust in the definition. The authors also report the challenges in designing and analysing DCEs and contingent valuation surveys and highlight the need for a more standardised approach to be able to compare results among studies. This could involve more standardised definitions of safety, guidance on the order of questions, relevant attributes and the level of information provided. A challenge particularly for contingent valuation was embedding effects with regards to safety interventions (176). Asking how much a participant is willing to pay for a pharmacist service that reduces the number of HPEs implicitly asks their WTP for the pharmacist service, a reduction of the probability of getting a hazardous prescription and a reduction in the risk of ADEs from HPEs. This was for example done in a Malaysian study that found 67% of 100 participants to be willing to pay for a pharmacist service and estimated a WTP of ten Malaysian Ringgit equivalent to £1.42 (177). A challenge reported for contingent valuation methods was that participants are often willing to pay the same for the combination of all these different effects, as for them individually (170, 178). Overall, studies valuating safety in healthcare are rare and so far, no studies used the available methods to assess the value of non-health benefits of safety using preference-based methods.

2.7 Overview of interventions aiming to reduce hazardous prescribing

This section provides a brief overview of the type of interventions that were developed to reduce hazardous prescribing in primary care with examples of how these interventions are constituted. Subsequently, the effectiveness of interventions and aspects contributing to varying results are briefly described. The aim is to create an understanding of existing interventions, before section 2.8 reports a systematic review of the literature around economic evaluations of interventions aiming to reduce hazardous prescribing. Interventions were grouped as (i) educational interventions, (ii) computerized alert systems, (iii) electronic audit and feedback interventions (e-A&F), (iv) pharmacist-led interventions, and (v) multifaceted interventions.

Educational interventions

Educational interventions provide interactive workshops, provision of educational material with audit data or practice visits to educate patients, carers or healthcare professionals (179). An example of an educational intervention was developed by Pimlott et al. (2003) that discusses the results of an audit of proton pump inhibitor (PPI) use with GPs to reduce inappropriate prescribing of these medications (180, 181). Another interventions provided audit data via mail aiming to reduce prescribing of benzodiazepines (181).

Computerised alert systems

Computerised alert systems utilize patient information, e.g., from electronic health records, and provide advice to the healthcare professional during encounters with the patient. The given advice can be variable, for example, on treatment alternatives, monitoring reminders or dosing suggestions, and is usually based on some form of best practice as proposed by guidelines or clinician consensus. Computerised alert systems are used at the point of patient care and alerts are flagged during the consultation.

Electronic audit and feedback interventions

In contrast to computerised alert systems that are used at the point of patient care, e-A&F are delivered outside of clinical consultations (182). Dealing with the computerised alert system during consultations can interrupt the workflow and can be time consuming. This can cause overriding of the alerts without considering them and reduces face-to-face time with the patient (183). A study by Gurwitz et al. (2014) investigated an e-A&F intervention that informed the GP about discharge medications of patients discharged from hospital including a medication review and an alert system to schedule a GP follow-up visit (184). Another e-A&F intervention screened medical records of patients with mild hypertension and displayed suggestions for optimal antihypertensive therapy based on the relevant guidelines (185).

Pharmacist-led interventions

Pharmacist-led interventions propose new roles of pharmacists to increase prescribing safety (179, 186-189). Medication review by the pharmacist is the most common pharmacist-led intervention (179, 186). Feedback can be given to either the patient with a subsequent consultation with the prescriber (190-192) or to the prescriber with subsequent information of the patient on agreed medication changes (193-196). The consultation with the prescriber can be oral (192), in written form (196) or both (193, 195, 197). Most pharmacist-led studies used defined sets of HPEs by Beers [2.1] (198, 199) or the medication appropriateness index criteria (MAI) criteria (192-195, 197-199) to guide medication reviews and measure the effect of the intervention.

Multifaceted interventions

Multifaceted interventions combine aspects of multiple intervention types. For example, in addition to a medication review by pharmacists, some studies included further education of the patient by the pharmacist on adherence or the care plan (193-195). Another example of a multifaceted intervention is PINCER that combines aspects of an e-A&F intervention in form of an audit report and a pharmacist-led medication review (44, 45). A software creates a report that highlights patients with HPEs by screening electronic patient records from the practices. The pharmacist intervention comprises a review of the medications of patients flagged in the report, meetings with GPs to discuss results of the HPE report and to agree actions required to solve the prescribing hazards in the report.

Effectiveness of interventions aiming to reduce hazardous prescribing

A large number of reviews has recently been published that compare the design and effectiveness of interventions preventing hazardous prescribing in primary care (200). Evidence on effectiveness of interventions aiming to reduce hazardous prescribing has been mixed (179, 186, 187, 201-203). Aspects contributing to this mixed effect were often the chosen outcome measures or challenges associated with the specific type of intervention.

Unspecific outcome measures that can be influenced by various other factors made it difficult to identify differences attributable to the intervention types (204). The systematic review and meta-analysis of interventions targeting medication errors in primary care by Khalil et al. (2017) reported that interventions often showed an improvement on process indicators, such as HPE rates or medication changes; however, changes in hospital admissions or mortality were small or not statistically significant (187). The authors reported that this is a result of small sample sizes and that further research is required with larger studies with higher quality study designs that test patient related outcomes of interventions aiming to reduce HPEs. A focus on more specific patient outcomes related to the prescribing hazards, such as drug related admissions instead of any admission, could be beneficial to demonstrate the impact of the interventions (204).

The challenges associated with implementing different types of interventions could also be contributing to the mixed effectiveness of interventions. Educational interventions showed varying effectiveness due to their heterogeneity and different outcome measures that often focused on testing prescribing skills rather than measuring HPEs avoided in routine care (205, 206). Computerised alert systems and e-A&F were found effective in reducing medication errors in only half of the interventions (202). Multiple studies investigated the challenges with implementing these types of interventions and found that embedding the interventions in existing routine care, as well as acceptance by users are key factors contributing to the varying effectiveness (207-209). Most effective were computerised alert systems and e-A&F interventions that focus on a limited number of hazards and clearly define these (202). Pharmacist-led interventions demonstrated effectiveness only in a few studies and reviews found the quality of evidence to be low (179, 186). Communication and integrating pharmacists into multi-disciplinary teams were reported as important factors for effective pharmacist-led interventions (186, 210). Multifaceted interventions that included aspects of multiple intervention types were found to be most effective in reducing hazardous prescribing (179, 186, 203). An RCT comparing the multifaceted intervention PINCER that combines aspects of e-A&F and pharmacist-led interventions with an intervention only providing the e-A&F aspect of the PINCER intervention identified PINCER as the more effective and cost-effective strategy (45).

Overall, effectiveness of interventions does not only vary between different types of interventions but also within these types. This chapter identified key aspects that were associated with beneficial interventions in primary care: (i) combining multiple intervention types; (ii) pharmacist-led services can contribute to the effectiveness of information technology interventions, e.g., clinical alert systems and e-A&F; (iii) outcome measures should be specific to the hazards the intervention aims to reduce.

2.8 Systematic review of economic evaluations of interventions aiming to reduce hazardous prescribing

This section begins with an overview of economic evaluations in general and ends with a detailed review of the existing evidence on cost-effectiveness of interventions aiming to reduce hazardous prescribing.

Economic evaluations in healthcare

Economic evidence is required to inform healthcare resource allocation decisions at the population-level. Economic evaluations typically compare alternative strategies in relation to costs and patient outcomes (4). The European Council highlighted the need for evidence especially on the cost-effectiveness of interventions in patient safety (16). Different types of economic evaluations exist. Cost-benefit analysis measures the benefits in monetary units. Cost-effectiveness analysis (CEA) evaluates the costs and health outcomes of interventions, where health outcomes are expressed as natural units. Results are presented as cost per unit of health outcome gained (4). A special case of CEAs are cost-utility analyses (CUA). The health outcome is described as a preference-based valuation of outcome, such as quality of life. QALYs can be used as a measure to assess the quality of life. On a scale from zero to one quality of life scores can be used to weight each year of life remaining for a patient in a specific health state (125). The use of QALYs is encouraged by decision maker bodies because it enables comparison across diseases and type of interventions (3). The CUA identifies the QALY gain by the intervention and the associated opportunity cost. If the intervention is less costly and more effective than the comparator, the intervention is considered to be the dominant strategy. A second possibility for cost-effectiveness is to

prove that the cost of the QALY gain is reasonable. In CEAs assessing cost-effectiveness of an intervention compared with standard care, this can be shown by the extra costs per additional QALY generated by the intervention, the incremental cost-effectiveness ratio (ICER). Decision makers, such as NICE, provide a threshold for this ICER to define the maximum WTP for each unit of effect (QALY) generated. The ICER enables decision makers to compare different interventions and to decide how to allocate resources. NICE recommends treatments for use in the NHS if the ICER is not greater than £20000 - 30000 per QALY (3).

Economic evaluations of interventions reducing hazardous prescribing

A systematic review was conducted by the author of this thesis alone to identify existing economic evaluations that assessed the cost-effectiveness of interventions designed to reduce medication errors. The aim of this review was to inform the *de novo* economic evaluation of SMASH in how to incorporate medication errors into the model. To meet the aim, the following objectives were set:

- Objective One: identify all published model-based economic evaluations of medication errors;
- Objective Two: compare models on the basis of structural assumptions to incorporate medication errors;
- Objective Three: compare approaches on the basis of how they incorporate harm from medication errors.

Search strategy

Inclusion criteria for the systematic review were developed to include all economic evaluations where different levels of medication error rates were modelled. This included medication errors in the whole medication use process: (i) prescribing errors, (ii) dispensing errors, (iii) administration errors, and (iv) monitoring errors. A study was included if the economic evaluation analysed comparators with different medication error rates using a decision-analytical model. The focus on decision-analytical models was based on the importance of these for the conceptualisation of a *de novo* model based economic evaluation. Articles were eligible if they were written in English or German and were full

text publications. Search databases were EMBASE and Medline because they are the only recently updated databases recommended by NICE and the Cochrane handbook (211). The NHS EED database stopped being updated in 2015 and was therefore not be used for this review. For the systematic review the databases Medline (1946 to December week 4 2019) and Embase (1974 to 3 January 2019) were searched on 4 January 2019.

To identify key concepts for the review search strategy, it was planned to use the PICO criteria (212) as recommended by NICE (213). The main concepts were based on adapted PICO-S criteria (Population-Intervention/exposure-Comparator-Outcome-Study type). The concept of population and comparator were not prespecified because they were not relevant in this search. The search strategy was based on search terms for the concepts: (i) Intervention/Exposure: 'Medication error'; (ii) study type I: 'decision-analytical model'; (iii) Study type II: 'Health economics'. The focus of the review was on the study design and therefore two concepts to precisely identify the study type of interest, which was economic evaluations based on decision-analytic models, were chosen. The search terms for each of the concepts consisted of free-text and index terms. Search terms for concept (i), medication error, consisted of a search strategy previously used by Elliott et al. (2018) (31) and Walsh et al. (2017) (146) with additional use of index terms appropriate for each database and free text search on hazardous prescribing. For concept (ii) 'decision-analytic model', a keyword list previously used in NICE guidelines was applied (214, 215). Search terms for concept (iii) 'Health economics' were combined from the Centre for Reviews and Dissemination (216) as suggested by NICE (213) and from previous NICE guidelines that used an identical search strategy for 'health economics' (214, 215, 217). Search terms for each category can be found in Appendix A.

Results of systematic literature review

The search based on the search terms for 'medication error' AND 'decision-analytic model' AND 'health economics' found 368 articles after excluding duplicates, conference abstracts, and non-English or non-German articles. A flowchart of the review process is presented in Figure 2.1. After the articles were screened by the author of this thesis, 13 studies were identified that fulfilled the inclusion criteria (149-151, 156, 218-226).

Figure 2.1: Flowchart of economic evaluations included in the systematic review

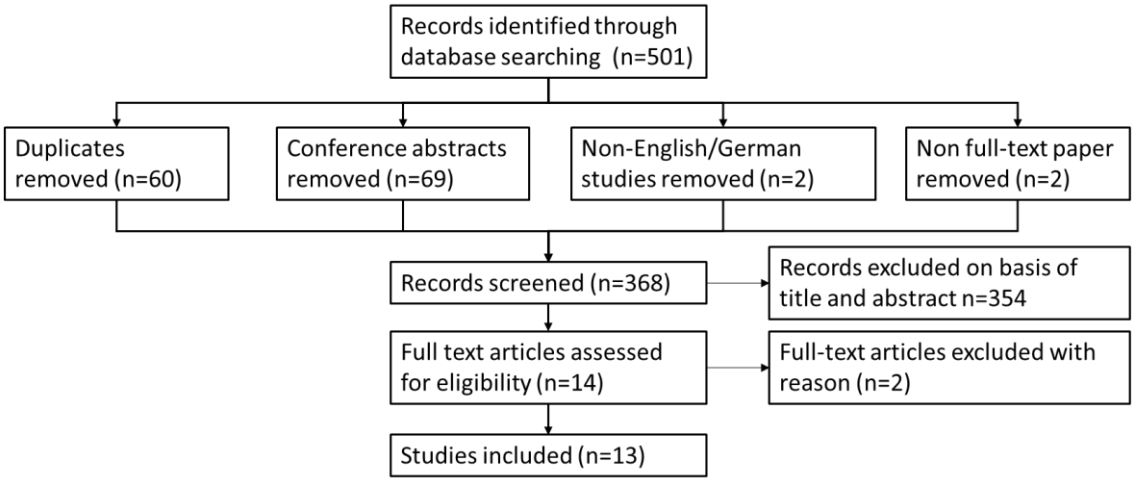


Table 2.3 summarises study characteristics from full text screening. Most economic evaluations investigated interventions in secondary care settings. Only Elliott et al. (2013), Forrester et al. (2014), Foster et al. (2018) and Moriarty et al. (2019) (4 out of 13) looked at medication errors in primary care. The majority of studies were from the US (4 out of 13) or the UK (4 out of 13). Intervention types were clinical alert systems (218, 222, 225), pharmacist interventions (219, 221) or multifaceted interventions (149, 156). Two studies looked at the effect of hypothetical interventions effecting medication error rates (151, 224). Roselli et al. (2014) and Forster et al. (2018) compare different drug administration processes and Maviglia et al. (2014) compared bar-coding-based dispensing with standard drug dispensing methods. In contrast to these studies looking at interventions or hypothetical interventions that aim to reduce medication errors, Samp et al. (2014) and Moriarty et al. (2019) investigated the difference between medication error and no medication error. The majority of economic evaluations used a decision tree (8 out of 13). State-transition models were used in only three studies (149-151). In the following sections, the identified studies are compared with regards to structural assumptions (i) made to incorporate medication errors [Objective Two], (ii) on how harm from medication errors was linked with health outcomes and costs [Objective Three], or (iii) on evidence on robustness of the results to assumptions around the consequences of medication errors.

Table 2.3: Study characteristics of economic evaluations identified in the review

Reference	Setting	Type of economic analysis	Perspective of cost	Intervention under investigation and comparators	Type of model	Error type and subdivision categories if any	Result measure
Elliott 2013 (149)	Primary care, UK	CUA	Healthcare provider	Pharmacist-led IT intervention vs. simple feedback	Two stage model decision tree and Markov model	Prescribing, monitoring (pre-defined set)	Incremental cost per QALY
Forrester 2014 (218)	Primary care, US	CEA	Healthcare provider	Electronic prescribing vs. hand prescribing	Decision tree	Prescribing (clinical or administrative error)	Incremental cost per error or per ADE
Forster 2018 (150)	Primary care, UK	CUA	Healthcare provider	Population with no handling errors vs current inhaler 1 vs current inhaler 2	Two state Markov model	Administration (one specific error)	Incremental cost per QALY
Ghatnekar 2013 (219)	Secondary care, Sweden	CUA	Healthcare provider	Multi-disciplinary team model for (i) review process at admission; or (ii) discharge report vs current practice	Decision tree	Any medication error	Incremental cost per QALY
Karnon 2009 (156)	Secondary care, UK	CUA	Healthcare provider	5 pharmacist-led IT medicines reconciliation interventions vs. standard care	Decision tree	Any medication error (error due to omission, commission or allergies)	Incremental cost per QALY
Maviglia 2007 (220)	Secondary care, US	CA	Healthcare provider	Bar-code assisted medication-dispensing vs normal dispensing	Decision tree	Dispensing	Net cost[a]
Moriarty 2019 (151)	Primary care, Ireland	CUA	Healthcare provider	Potentially inappropriate prescribing versus appropriate prescribing	Markov model	Prescribing error (pre-defined set)	Incremental cost per QALY

Reference	Setting	Type of economic analysis	Perspective of cost	Intervention under investigation and comparators	Type of model	Error type and subdivision categories if any	Result measure
Nerich 2013 (221)	Secondary care, France	CA	Healthcare provider	Pharmaceutical analysis, vs no pharmacist involvement	Decision tree	Prescribing (in antineoplastic drugs)	Net cost [a]
Nuckols 2015 (222)	Secondary care, US	CUA	Societal perspective	CPOE vs paper ordering	Decision tree	Any medication error	Cost and QALYs gained
Rosselli 2014 (223)	Secondary care, Colombia	CC	Healthcare provider	4 drug administration systems	Decision tree	Administration (only with dopamine)	Errors leading to harm, errors causing death
Samp 2014 (224)	Secondary care, US	CC	Healthcare provider	Medication error vs. no medication error	Decision tree	Any medication error	Cost per error
Westbrook 2015 (225)	Secondary care, Australia	CEA	Healthcare provider	Electronic medication management system vs paper-based prescribing	Decision tree	Prescribing	Incremental cost per ADE avoided
Yao 2012 (226)	Secondary care, UK	CUA	Healthcare provider	Educational intervention for patient discharge vs no intervention	Decision tree	Any medication error	Incremental cost per QALY

[a] Aggregated cost including cost savings due to avoided ADEs and cost of intervention; ADE: adverse drug event; CA: cost analysis; CC: cost consequence; CEA: cost-effectiveness analysis; CUA: cost utility analysis; ENB: expected net benefit; EMB: expected monetary benefit; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life-years

Structural assumptions to incorporate medication errors [Objective Two]

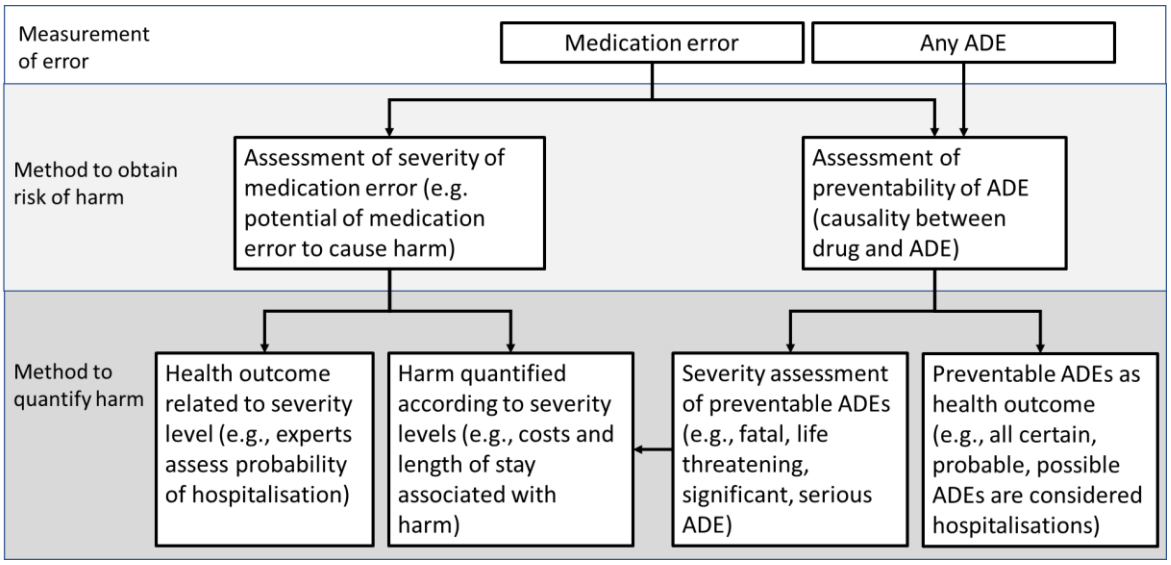
This review looked at medication errors in any part of the medication use process. Types of medication errors used in the models were in most of the cases either any medication error (156, 219, 222, 224, 226) (5 out of 13) or any prescribing error (218, 225) (2 out of 13). Some studies focused on a subset of a specific error group (149-151). Elliott et al. (2013) and Moriarty et al. (2019) analysed a predefined subset of medication errors based on likelihood and seriousness of error types, similar to Forster et al. (2018), who investigated one error that had shown an effect on harm outcomes in a previous study. Other studies focused on only one drug or drug group that was representative of the setting (221, 223). Karnon et al. (2009) combined prescribing, dispensing and administration errors into one group but categorised them by the source of error whether it was an act of omission, commission or due to an undetected allergy (156). The studies used different effect sizes of the interventions. Most often the impact of the intervention on error rates was used (149, 150, 156, 218, 221-225). However, Gathnekar et al. (2013) and Maviglia et al. (2007) applied a more indirect measure of medication errors and used ADEs caused by or potentially caused by the medication error.

Methods to link medication errors with health outcomes [Objective Three]

Objective Three incorporates how medication errors were linked to harm in form of health outcomes or costs in decision-analytic models in the literature. None of the studies prospectively followed patients with a medication error to observe their direct effect on ADE rates or applied results from such a study from the literature. All estimations of ADEs were generated retrospectively or were projections of potential future harm by healthcare professionals even if evidence was derived from a single primary study. Medication errors were linked with patient outcomes either indirectly or directly. Direct measurement refers to studies that directly link the error with a specific ADE that can be clearly associated with the medication error (increased likelihood of specific ADE). The specific ADE, such as GI bleeding, has a measurable impact on costs and patient outcomes that can be included in the economic evaluation. Indirect measurement refers to estimating the potential severity of the identified medication errors (assessment of severity) or identifying all ADEs and estimating whether these can be linked to the error and were preventable (assessment of

preventability). These two methods were described in sections 2.4.1 and 2.4.2. The difference of these indirect measurements compared with direct measurements was that multiple steps were required to interpret the severity levels or the preventable ADEs. These usually involved several assumptions that induced further uncertainty. Methods applied in the economic evaluation are reported in Figure 2.2. The risk of harm was required either by assessing the potential severity of errors or the preventability of ADEs, as well as some form of quantification of the impact of harm on health outcomes or costs. This second step referred to as the method to quantify harm, requires an assumption on consequences of each severity level, for example, what severity levels are considered to cause a hospital admission, or estimating costs associated with different severity levels. The method used to estimate the risk of harm from the medication error was reviewed separately from the method used to quantify harm or to generate health outcomes related to patient harm.

Figure 2.2: Flow diagram of potential pathways to quantify patient harm indirectly



Only three of 13 studies incorporated the increased likelihood of specific ADEs associated with the medication errors [Table 2.4] (150, 151, 227). The advantage of the availability of specific ADEs, such as the increased risks of asthma exacerbations in Forster et al. (2018) associated with the administration error, is that quality of life and costs can be attached to these ADEs very specifically. The three studies by Forster et al. (2018), Elliott et al. (2013) and Moriarty et al. (2019) used data from the literature on the increased likelihood of specific ADEs.

Table 2.4: Details on incorporation of medication errors and linkage with patient outcomes estimated from economic evaluations from literature review

Reference	Measurement of error	Source for harm	Method to obtain probability of harm	Method to quantify harm estimate	Measurement of cost of error	Measurement of effectiveness	Robustness of link between error and harm
Elliott 2013 (149)	Error rate	Literature	Increased likelihood of specific ADE	Not required	Healthcare resource use by health state from literature	Utilities for each health state	PSA, OSA
Forrester 2014 (218)	Error rate	Expert elicitation	Assessment of preventability of ADEs (Naranjo score)	Severity of pADEs assessed by experts (NCC MERP categories)	No cost for ADEs or harm	Incremental errors or ADEs	PSA
Forster 2018 (150)	Error rate	Literature	Increased likelihood of specific ADE	Not required	Healthcare resource use by health state from literature	QALY from Asthma Quality-of-Life Questionnaire scores	PSA, OSA
Ghatnekar 2013 (219)	ADE related hospital admission	Expert elicitation	Assessment of preventability of ADEs (WHO-UMC criteria)	Certain, probable and possible pADE are used as hospitalisation	Cost of additional hospitalisation for ADE from literature	Disutility of hospitalisation or outpatient contact	PSA
Karnon 2009 (156)	Error rate	Literature	Assessment of preventability of ADEs (preventable, not preventable, no harm expected) excluding intercepted errors	Weighted average of two published severity distribution of ADEs	Cost by ADE severity from literature	QALY decrement by ADE severity (hypothetical estimates)	PSA

Reference	Measurement of error	Source for harm	Method to obtain probability of harm	Method to quantify harm estimate	Measurement of cost of error	Measurement of effectiveness	Robustness of link between error and harm
Maviglia 2007 (220)	Potential ADE rate	Literature	Assessment of preventability of ADEs (preventable, not preventable, no harm expected) excluding intercepted errors	All pADE are used as hospitalisation	Cost of additional hospitalisation for ADE from literature	Cost of pADEs	PSA and relevant OSA (probability and cost of harm)
Moriarty 2019 (151)	Error rate	Literature	Increased likelihood of specific ADE	Not required	Healthcare resource use by health state from literature	Utilities for each health state	PSA, OSA
Nerich 2013 (221)	Error rate	Expert elicitation	Assessment of severity (Hatoum scale)	Two panels assessed probability of hospitalisation following medication error	Cost of additional hospitalisation from literature (length of stay and ward elicited by experts)	Benefit based on avoided cost and hospitalisation	PSA and relevant OSA (changing category definitions)
Nuckols 2015 (222)	Error rate	Literature	Assessment of preventability of ADEs (meta-analysis of 7 studies)	Distribution of severity of ADE from 4 studies (Folli scale or NCC MERP)	Cost by ADE severity from literature	QALY decrement by severity of ADE and age from literature and assumptions	PSA and relevant OSA
Rosselli 2014 (223)	Error rate	Literature	Severity categories developed from expert panel and probabilities of categories from different sources in the literature	Harm attached to severity categories based on NCC MERP categories	Cost by ADE severity from literature and adjusted by experts	None	PSA and relevant OSA (probability of harm)

Reference	Measurement of error	Source for harm	Method to obtain probability of harm	Method to quantify harm estimate	Measurement of cost of error	Measurement of effectiveness	Robustness of link between error and harm
Samp 2014 (224)	Error rate	Literature	Assessment of severity (temporary or permanent harm, prolonged hospitalization, harm that required life-sustaining intervention)	Severity levels transformed into categories [probability of hospitalisation with errors reaching the patient with permanent or temporary harm from pharmacist assessment and based on assumptions]	Cost of additional hospitalisation from literature and medication changes	None	PSA and relevant OSA (probability and costs of harm)
Westbrook 2015 (225)	Error rate	Expert elicitation	Assessment of severity by pharmacist (severity assessment code) and recording of intercepted errors	Moderate, major and serious medication errors were seen as potential ADE. Proportion of potential ADE resulting in actual ADE and weighted average of two published severity distribution of ADEs	Cost of ADE and excess length of stay by severity level from literature	Incremental ADEs	Relevant OSA (probability of ADE and severity distributions)
Yao 2012 (226)	pADE	Literature	Expert elicitation of potential reduction of pADEs by the intervention	Distribution of ADE severity based on level of impairment (two studies)	Cost by ADE severity from literature	QALY decrement by severity level from literature	No relevant OSA

ADE: adverse drug event; OSA: one-way sensitivity analysis; pADE: preventable ADE; PSA; probabilistic sensitivity analysis; QALY: quality-adjusted life-year

All other studies estimated the risk of harm either by assessing severity of medication errors or preventability of ADEs. Of 13 studies, four estimated the risk of harm within the study using expert elicitation (218, 219, 221, 225). Experts either elicited preventability and casual relation to the medication error of identified ADEs (218, 219) or assessed severity of the medication errors (221, 225). In Forrester et al. (2014), clinical data was reviewed in the six months after the medication error was detected and searched for admissions based on ICD10-codes that might have been medication related. (218). The Naranjo algorithm [2.4.1] was used to assess the probability that the drug contributed to the hospitalisation (228). The intervention by Gathnekar et al. (2013) consisted of two parts: (i) An admission-based review process of patient records to reduce re-admission due to medication error (ii) a medication report at discharge that aimed to reduce re-admissions due to medication error after discharge. Medication error rate was linked to patient harm by assessing the likelihood of events to be due to medication error from other studies of the same research group in the same area of Sweden. It was not clear if this was the same hospital. The first part of the intervention was evaluated by Hellstrom et al. (2011) (229). A team of clinicians and senior researchers estimated the causality of re-admissions within three months and medication error in the intervention using the WHO causality scale [2.4.1] (108). The effect of a discharge report on required medical care in hospitals from the second part was assessed in a study by Midlov et al. (2008) (230). Any primary and secondary care contact within three months of discharge was evaluated and the association with medication errors assessed using the WHO causality scale. For the pharmacist medication review, evaluated by Nerich et al. (2013), physicians assessed the severity of each medication error identified in terms of their clinical impact using a scale by Hatoum et al. (1998) (231). According to Hatoum's categories, medication errors were significant if stopping them either saved a patient's life or a major organ function was preserved. A not clinically significant medication error was judged to improve drug therapy to an acceptable level based on current guidelines or evidence standards.

Most studies based their risk of harm on studies from the literature unrelated to their intervention (149, 150, 156, 219, 220, 222-224, 226). Two of 13 studies used data derived from experts that estimated the potential harm of each medication error based on severity assessment (223, 224). Other models retrieved the potential risk of harm from studies that

assessed likelihood of ADE to be due to a medication error, hence preventable (219, 220, 222, 226). For example, Yao et al. (2007) applied results from a study that based the causality assessment on expert judgement without using an algorithm (232). Nuckols et al. (2015) used results from a meta-analysis of different studies linking errors with patient harm. The meta-analysis included six studies analysing the probability of preventable ADE from medication errors, excluding, if reported, medication errors with no or almost no potential for harm.

Methods to quantify the impact of harm on costs and health outcomes

In Figure 2.2, this step is shown in the lower part described as 'method to quantify harm'. It needs to be considered, which severities should be included and how consequences are quantified. Some studies estimated the probability of each severity level to lead to hospitalisation as a quantifiable health outcome related to harm (221, 224). After rating the severity of medication errors in Nerich et al. (2013), further clinical expertise of six specialist physicians was acquired in order to estimate probabilities for hospitalisation for each severity level. Rosselli et al. (2014), on the other hand, allocated costs to each of the levels of severity based on expert recommendations.

When preventable ADEs were linked with the medication errors, authors included them in different ways in the model. Forrester et al. (2014) used the number of ADEs itself as the effectiveness outcome and reported results in cost per preventable ADE avoided without accounting for costs generated by the ADE (218). Hospitalisation was used as the health outcome to quantify harm by considering all preventable ADE to be hospitalisations (220) or only ADEs with certain, probable and possible causality (219). The more common approach was to apply severity distributions of ADEs from the literature. Severity scales were used that are based on levels of impairment, such as the scale by Brennan et al. (233) in the economic evaluation by Yao et al. (2012) (226). Other severity categories applied in for example Nuckols et al. (2015) and Karnon et al. (2009) were based on the probability to cause harm, such as the Folli scale (234) or the American NCC MERP categories (235). The NCC MERP categories were developed to assess severity of medication errors, but they were also used to assess severity of preventable ADEs. The categories are grouped into

different levels of harm (death, harm, no harm). The three economic evaluations assessing severity of preventable ADEs all used combined estimates from multiple studies. Nuckols et al. (2009), for instance, calculated the distribution of severity of preventable ADEs from four different studies (19, 236-238) using either the severity criteria by Folli et al. (1087) or the US NCC MERP categories.

In summary, the two-step process of linking medication errors with harm by assessing the potential risk of harm and subsequently making assumptions on how to interpret this potential risk to generate health outcomes and costs, heavily relied on assumptions and expert opinions. No gold standard on how to interpret severity levels exists, and each study interpreted it differently. Compared with estimates on the increased likelihood of a specific ADE associated with a specific medication error, the assumptions underlying the two-step process introduce a lot of uncertainty.

Measurement of the cost of error

This section describes if and how studies incorporated the impact of medication errors on costs in their cost assessment. All but one study accounted for the cost from patient harm associated with medication errors (218). The studies that estimated the increased likelihood of specific ADEs associated with the error attached costs to the specific health states, such as costs of uncontrolled asthma (149-151). The other nine studies attached costs to severity levels of preventable ADEs (156, 222, 223, 225, 226) or attached costs to additional hospitalisations from preventable ADEs (219-221, 224). The latter followed some form of assumption of which preventable ADEs (e.g., probable and definitely preventable ADEs) or which severity levels (e.g., serious, fatal ADEs) result in hospitalisations.

Of these nine economic evaluations, four applied costs from a case-control study in the US from 1997 by Bates et al. (156, 220, 224, 225). Bates et al. (1997) provided an estimate of cost of preventable ADEs, used by Maviglia et al. (2007) and Samp et al. (2014), and cost of preventable ADEs by severity levels of ADEs, applied by Westbrook et al. (2015) and Karnon et al. (2009). The cost of preventable ADEs were not significantly different ($p > 0.05$) in Bates et al. (239). This introduces further uncertainty into probabilistic analyses and might

bias the results of deterministic analyses when only mean estimates are used. The study by Nuckols et al. (2015) (222) applied more recent estimates by ADE severity from a US study in 30 community hospitals (240) with ranges based on estimates from other studies. Only two of the nine studies used country specific cost estimates from the literature on cost of ADEs (220, 224).

Measure of effectiveness of interventions

The measurement of effectiveness in the economic evaluations was either incremental ADEs (218, 225), cost savings from prevented ADEs (220, 221), QALY decrements from ADEs (156, 219, 222, 226) or QALYs gained (149-151) from the intervention. None of the models measured non-health benefits as described in section 2.6. Of the four studies applying QALY decrements to severity levels of ADEs, two economic evaluations used hypothetical estimates from Karnon et al. (2009) (156, 222). Yao et al. (2012) estimated a utility by simulating 'typical EQ-5D states' for the different severity levels (226). The typical utility for the most severe ADE level, was for example based on guessed utilities for stroke. Gathnekar et al. (2013) interpreted preventable ADEs as hospitalisations and attached a weighted average of utilities based on utilities from the literature for the most common diagnoses for preventable ADEs identified in the study.

The studies simulating the increased likelihood of specific ADEs associated with specific medication errors attached QALYs from the literature that were associated with the specific ADE, such as quality of life with uncontrolled asthma (149-151).

Robustness of results to estimates around harm assessment

Eleven of 13 studies used probabilistic sensitivity analysis (PSA) to test robustness of the model results (149-151, 156, 218-224). This review was particular interested in the sensitivity of the results to estimates associated with assessing harm. Sensitivity to relevant individual parameters was investigated by six of the ten studies not directly measuring actual harm. In the one-way sensitivity analyses, the results were most sensitive to estimates around harm from medication errors, such as probability of harm from medication error and cost of preventable ADEs (220, 222, 224, 225). Changing the severity

categories in Nerich et al. (2013) almost decreased the benefit to cost ratio by half. Westbrook et al. (2015) reported that the results were more sensitive to the probability of harm parameters than to changes of the applied severity distributions. Only the economic evaluation from Colombia found staff costs to have a higher impact on the results than the estimates of probability of harm (223).

Conclusion from literature review of economic evaluations

The review showed that most research has been conducted in secondary care settings, and that primary care in the UK has been under researched as reported in other reviews (31, 146, 241). One of the key limitations of most of the identified models were the assumptions required to estimate the impact of error related harm on health outcomes and costs. In sensitivity analyses, uncertainty around harm estimates was found to have a substantial impact on the estimates of relative cost-effectiveness. Studies often relied on weak estimates from studies, where no confidence intervals, ranges or significance levels were presented, such as Bates et al. (1995) that has been used directly or indirectly in four of the 13 identified models (156, 220, 222, 225). A review by Nuckols et al. (2014) also highlighted the weak design of studies assessing the risk of preventable ADEs (242). Those studies that used expert elicitation within the study (218, 219, 221) and those applying estimates on the risk of harm from the literature used methods that were dependent on judgement by healthcare professionals (107, 127). The use of potential severity of HPEs elicited by experts does not account for the probability that errors might be intercepted before they reach the patient.

Errors were often grouped together as any medication error or by severity of error. The underlying assumption that an error always has the same probability of leading to harm, is a simplification for pragmatic reason that over simplifies the problem. If the intervention only reduces a specific set of medication errors, these might result in different risks of ADEs than errors prevented by a second intervention.

The second step, after the risk of ADEs is assessed, is the quantification of harm. To generate quantifiable estimates of harm, often depended on expert elicitation or vague

estimates from the literature. The more severity levels the model contained, the more challenging the data availability became. In Nuckols et al. (2015), for example, probabilities for severity levels were taken from four different studies (19, 236-238) because no study provided probabilities for all levels, which impedes reliability of the results.

While estimates on the risk of ADEs and to quantify harm already introduced substantial uncertainty, attaching costs and utilities proved difficult as well. The estimates were often derived from studies in different settings or countries, were not associated with the intervention under investigation, and were outdated. For example, the US study by Bates et al. (1997) that was widely used to estimate cost of preventable ADEs was applied in an Australian (225) and a UK setting (156). The study is over 20 years old and cost structures and payment systems have potentially changed. A more recent review of economic impact of ADEs and preventable ADEs found the cost of ADEs to vary considerably between healthcare setting, population, methodology and the publication date (30). QALY estimates, when used, were usually based on hypothetical estimates. This was probably due to the lack of availability of any such data in the literature (151, 232, 241). A review mandated by the European Commission supported these findings and criticised the suitability of studies used for cost and quality of life estimates (241).

In conclusion, the lack of robust data linking patient harm and medication errors, as well as on applicable costs identified in sections 2.4 and 2.5 impacted the quality of results from modelling studies. The only economic evaluations, where more precise input parameters were applied, focused on a specific subset of errors (149-151). However, these only cover a small number of medication errors. This underlines the need for further research on the impact of medication errors on health outcomes and costs.

2.9 The Safety Medication dASHboard (SMASH)

This dissertation explores a method to assess the cost-effectiveness of SMASH in reducing hazardous prescribing. In this section, the intervention itself is described, with a specific focus on the set of HPEs targeted by SMASH, and the choice of study design to evaluate the effectiveness of SMASH in Peek et al. (2020) is critiqued (48).

SMASH is an e-A&F intervention that aims to reduce hazardous prescribing in primary care (40). It is a system level DHI with three key components: (i) training of clinical pharmacists to deliver SMASH; (ii) a web-based dashboard providing actionable, patient-level feedback; and (iii) pharmacists reviewing individual at-risk patients and initiating remedial actions or advising GPs on doing so. SMASH is a local development of the previously described PINCER intervention [2.7], developed and tested in Salford. Compared with PINCER, SMASH uses continuous feedback instead of reports at set points in time. SMASH can be accessed at any time via a secure login and is updated every 24 hours. A clinical pharmacist supports the GP in responding to the identified patients and to find solutions similarly to the PINCER intervention, but the pharmacist visits the practices constantly in SMASH (243). Policy developments encouraging the employment of clinical pharmacist in general practices, supported the implementation of this continuous surveillance in SMASH (244). Salford Clinical Commissioning Group (CCG) was a forerunner to this and decided to fund up to 20 clinical pharmacists to work in general practice across the CCG. SMASH works with these pharmacists that are continuously linked with general practices. PINCER on the other hand, used a pharmacist outreach approach where they could spend up to 12 weeks with a practice.

SMASH contains a web-based electronic system that combines active aspects of an e-A&F interventions with aspects from an educational intervention. The educational component of the dashboard provides generic information on each of the types of HPEs and HMEs targeted by SMASH. The provided information educates the user on available evidence on the risks associated with the specific HPE types and recommendations how to resolve it. The e-A&F component of the dashboard actively reports on the patients currently exposed to HPEs and HMEs in the general practice. The electronic health records are screened for specific prescribing safety indicators that describe 12 different HPEs or HMEs. How these were developed is described in section 2.9.1. The dashboard has a user interface that provides a visual presentation of performance indicators, in this case the number of patients potentially exposed to HPEs. Examples of hypothetical aggregated patient lists are reported in Williams et al. (2018) (40). The patients exposed to an HPE are presented as affected patients (numerator), and the patients at risk of the HPE denoted as eligible patients (denominator). The patient list is linked via NHS number to the electronic health

records of the patients. The lists of patients for whom potential HPEs were identified are updated every 24 hours and accessible at any time via a remote log in through a secure server. In addition, the dashboard enabled users to track changes in number of patients potentially exposed to HPEs within the practice as well as in comparison to other practices. These changes were presented as numbers or in graphical form. This form of multimodal presentation of feedback, tracking of performance, as well as clear illustration of actions required is considered best practice when designing e-A&F interventions (245).

Practice-based pharmacists interpret the feedback from the dashboard by reviewing each patient identified with a potential HPE (246, 247). The review incorporated an assessment of the validity and relevance of the HPE for the specific patient based on the patient's history. The relevance for the patient is essential to select those cases where the hazardous prescription was correct and no alternative treatment options were available for this patient. If the pharmacist assessment found the HPE to be valid and relevant, they consulted with the GP regarding appropriate actions to solve the HPE. For some of the HMEs the pharmacist ordered the pending monitoring test without consulting the GP. Actions varied depending on the HPE or practice and could include the removal of a drug, an additional drug or a reminder for the patient to collect prescriptions (246). Actions were pursued by the GP or the pharmacist if the pharmacists were trained as independent prescribers.

In summary, SMASH is a complex, multifaceted intervention that combined elements from educational interventions (evidence summaries in dashboard), e-A&F (dashboard screening patient records) and pharmacist-led interventions (medication review). The educational aspect of SMASH was in the form of the HPE audit function within the dashboard and the provision of evidence summaries on the impact of the HPE and possible actions to resolve it. Kaur et al. (2009) mentioned the importance of reinforcing contact with the prescriber in educational interventions to enhance effectiveness, which is done in SMASH with the incorporation of the pharmacist service. To encourage social interactions when giving feedback was also identified as a key feature required for sustained effects of electronic feedback interventions (46). The e-A&F function of the intervention in form of the interactive dashboard supported pharmacist in identifying patients with HPEs from

electronic health records and provided performance reviews. A panel of clinical and policy experts judged these digital technologies to be most effective and cost-effective (169). The use of electronic health records in safety interventions was found to be promising in reducing the occurrence of ADEs (241). This potentially allowed pharmacists to review more patients compared to other pharmacist interventions (190-195, 197-199, 248). A common challenge with e-A&F interventions was alert fatigue (46, 249). Alert fatigue is defined as: 'Mental fatigue experienced by healthcare providers who encounter numerous alerts and reminders from the use of clinical decision support systems. As the numbers of alerts and reminders designed to provide meaningful assistance to the patient care process increases, many health personnel may ignore them' (250). SMASH only focused at a specific set of HPE types based on likelihood and seriousness to avoid alert fatigue, as it is best practice for e-A&F interventions (245). Similar approaches were found in the literature, where interventions focus on specific sets of HPEs, e.g., Beers (198, 199) or specific conditions. The approach to target specific burdensome ADEs on a clinical level, was considered most effective and less costly compared with national or organisational level interventions by clinical and policy experts (169). Overall, the development of SMASH was in line with best practice guidance on designing and implementing e-A&F interventions (46, 47, 245).

2.9.1 Development of a set of HPE types

This section starts with the rationale for using a specific defined set of HPE types followed by a description of how the set of HPEs targeted by SMASH was developed. The challenges with estimating HPE prevalence were summarised in section 2.2. Focusing on a specific set of HPEs with specific criteria, such as prescribing safety indicators, seems to be a valid alternative to solve the challenges around varying definitions, measurement methods, qualification of reviewers and inter-rater reliability. These specific criteria define HPEs that can be transformed into something computable in order to reduce variability and subjectivity in the identification process. The approach to use specific criteria has been widely used, particularly in older populations, using the Beers or STOPP criteria (93) as shown by the literature review by Assiri et al. (2018) that found that the majority of studies analysed PIMs using these criteria. A policy by the Short Life Working Group, commissioned

to inform the Department of Health and Social Care on measures for England to reduce medication errors, suggested to use a specific set of HPE types to identify HPEs and to track effectiveness of measures to reduce HPE prevalence (35).

Prescribing safety indicators are generally used to define HPEs, representing potential deviations from best practice. The challenge with defining a specific set of HPE types is to cover those HPEs that are relevant based on likelihood to occur and to cause harm (183). Over the past decade progress has been made in defining patient safety indicators. The Royal College of General Practitioners commissioned the development of prescribing safety indicators for a UK setting. Based on indicators developed by Morris et al. (2004) in the US (120), UK specific indicators were defined using a RAND appropriateness measure (251) to achieve a consensus among GPs (42). The PINCER trial (45) was the first intervention using the developed prescribing safety indicators to target HPEs in a pharmacist-led information technology intervention. This first set of prescribing safety indicators used in the original PINCER trial is referred to as PINCER 1 indicators [Table 2.5]. The PINCER trial assessed prevalence of the PINCER 1 indicators in 72 English general practices and showed a significant reduction of HPEs (OR 0.71, 95% CI 0.59 to 0.86) and HMEs (OR 0.56, 95% CI 0.44 to 0.70) after six months in practices using the complex intervention compared with simple feedback (45). The PINCER 1 intervention was accessible free of charge for all general practices in the NHS until July 2015 and can still be purchased from PRIMIS (252). A study by Spencer et al. (2014) updated these prescribing indicators based on other sets in the literature, using the RAND appropriateness measure and asking GPs to weigh the importance of different prescribing safety indicators (43). The PINCER 2 indicators were based on this new set and were used for pilot studies for the PINCER rollout in Rushcliffe CCG. Results from the pilot study were combined with results from a large epidemiological study on prevalence of the HPE types by Stocks et al. (2015) (56) to derive the final PINCER 3 indicators for the PINCER rollout in the East Midlands (253).

For SMASH, that was rolled out in Salford, Greater Manchester, specific additional HPE types related to chronic kidney disease and treatment monitoring indicators were requested by the CCGs at the development stage. The Salford Integrated Record (SIR), combining health records from primary and secondary care, used by the practices in Salford

does not provide data on mental health. As a result, PINCER 3 HPE types related to mental health were excluded from the SMASH set [Table 2.5]. A description of each prescribing safety indicator divided in to definitions on patients at risk and patients exposed by the HPE type is reported in Appendix B. Overall, the HPE types in SMASH and PINCER 3 are based on almost the same prescribing safety indicators. The set was developed based on extensive literature reviews and consensus processes (42, 43).

Table 2.5: Overview of prescribing safety indicators used in different interventions

Prescribing Safety Indicator	PINCER 1 (45)	PINCER 3 (253)	SMASH (40)
Patients aged ≥75 years who have been prescribed an ACE inhibitor or loop diuretic long-term who have not had a recorded check of their U+E in the previous 15 months	x		
Patients receiving amiodarone for ≥six months who have not had a THY within the previous six months	x		x
Patients receiving methotrexate for ≥three months who have not had a FBC and/or LFT within the previous three months	x		x
Patients receiving lithium for ≥three months who have not had a recorded check of their lithium concentrations in the previous three months	x		
Patients receiving warfarin for ≥three months who have not had a recorded INR check within the previous 12 weeks	x		
Prescription of an oral NSAID, without co-prescription of GPA, to a patient aged ≥65 years		x	x
Patients with a past medical history of peptic ulcer who have been prescribed a non-selective NSAID without a GPA	x	x	x
Prescription of an antiplatelet drug without co-prescription of a GPA, to a patient with a history of peptic ulcer or GI bleed		x	x
Prescription of OAC in combination with an oral NSAID		x	x
Prescription of OAC and an antiplatelet drug in combination without co-prescription of a GPA		x	x
Prescription of aspirin in combination with another antiplatelet drug without co-prescription of a GPA		x	x
Prescription of antipsychotics for >6 weeks in a patient aged ≥65 years with dementia but not psychosis		x	
Prescription of a CHC to a woman with a history of venous and/or atrial thrombosis	x		
Prescription of a long-acting beta-2 agonist inhaler to a patient with asthma (unresolved) who is not also prescribed an inhaled corticosteroid		x	x
Prescription of a non-selective beta-blocker to a patient with asthma (unresolved)	x	x	x
Prescription of an oral NSAID to a patient with a history of HF		x	x
Prescription of an oral NSAID to a patient with eGFR <45		x	x

FBC: full blood count; GPA: gastroprotective agent; HF: heart failure; INR: international normalized ratio; LFT: liver function test; NSAID: non-steroidal anti-inflammatory drugs; OAC: oral anticoagulant including warfarin and direct oral anticoagulants (DOACs) THY: thyroid function; U+E: urea (renal) and electrolytes; ICS: inhalative cortisone

2.9.2 SMASH effectiveness study in Salford

SMASH was evaluated using a quasi-experimental study design. This section aims to create an understanding of quasi-experimental designs in general, the strengths and limitations of specific study designs, and then focuses on why this was an appropriate design to estimate effectiveness of SMASH in reducing hazardous prescribing.

Risk of bias in quasi-experimental study designs in general

RCTs are the gold standard to evaluate interventions in healthcare and to establish causal relationships (3, 254). However, RCT designs are not always practical or possible (255). Quasi-experimental studies manifested their position as a valid alternative in cases were RCTs are not an option (7). By definition, quasi-experimental studies lack the randomisation process utilised in RCTs (255). Randomisation of the intervention reduces the risk of selection bias in assignment of treatment and it avoids imbalance of the intervention and comparator with regards to patient characteristics (256). Without randomisation, it cannot be guaranteed that the estimated effect of the intervention is not distorted by unmeasured confounders. The vulnerability to unmeasured confounding depends on the study design of the quasi-experimental study (13, 14, 257).

Quasi-experimental study designs can be divided into those without a concurrent comparator and those with a concurrent comparator. This counterfactual can be a similar local, regional, national or synthetic control were the intervention was not implemented (concurrent) or can be a historic control (subjects serve as own control) (13). In this section, an overview of quasi-experimental study designs using concurrent controls is provided with a brief summary of typical biases and different methods used to construct comparison groups. Subsequently, the use of historic controls is described focussing on potential bias associated with this design. After potential biases with and without concurrent controls are explained, common data analysis methods used in quasi-experimental studies are described. It is distinguished between design methods for constructing comparison groups and methods for data analysis as it has been elsewhere (256). Other studies compare methods for constructing comparison groups, for example, matching and synthetic

controls, and analysis methods, for example, Difference-in-Difference (DID) all as different evaluation methods (258).

Quasi-experimental studies with concurrent controls

Concurrent controls can be used in quasi-experimental studies if data are not only available for the intervention group but also for a comparator that did not implement the intervention. A concurrent control has the advantage that it minimises confounding due to unobserved time-variant confounding. These could be changes in the population over time or intercurrent events affecting the effectiveness outcome occurring around the same time as the intervention. The latter is referred to as history bias (14, 259). As a consequence of the non-randomised assignment of the intervention, observed and unobserved time-invariant confounding could still be present. Observed treatments effects can then be biased because of systematic differences between the comparator groups.

There are two main study designs with concurrent controls: those with pre-test (pre-intervention) and post-test (post-intervention) observations and those with only post-test observations. The aim is to select a comparator that is similar to the intervention group and is said to predict what would have happened to the intervention group if the interventions had not been implemented (256). With only post-test data available a suitable comparator can only be identified based on the observed baseline characteristics. Without any pre-test data, it is difficult to test if the comparator is appropriate and historic trends cannot be included in the assessment. The availability of pre-test observations from both strategies with and without the intervention enables to determine the comparability of the two groups beyond the baseline characteristics at initiation of the intervention (13). With no reasonable comparator, the observed effect can be a result of differences in the compared groups (13, 257). By comparing pre-test observations with regards to levels and trends of the effectiveness outcome, reasonable comparators can be identified.

Various methods are available that aim to reduce bias due to time-invariant observable confounding by identifying reasonable comparators with balanced characteristics. These methods are referred to here as methods to construct comparison groups. A common

method to construct comparison groups is by matching subjects with characteristics similar to those of the intervention group. The matching methods, such as propensity score matching (PSM), can be used to increase balance of observed baseline characteristics between the compared groups (256, 260). Even if observed baseline characteristics are balanced between the comparators, it cannot be guaranteed that unobserved characteristics are balanced too. The estimated effect could be explained away by differences in unobserved baseline characteristics (time-invariant confounder).

A form of matching that matches not only on characteristics at one point in time but on trends in outcomes in the pre-test period is the use of synthetic controls. Synthetic controls are increasingly used when evaluating policy interventions (261, 262). It is a data driven approach that aims to generate a comparator that mimics characteristics and outcomes of the intervention group before the intervention was implemented. The actual observed effectiveness outcomes can then be compared to the effectiveness outcomes predicted by the synthetic control in the post-test period. The synthetic controls are generated by linear programming from actual controls without the intervention. A weighted average of some of the potential control units is used to build the synthetic control group. The potential control units are matched based on pre-defined characteristics thought to impact the effectiveness outcome of interest from the pre-test period (261, 262). As with normal matching, synthetic controls only generate controls based on observable characteristics and cannot fully account for unobserved confounders.

Another method used to construct comparison groups is regression discontinuity (256). This option is possible when exposure or assignment of the intervention is bound to a defined threshold or cut-off. Assuming subjects or units just below and just above the threshold have similar characteristics, this method is thought to reduce the risk of time-invariant observed and unobserved confounding.

Quasi-experimental studies with historic controls

In longitudinal datasets with at least one pre-test observation, this pre-test observation can be used as a historic control. Each subject or unit serves as its own control. A historic control

minimises confounding due to time-invariant confounding (257). Because the intervention and comparator population are the same and observed and unobserved base line characteristics are thought to be distributed evenly when each subject serves as its own control. The use of historic controls, however, is susceptible to maturation effects and time-variant confounding. Maturation effects are described by secular trends over time that can be mistaken as a treatment effect (14). It can also not be ruled out that the observed effect is due to unobserved concurrent interventions or events occurring at the same time (history bias). Another problem in using historic controls is regression to the mean that can lead to wrong interpretation of an observed effect as a treatment effect that is due to chance (14). This is especially problematic in simple pre-test/post-test designs with pre-test being a period before the intervention and post-test after the intervention, such as cost before and cost after an intervention. Where multiple pre-test and post-test datapoints are available the risk of regression to the mean can be minimised (257).

Data analysis methods used in quasi-experimental study designs

Different evaluation methods are used in impact evaluation studies with concurrent and historic controls. While here matching methods and identification of controls were considered as design methods, the actual analysis of the data after the cohorts were identified is referred to in this section on data analysis methods. Table 2.6 compares key features and associated biases of different methods for data analysis.

The analysis method is driven by the availability of pre-test data and a reasonable control (13). Hinde et al. (2019) and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Research Practice Task Force report on quasi-experimental study designs by Berger et al. (2012) provide an informative overview of possible methods, their strengths and limitations.

In a simple pre-test/post-test or before and after study, the difference between observed outcomes before and after the intervention are calculated. As described before, this method cannot account for time-varying confounding, maturation effects, such as secular trends and seasonality, or regression to the mean. If a control group is available, observed

mean outcomes post-intervention can be compared. The comparator minimises the risk of time-variant confounders, but this method is susceptible to time-invariant confounding and it can be challenging to identify reasonable comparators (13).

Table 2.6: Comparison of data analysis methods in quasi-experimental study designs

Method of analysis	Concurrent control (yes, no)	Pre-test data available (yes, no; single, multiple observations)	Risk of confounding and potential bias	Can bias be mitigated?
Simple before and after comparison	No	Yes; single observation	Time-variant confounding, history bias[a], maturation bias[b], regression to the mean	No
Simple before and after comparison using average of historical data	No	Yes; multiple observations	Time-variant confounding, history bias[a], maturation bias[b]	No
Comparison of control and intervention outcomes	Yes	No; single observation	Time-invariant confounding, selection bias[c]	No
Interrupted time series analysis (ITSA)	No	Yes; multiple observations	Time-variant confounding, history bias[a]	No
Difference-in-Difference analysis (DID)	Yes	Yes; multiple observations	Time-invariant and variant unmeasured confounding affecting trend in outcomes, selection bias[c]	Yes
Controlled interrupted times series analysis	Yes	Yes; multiple observations	Time-variant confounding, selection bias[c]	Yes

[a]History bias, e.g., intercurrent events; [b]maturation bias, e.g., seasonality or inter-annual trends, secular trends in outcomes or population characteristics; [c]systematic differences between compared groups, unreasonable comparators

When multiple pre-test observations exist, time series techniques can be used. Interrupted time series analysis (ITSA) utilises the availability of pre-test data to account for potential pre-intervention trends. The advantage of this methods is that it can account for secular trends and seasonality and is not susceptible to common time-invariant confounding and regression to the mean [Table 2.6]. The most common ITSA method is segmented regression (263, 264) that can be used without a control available. More advanced ITSA methods use multistage regression modelling or when a regional, national or synthetic

control is available, controlled ITSA methods (7, 265). The controlled design allows to account for history bias if an appropriate control was identified (264). Hence, the time-invariant confounding can be mitigated.

When control area data are available, a frequently used statistical methods to analyse longitudinal data is DID (258). The core assumptions in DID is the assumption of parallel trends (266). As described earlier, with pre-test data available, trends can be used to assess if a comparator is reasonable. The underlying assumption is that the trend in the post-test-period is the same in intervention and control group. In DID the difference before and after the intervention starts are estimated for the intervention and control group. The treatment effect is then estimated by subtracting the before and after difference of the control group from the before and after difference in the treatment group. Consequently, DID can mitigate biases due to potential intercurrent events (history bias) that impact the observed effectiveness [Table 2.6] that is not possible in designs without concurrent controls (257). Time-invariant confounding is compared to simple comparison of control and intervention only a problem if it affects the trend in outcomes post-intervention. If trends pre-intervention are comparable and not distorted by differences in baseline characteristics, the risk of these differences to impact the post-intervention trend in outcomes is considered low.

In summary, various study designs and data analysis methods are available that each introduce different biases. A key challenge in designs with and without concurrent controls is unmeasured confounding (14, 267, 268). An understanding of the different types of bias relevant for the specific design is necessary to interpret the results correctly. Overall, controlled designs that account for trends in the pre-test period, e.g., DID and controlled ITSA, are found to be stronger than designs using historic controls alone (13, 257, 259, 264). The key strength is that they are more robust against history bias and time-variant confounding because of the controlled design (264). However, controlled designs only produce robust results, when the control group is similar to the intervention group or differences can be adjusted for in the analysis (selection bias). Depending on the data analysis method, a reasonable control can be a control with no systematic difference with regards to base line characteristics (256) or with a common trend and the same exposure

to events happening at the same time as the intervention (259, 266). When no reasonable concurrent control is available, and methods to construct the control group do not produce a control similar to the intervention group, the risk of confounding is high and ITSA designs are the next best alternative (269).

Observational studies as an alternative where RCTs are not an option

A common criticism of randomised trials is the lack of generalisability to real world settings and quasi-experimental designs allow to test the effect of the intervention in a real world population (7, 268). Other reasons why RCTs are not practical can be (i) the higher cost of randomised trials (7, 267, 270), (ii) the implications of limited numbers of eligible participants on the validity of the results (14), and (iii) the intention not to withhold a beneficial intervention from a random sample of practices (14, 270).

NICE acknowledged challenges with randomised trials and included quasi-experimental study designs in their recent evidence framework (12). The 'Evidence Standards Framework for Digital Health Technologies' published by NICE in 2019 stated clear rules for when quasi-experimental evidence is appropriate to use (12). Evidence standards were classified depending on the risk for the patient associated with the intervention defined by evidence tiers. For each evidence tier NICE provides study design options and minimal evidence standards that need to be fulfilled. With increasing risk of the intervention to harm the patient, the required evidence level increased proportionally. The evidence levels ranged from providing evidence of successful implementation of the intervention to providing evidence from high quality RCTs.

NICE provides a detailed list of criteria to stratify interventions into the appropriate evidence tier. Evidence Tier 2 interventions are defined as simply providing information or allowing simple monitoring or communication services. SMASH provided these services within the dashboard, but the intervention went beyond that. The involvement of pharmacists and the intention to achieve measurable reduction in HPE rate that might incorporate behaviour changes for the GP and the pharmacist differs substantially from the definition of a Tier 2 intervention. According to NICE, Tier 3a digital health technologies aim

to prevent and manage diseases, may be used alongside treatment and are likely to have measurable user benefits as the SMASH intervention does. For an intervention to also qualify for evidence Tier 3b, it needs to directly impact or provide treatment or diagnosis or 'guide treatment decisions'. SMASH aimed to reduce HPE rates (measurable benefit) by guiding decisions to resolve HPEs, change treatment and was used alongside normal treatment (40). Consequently, SMASH classifies as a Tier 3b digital health technology according to the classification system. The evidence standards framework recommends RCTs as the best practice standard for Tier 3b interventions. If this is not an option, NICE suggests high quality intervention studies (experimental or quasi-experimental) as the minimum evidence standard.

Salford CCG decided an RCT was not appropriate to implement and evaluate SMASH. Firstly, because of scarce resources, conducting an expensive and long randomised trial was not an option for SMASH (255). Secondly, SMASH required a specific data sharing infrastructure that was only available in Salford at the time. The limited number of practices in Salford restricted the sample size and could have limited the power of an RCT (267). Thirdly, the SMASH intervention was an adaption of the already successfully implemented PINCER trial that showed effectiveness (45) and cost-effectiveness (227) in general practices. CCGs in Salford considered random withholding of SMASH for some practices was not justifiable (14). Therefore, SMASH was implemented in a way that facilitated a quasi-experimental study design to estimate its effectiveness. Because no control area was available with a similar healthcare infrastructure as in Salford, controlled data analysis methods (controlled ITSA or DID) were not an option. The availability of multiple pre- and post-test observations allowed an ITSA design using each practice as its own historic control.

2.10 Conclusion

The problem of hazardous prescribing in England

The prevalence of HPEs in England seems considerably low with 4.0% (CI 95% 3.5% to 4.5%) of prescription items being identified as an HPE [2.3]. In relation to the volume of prescriptions in England, this was found to result into almost 45 million HPEs per year. This

imposes a considerable risk to the health of the patients and requires actions to minimise HPE rates in England (35).

The lack of quantifiable evidence on harm from hazardous prescribing

Not all of these 45 million HPEs actually cause harm to the patient. They could be of low severity and not impact patients health or be intercepted before they reach the patient (235). To estimate the true burden related to patient outcomes and costs of HPEs, information on how many HPEs reach the patient and cause harm is required. In section 2.4, the methodological limitations of the two most common methods to estimate harm from HPEs are highlighted. Instead of directly linking HPEs to harm, they either identify ADEs and assess if they are related to the HPE (causality) and if they were preventable (preventability), or they estimate the potential future harm by eliciting the severity of identified HPEs. Both methods rely on expert judgement and considerable staff resources to elicit causality and preventability, on the one hand, and potential severity, on the other hand, on a case by case basis. Both methods generate informative measures to get an understanding of the burden of harm from HPEs, but they do not provide a direct link between HPE and harm to the patient. Further research on the direct link between HPE and harm is needed (31).

In this dissertation this knowledge gap is addressed. The focus on specific HPE types enables an analysis of more specific outcomes directly linked with the HPEs. In Chapter Four, a case study of one of the HPE types targeted by SMASH demonstrates how harm from hazardous prescribing can be estimated. Large electronic health record datasets are used that allow a retrospective follow-up of patients exposed to a specific HPE type. From onset of the HPE, these patients can be followed based on their electronic health records in primary and secondary care and occurring ADEs can be detected. Instead of identifying all ADEs and assessing causality and preventability, only ADEs associated with the specific type of HPE are assessed. An assessment of any hospitalisation might have diluted the effect, especially when the population at risk of HPEs are considerably old and multiple other comorbidities could have caused the hospitalisation.

The role of SMASH in the context of patient safety interventions

A large number of interventions are reported in the literature aiming to reduce hazardous prescribing (200). Results on effectiveness of educational interventions, computerised alert systems, e-A&F and pharmacist-led interventions have been mixed [2.7]. There is need for future research to demonstrate not only a reduction of HPE rates but also the impact of interventions on patient outcomes, such as hospitalisations or mortality (187). Based on changes in HPE rates alone, conclusions on the impact on patient outcomes cannot be made (179, 186, 200, 203). Evidence on cost-effectiveness of interventions reducing HPEs was subject to a lot of uncertainty. The key factor driving uncertainty was the link between the medication error and harm [2.8]. This covered the probability of harm, the health outcomes associated with harm and the costs and utilities attached to these health outcomes. The lack of high quality evidence on resource use [2.5] associated with harm from HPEs, resulted in many economic evaluations not incorporating these at all.

A new type of multifaceted interventions that was first introduced with PINCER, combines aspects from pharmacist-led interventions and educational outreach with information technology interventions. The need for interventions combining different aspects of each intervention type was highlighted in different reviews on interventions aiming to reduce hazardous prescribing (179, 186, 203). The use of computerised screening tools to search electronic health records enabled a quick and objective measure of effectiveness of the study. By searching for explicit criteria that are defined by patient safety indicators, the HPE rates can be assessed automatically and the labour intensive and often subjective measurement of error rates as described in section 2.2 is not required. In the evaluation of PINCER, a new approach to investigate cost-effectiveness of interventions aiming to reduce HPE rates was used (149). One state-transition model was built for each HPE type to acknowledge the varying effects on patient harm, healthcare costs and quality of life of each individual HPE type. This approach was taken up by Moriarty et al. (2019) and Forrester et al. (2014) and was found to be more robust than economic evaluations not incorporating the increased risk of ADEs associated with HPEs [2.8]. The PINCER RCT has been evaluated and was found to be effective and cost-effective (45). Since then, the prescribing safety indicators have been developed further [2.9.1]. The local development of PINCER that is SMASH, the intervention under investigation in this thesis, contains an

updated set of prescribing safety indicators. SMASH already showed a reduction of these new HPE rates in the primary rollout in Salford (48), but the economic impact of these new HPE types still needs to be estimated.

In summary, SMASH represents a new type of intervention combining various aspects of previous interventions, in an area where not many economic evaluations were conducted. This dissertation aims to explore how to conduct an economic analysis of a quasi-experimental study that relies on routinely collected health data to measure exposure and outcomes using one particular type of HPE as an example.

Implications for thesis

Chapter Two identified different evidence gaps that were addressed in subsequent chapters: (i) the lack of quantifiable evidence on the increased risk of ADEs associated with HPEs, (ii) the lack of evidence on the economic impact of HPEs on patient outcomes and healthcare costs, and (iii) the lack of an economic evaluation of SMASH. This thesis addressed these evidence gaps by generating quantifiable evidence for one of the HPE types targeted by SMASH. In Chapter Four, the increased risk of specific ADEs associated with this HPE is assessed. In Chapter Five, the economic impact of this HPE on quality of life and healthcare costs is estimated. The last evidence gap on the economic evaluation of SMASH is addressed in Chapter Three and Chapter Six. While Chapter Three investigates the cost per HPE avoided, Chapter Six takes this analysis further by assessing the cost per QALY.

Chapter 3 - Cost-effectiveness analysis of SMASH

Chapter Three reports on a study that estimates the incremental cost per HPE avoided by the SMASH intervention compared with standard care. It entails detailed descriptions of the methods used to cost the intervention, as well as how the effectiveness estimates were derived. The total cost of the intervention is reported relative to the number of HPEs avoided, and the findings are discussed in the context of this thesis.

3.1 Introduction

System-level DHIs, such as e-A&F, are used increasingly with many new interventions being introduced in healthcare practice. The recent WHO guideline on DHIs highlights the need for countries to 'be guided by evidence to establish sustainable harmonized digital systems, not (to be) seduced by every new gadget' (271). An example of an e-A&F is the pharmacist-led information technology intervention SMASH. SMASH interrogates the electronic health record database of general practices and feeds back practice and patient-level information on potentially hazardous prescribing and inadequate monitoring to clinicians via a web-based dashboard (40). The aim of SMASH is to reduce the number of HPEs that are associated with an increased probability of harm for the patient (37, 113, 120).

The UK government recently made the reduction of HPEs, often referred to as medication errors, a policy objective (35) following recommendations from the WHO (33) and the EDQM (36). The policy encourages the implementation of interventions aiming to reduce HPEs. The effectiveness of SMASH has previously been demonstrated (48), but the cost of providing the intervention is not known to date. This chapter aims to (i) assess the cost of providing SMASH and (ii) combine the cost and effectiveness estimates to assess the cost-effectiveness of SMASH compared with standard care.

(i) Assess the cost of SMASH

In Chapter Two [2.9], the complex structure of SMASH that entailed an e-A&F component in form of the dashboard, the pharmacist services provided and an educational aspect was described. SMASH generates costs for the dashboard, pharmacist service, GP involvement, training and meeting costs. The different components of SMASH have their own specific challenges. DHIs are often implemented across a group of providers such that the cost is shared. How to allocate costs associated with the dashboard is therefore an important decision in the cost assessment. While the intervention is led by pharmacists, other staff is involved in the management of the HPEs, such as the GP, or in the support with the dashboard usage. Quantifying the involvement of each of these parties is challenging. As often with DHIs, there are differences in the delivery of the intervention between practices, for instance, the types of staff involved, methods of feedback and interactions among the involved staff, usage and acceptance of the intervention. In addition, these factors are likely to change over time in DHIs, when specific actions are taken over by other types of staff or processes become more efficient (272). Overall, there are various components of this complex intervention that need to be accounted for when assessing the cost of SMASH.

The cost of similar interventions targeting HPEs were reported by Elliott et al. (2014) for PINCER and Foy et al. (2020) for Action to Support Practices Implementing Research Evidence (ASPIRE) (149, 152). Another comparable intervention was the DQIP study in Scotland from 2012-2014 that planned to conduct a cost-effectiveness study according to the study protocol that has not been reported yet (51). The studies aimed to reduce the incidence of a similar set of HPEs. The studies were used to inform a costing framework and provided a helpful structure for cost components of such interventions. However, costs for these interventions were very specific to the interventions itself and could not be transferred to this study. PINCER did not entail a dashboard function that was always accessible as SMASH did and neither did ASPIRE. The DHI component was therefore a new aspect that needed to be included in this chapter.

While evidence on effectiveness of e-A&F exists as reported in a recent systematic review (273), evidence on cost-effectiveness was found to be scarce and/or of low quality in various systematic reviews (274-277). Unit costs, sources for estimates and calculations

were often not described sufficiently (274). Another review reported that key cost components, such as for training needs or personnel expenses, were not included (277). A more transparent reporting of costing studies is generally considered to be required for future CEAs (274, 276).

(ii) Combine cost and effectiveness estimates

One of the potential reasons for the low quality of existing economic evidence for DHIs is the lack of guidelines on how to conduct economic evaluations in DHIs (272). Where there is limited incentive or opportunity to implement clinical trials, evidence of effectiveness often relies on quasi-experimental methods (12, 272), such as it was the case for SMASH (48). NICE acknowledged this with the publication of the evidence standard framework described in Chapter Two [2.9.2] (12). While the evidence standard framework from NICE provides guidance on what minimal evidence standards should be fulfilled depending on the potential risk of harm for the patients involved in the interventions, there is no clear guidance on methodology of how to apply the new evidence standards or what impact the use of quasi-experimental studies has in the context of economic evaluations. Challenges that impact the interpretation of casual inference in different quasi-experimental designs were described in detail in Chapter Two [2.9.2]. Keeping in mind the limitations of quasi-experimental methods, this chapter explored how the effectiveness estimates can be used in the economic evaluation and what implications the specific quasi-experimental study design had.

Applying the new NICE evidence standards, the aim of this study was to assess the cost-effectiveness of SMASH as cost per HPE avoided. This adds to the identified gap of cost-effectiveness studies in e-A&F interventions. The chapter focuses on a transparent presentation of the cost of SMASH and the impact the use of effectiveness estimates from the quasi-experimental study had.

3.2 Methods

3.2.1 Design of economic evaluation

This chapter presents a cost-effectiveness analysis of SMASH compared with current practice over a time horizon of 12-months. The 12-months time horizon was chosen because effectiveness data was available up to 12 months for this study. Extrapolating the effectiveness of the intervention beyond the available data was not considered appropriate because of the chosen method of analysis that assumes a linear effect. The linear trend observed in the first 12 months after start of the intervention is not assumed to continue in the subsequent months (48). The economic evaluation followed guidelines from the Consolidated Health Economic Evaluation Report Standard (CHEERS) (278). The CHEERs checklist that reports where key criteria are addressed in this document is presented in Appendix C. Where guidance more specific to DHIs was required, the NICE 'evidence standards framework for digital health technologies' (12) was referred to.

3.2.1.1 Target Population

The target population of SMASH were patients at least 18 years of age in an average practice in Salford, Greater Manchester. The unit of analysis was each practice itself. The average number of patients at risk of HPEs in an average practice was estimated by dividing the total number of patients at risk (n= 47163) by the number of practices SMASH was implemented in (n=43). Patients at risk of a HPE fulfil the denominator requirements of at least one of the HPEs described in Appendix B. Patients can be at risk of multiple HPE types. The number of patients at risk of an HPE in a practice fluctuates over time. An average practice in Salford had 1097 patients at risk of an HPE based on the number of patients at risk at 12 months after intervention start.

3.2.1.2 Strategies compared

The intervention was a system level DHI where an e-A&F intervention was combined with a medication review by a pharmacist. SMASH was rolled out in 43 practices in Salford, Greater Manchester. The intervention is described in more detail in Chapter One [2.9]. Williams et al. (2018) and Peek et al. (2020) report further details of the rollout (40, 48).

Standard care was chosen as the alternative strategy and describes current practice. It was assumed that SMASH was an addition to standard care, and it did not substitute other activities in the GP practices to resolve HPEs, for instance, annual patient reviews by the GP. It was further assumed that current practice was constant and did not change in standard care or with implementing SMASH. Hence, the economic evaluation compared SMASH with standard care as it would have been without SMASH.

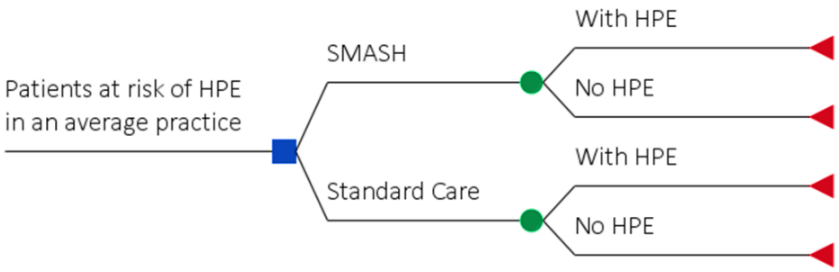
3.2.1.3 Primary outcomes

The primary effectiveness outcome was the number of HPEs avoided by SMASH. Costs were estimated from the NHS and Personal Social Services (NHS/PSS) perspective with regard to direct costs in providing SMASH in general practices. Costs were estimated in 2019 British pounds (£) and were not discounted because the time horizon did not exceed one year. No quantifiable estimates on resource use for services aiming to reduce HPEs before the intervention was implemented were available. There was no indication that these services stopped after SMASH was implemented. Due to lack of more specific information and quantifiable estimates, it was assumed that any service aiming to reduce HPEs in the comparator would also be available in practices with SMASH. Consequently, no additional costs were included for these services in the intervention or comparator. Cost per HPE avoided was used to express relative cost-effectiveness of SMASH versus standard care.

3.2.1.4 Structure of Decision-analytic model

A decision-analytic model (decision tree) followed the patients at risk of HPEs in an average practice in Salford. Two branches of the decision tree compared a practice using SMASH with standard care [Figure 3.1]. The full list of all HPEs reviewed by SMASH can be found in Appendix B. The CEA was conducted based on the relative effectiveness (difference in HPE rate) and mean cost of SMASH and standard care in order to generate cost per HPE avoided at practice level. The generation of cost per HPE avoided consisted of three parts: (i) estimating the relative effectiveness of the intervention compared with standard care [3.3.1], (ii) estimating the total cost of the intervention and control [3.3.2 and 3.3.3], and (iii) combining cost and effectiveness estimates to generate cost per HPE avoided by practice [3.3.4].

Figure 3.1: Decision tree describing the economic model used to assess cost-effectiveness of SMASH compared with standard care



3.2.2 Effectiveness of SMASH

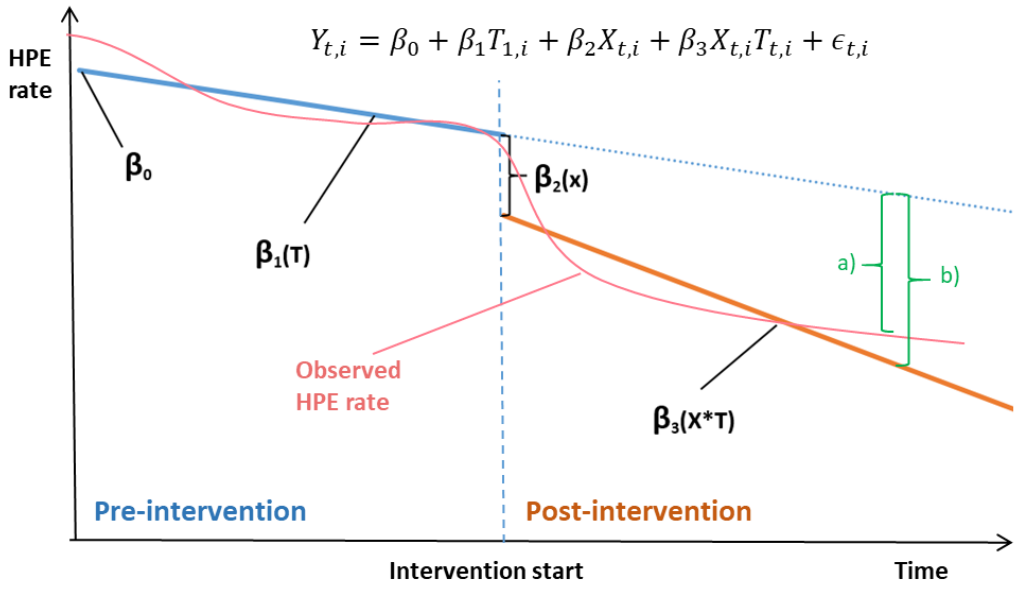
The effectiveness of SMASH was described by the reduction in HPE rates. HPE rate was defined as the proportion of actual HPEs (numerator) within patients at risk of HPEs (denominator) [Appendix B]. The denominator is not equal to the practice size but described by the number of patients at risk.

The effectiveness of SMASH was estimated outside of this thesis, and the method used was a practice-based ITSA using segmented regression and subsequent meta-analysis (48). The author of this thesis had no influence on the analysis. The outcome measures in Peek et al. (2020) were prevalence of exposure to any HPE (composite of ten types) and exposure to any inadequate blood-test monitoring (composite of two types). This chapter focuses on probabilities of HPEs. HPE rates were measured daily from 12 months before until twelve months after SMASH was initiated.

Because of the unique healthcare infrastructure in Salford and the rollout of the intervention in all eligible practices, there is no suitable contemporary comparator and the pre-intervention period was used as a historic comparator. Without a contemporary comparator, methods, such as DID or controlled ITSA, were not an option. Details about these methods are provided in Chapter Two [2.9.2]. Hinde et al. (2019) compared different methods to analyse time series datasets to inform decision making and suggested ITSA for analyses where a historical comparator is available (13). The advantage of ITSA over other methods, for example, simple before and after analysis, is that it allows accounting for the observed decreasing trend in HPE rates before the intervention and not just for the mean

before intervention start (7). The underlying regression model of the ITSA is visualised in Figure 3.2. No adjustments were made for transitional noise. It was assumed that the intervention effect will be observable immediately after the intervention starts because trainings and introductory meetings took place before the intervention start. In Peek et al (2020) or any related work no explorations were made to investigate learning or decay effects.

Figure 3.2: Graphical presentation of fictitious results of regression analysis from ITSA and their potential relation to HPE rates observed during the rollout



$Y_{t,i}$: Proportion of HPE in at risk patients (HPE rate) at time t for practice *i* [*i*=1, 2, …., 43]; β_0: intercept; β_1: pre-intervention slope, change per unit t; β_2: difference in HPE rate immediately after intervention implementation, level change; β_3: difference between pre-intervention and post-intervention slope; X: dummy variable for intervention; T: time since intervention start; a) difference based on pre-intervention trend and observed HPE rate; b) difference based on expected values from β_3(X*T)

The ITSA was performed on each practice separately to account for and quantify practice variation regarding the effectiveness of SMASH and the nested structure of the data. The HPE rate at six and 12 months after intervention start was estimated from the pre-intervention trend generated by the expected values for β1(T) at these time points [dotted line in Figure 3.2]. The extrapolated pre-intervention trend estimated the HPE rate in the comparator and is referred to as the counterfactual. The HPE rate for SMASH could be described by the ITSA results from the expected values of the regression or the observed values. Therefore, the absolute risk reduction could be described by two methods,

illustrated in Figure 3.2, as the difference in HPE rates between: '(a)' the observed HPE rate and standard care extrapolation (counterfactual), or by '(b)' the expected HPE rate from of the regression values denoted as $\beta_3(X*T)$ and the counterfactual. Depending on how well the predicted trend fits the observed data and the time point at which the effectiveness is assessed, results from the two methods can differ. A continuous linear decrease of the observed HPE rates over time would result in the same estimates for the absolute difference between observed and extrapolated HPE rates over time (a=b). This would be the ideal scenario where the regression models predict the HPE rate with SMASH precisely. In a scenario where the observed HPE rate reaches a floor effect, the absolute difference measured with method '(b)' increases over time (b>a). Because the HPE rate cannot decrease below 0 and a continuous decreasing trend is assumed for the hypothetical comparator, this does occur eventually.

The ITSA per practice was followed by a random effects meta-analysis. The random effects meta-analysis acknowledged practice heterogeneity that could have been derived from varying patient characteristics or the implementation of SMASH between practices. For SMASH, a bootstrap version of the DerSimonian-Laird method was carried out in order to assess this practice heterogeneity. In a re-analysis of almost 60000 meta-analyses identified from the Cochrane database, this method was found to be the most precise to identify between study variance and to estimate the effect size (279). Details about the methods for the ITSA and the meta-analysis used in Peek et al. (2020) that were not described in this paragraph or the publication can be found in Appendix D. A comparison of the ITSA with other data analysis methods is provided in Chapter Two [2.9.2].

The analysis team provided observed HPE rates with SMASH, the absolute difference between regressed estimates, regressed HPE rates with SMASH, HPE rates for the counterfactual for each HPE separately and for the composite of all HPEs. The number of HPEs avoided was calculated by multiplying the number of patients at risk of HPEs with the absolute difference in HPE rates at that time point between SMASH and the hypothetical comparator. Data was provided for the base case analysis at 12 months after intervention start and also for a six month scenario. The HPE rate in SMASH ($HPErate_{SMASH}$) was based on the observed results for SMASH and the denominator ($No\ of\ patients\ at\ risk$) based

on the observed number of patients at risk of the HPE at 12 months. The absolute difference from meta-analysed results of the ITSA was added to the observed HPE rates to generate the HPE rate for standard care ($HPErate_{Standard\ care}$). The absolute difference [scenario '(a)' in Figure 3.2] was adjusted for the pre-intervention trend and accounted for the heterogeneity between practices.

To get an understanding on how well the regression model predicts the observed values over time, the number of HPEs avoided in scenario '(a)' and '(b)' were compared. The number of HPEs avoided was generated by multiplying the absolute difference between SMASH and standard care for both scenarios with the denominator. The smaller the difference the more closely the regressed HPE rate for SMASH matched the observed values.

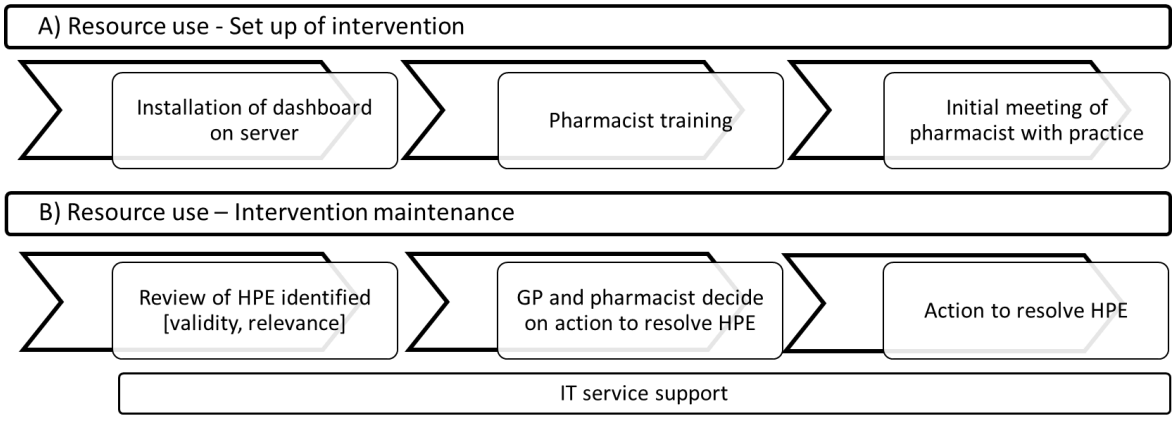
3.2.3 Cost of SMASH

The cost of SMASH was assessed retrospectively after SMASH had been rolled out across Salford, Greater Manchester. The unit of analysis was cost of SMASH in an average practice. Resources consumed implementing SMASH were assessed using a micro-costing approach, where each step of the service pathway was incorporated (280). For standard care, the cost was constant and, thus, no additional resource use was assumed to have been generated that was not also generated with current practice in SMASH. SMASH was seen as an addition to standard care. First, resource components that were necessary to deliver SMASH were identified based on a conceptualisation of the service pathway. Published qualitative (208, 243, 246, 247) and quantitative data relating to the delivery of SMASH (40) were used to identify a service pathway from which key cost components relevant for the resource use estimation could be derived. Subsequently, the amount of each resource component was estimated using micro-costing methods for each practice. For resource use of SMASH, a mixed bottom-up and top-down approach was chosen. Where activity or usage records were available on practice level, these were used to quantify resource use (bottom up costing). If only aggregated resource use across all practices was available, these were allocated between the practices enrolled in the study (top down costing) (281, 282).

3.2.3.1 Conceptualisation of resources

In order to identify relevant cost components of the intervention, the processes of deploying and using the intervention in practice needed to be fully understood (283). Resource use and cost were divided into set-up and maintenance costs. Development costs to design the dashboard were not considered in this analysis because they would not be relevant for decision-makers considering a wider rollout of SMASH (272, 284). Figure 3.3 illustrates the resource components that were identified to deliver SMASH from the conceptualisation of the service pathway.

Figure 3.3: Components of the service pathway relevant to estimate resource use of SMASH required for (A) set-up of the intervention and (B) intervention maintenance



Set-up costs entailed one time fixed costs required to initiate the intervention, such as resources for training needs or equipment requirements (283). The software was installed on the Salford Royal NHS Foundation Trust server. After installation, the dashboard was accessible for all practices via web log in (40). Each practice had one pharmacist advising on SMASH that received training to use SMASH. An initial practice meeting took place where the pharmacists introduced the dashboard to the practice and explained the dashboard (40). Cost related to server installations, pharmacist training and the initial practice meeting were identified as fixed set-up cost [Figure 3.3].

All recurrent costs were considered as maintenance costs, e.g., personnel, supplies, information and communication costs (283). Maintenance costs also included an IT service that covered (i) adding new users to the dashboard, (ii) responding to queries regarding

specific identified HPEs and (iii) fixing issues with the dashboard. The cost for IT service support were fixed costs. HPEs identified by the dashboard needed to be managed. This involved a review of identified HPEs, a GP contact and actions to resolve HPEs [Figure 3.3]. The management of HPEs was the only variable costs that changed with the volume of HPEs that were identified by the dashboard.

Based on SMASH design and details on the rollout of SMASH the following key cost components were identified for set-up costs: (i) server costs; (ii) training of pharmacists; (iii) initial staff meeting; and for maintenance costs: (i) management of HPE alerts by the dashboard; (ii) IT support [Figure 3.3].

3.2.3.2 Quantifying the resource use

Resources consumed for each cost component were estimated retrospectively by using data collected routinely during the rollout of SMASH (pharmacists' field notes) and expert elicitation (interviews with staff). Individuals were selected purposively for the interviews to obtain insight into three elements of resource use for SMASH: training of staff, delivery of SMASH, and IT infrastructure. A snowballing approach was used to identify subjects that could provide the information required. All interviews were conducted by the author of this thesis. Two pharmacists [participant 1 and 2] (involved in the delivery of SMASH) and a software engineer [participant 3] (who developed the dashboard, was involved in the implementation phase of SMASH and is responsible of the server maintenance and IT service) were interviewed. Where the required values on resource use were not available from the published papers, questions for experts to estimate these values were developed by the author of this thesis. This semi-structured schedule of questions was followed to guide the interviews [Appendix E]. Questions were asked around delivery of the intervention with the focus to assess cost for an average practice.

3.2.3.3 Resource use items of cost components

Each cost component identified in 3.2.3.1 entailed various resource use items and the methods how they were derived are reported in Table 3.1.

Table 3.1: Methods used to quantify resource use by cost component and details of the resource use items relevant to estimate

Resource use	Method	Resource use items	Source
Server costs			
Installing software	Total cost of installing the server was allocated between practices and distributed over the minimum lifetime of SMASH	Number of practices sharing costs, expected minimum lifetime of server	(40), assumption on expected lifetime
Pharmacist training			
Number of training events	Multiple trainings were provided with an average number of attendees. The number of trainings conducted to train all pharmacists was calculated by the number of pharmacists delivering SMASH divided by the average number of attendees per training. Total number of trainings was allocated between practices	Number of attendees per training, length of training, number of pharmacists requiring training, salary bands of trainer and trainees, expenses, Number of practices sharing cost	Interview [participant 1,2] (Williams et al. 2018)
Staff resources	The time trainees and the trainer spent per training were based on salary bands and training length		
Initial meeting			
Staff resources	The time attendees spent were based on salary bands and meeting length	Attendees of meeting, length of meeting, salary bands of meeting attendees, expenses	Interview [participant 1,2]
Management of HPEs			
Number of HPEs reviewed	Sum of all HPEs at baseline and new HPEs recorded over 12 months.	Number of HPEs reviewed, percentage of HPEs requiring patient contact, time pharmacists and GPS spent with managing HPEs requiring and not requiring patient contact	Dashboard records, interview [participant 1,2,3], field notes
Percentage of HPEs requiring patient contact	Mean of suggested percentage of participant 1, participant 2 and field noted written by pharmacists during intervention		
Time spent managing HPEs	Managing HPEs requiring no patient contact were managed by pharmacists. HPEs requiring patient contact were managed by pharmacists and GPs. The average times were reported in the interviews		
IT services			
Staff resources	Average time of IT services provided per week and salary bands of the provider were reported	Weekly support, salary band of provider of IT services	Interview [participant 3]

GPs: general practitioners; HPE: hazardous prescribing event; IT: information technology; SMASH: Safety Medication Dashboard

Server access costs included two servers on which SMASH was installed and any further support, for example, for backups and maintenance, that was required [participant 3]. Once the dashboard was installed on the servers, it was accessible even after the follow-up of the SMASH implementation study ended without any additional costs. Allocating the server costs to the 12-months time horizon only would overestimate the true cost. According to the tutorial on costing of eHealth interventions by Bergmo et al. (2015) this type of costs can be distributed over the 'expected lifetime' using an appropriate discount factor (280). The minimum lifetime was used as a conservative estimate for the expected lifetime in the base case. Minimum lifetime was defined as the years since start of SMASH up to the time of this study resulting in four years. The 12 months cost estimate is not discounted. Only for longer time horizons the subsequent yearly payments would be discounted.

Resource use for pharmacist training and the initial meetings were quantified by assessing time that staff spent with this activity as proposed by Bergmo et al. (2015). This required evidence of the type of staff and time investment of the trainer, the attendees and the number of events that took place. The resource use of training of pharmacists was estimated based on the total number of pharmacists working on SMASH (40) and the average duration and number of attendees for the training sessions [participant 1, 2]. The initial meeting incorporated resource use for the time the pharmacist spent at the meeting and the number and type of participants from the practice staff. Additionally, resources for room bookings or travel expenses for staff were included.

The quantity of resource use items for the cost for managing HPE constituted the number of HPEs reviewed, the time inputs for pharmacist and GPs per HPE and the percentage of HPEs reviewed that required patient contact. During the study period of the SMASH effectiveness study, the time staff spent managing HPEs was not recorded. Pharmacists reported that the time spent on reviewing HPEs varied considerably between practices and between weeks or months. As a result, an average time per practice was not possible to quantify. Pharmacists did recall time spent per HPE on average [participant 1, 2]. This was used as a proxy together with the number of HPEs reviewed. The average time per HPE was multiplied with the number of unique HPEs identified by the dashboard assuming each

unique HPE was reviewed once. The dashboard recorded all HPEs at any day during follow-up but not how many of these were actually reviewed by pharmacists. 'All HPEs' includes those that remained unresolved that re-occurred the following days. Assuming each patient with a HPE identified was reviewed exactly once, the HPE at baseline and on each following day only the new HPEs identified were counted.

Pharmacist and GP time spent per HPE in minutes was estimated based on experiences of the pharmacists [participant 1 and 2]. Pharmacists distinguished between HPEs requiring patient contact and those that did not. HPEs not requiring patient contact were solely dealt with by the pharmacist. When patient contact was required, an action to resolve the HPE was agreed on with the GP, and the patient was contacted. Consequently, HPEs requiring patient contact involved time of the pharmacist to review the HPE and time for the consultation of the pharmacist and the GP.

To assess the percentage of reviewed HPEs that required patient contact, expert elicitation [participant 1, 2] and field notes from pharmacists, written during SMASH follow-up, were analysed. During the rollout, pharmacists were encouraged to take field notes on each HPE they reviewed with key points on relevant patient history, relevance of HPE for this patient and suggested or taken actions. Each HPE was screened individually by a clinical pharmacist, the author of this thesis, and grouped to either not requiring patient contact or requiring patient contact. The percentage of HPEs that required patient contact was estimated as the mean of the values elicited from participant 1 and 2 and the results from the field notes.

Assumptions about the resource use on the basis of the interviews and/or the field notes were fed back to the healthcare professionals involved in SMASH for face validation. This entailed confirming that the assumptions and estimated were reasonable.

3.2.3.4 Unit costs

Unit costs were used from the most recent published reference costs at the time of analysis. Staff salary levels (GPs and pharmacists) were based on the NHS pay scales or

University of Manchester salaries. The unit cost from the University of Manchester pay scale are analogous to those that would occur in a wider roll-out of SMASH. The cost can be seen as generalisable because, e.g., in the current roll-out of SMASH to Greater Manchester the University of Manchester is providing the IT service. Where multiple hourly salaries were available per band, the midpoint of the band was used. NHS reference costs were taken from the 2019 Personal Social Services Research Unit (PSSRU) publication (285) as recommended by NICE (12). The unit costs are reported in the results section [3.3.2].

3.2.4 Incremental economic analysis

Effectiveness

The number of HPEs were generated from the HPE rate with SMASH (HPE_{SMASH}) or with standard care ($HPE_{Standard\ care}$) multiplied by the number of patients at risk. For the base case analysis, the HPE rate with SMASH was the observed proportion of patients with a HPE ($HPErate_{SMASH}$). The HPE rate in standard care ($HPErate_{Standard\ care}$) was calculated by adding the absolute difference generated in the ITSA and meta-analysis to the HPE rate with SMASH ($HPErate_{SMASH}$).

Costs

The cost components for set-up and maintenance costs were summed up to estimate the cost of delivering SMASH in the 12 months of follow-up ($Cost_{SMASH}$). The cost of standard care was assumed to be zero ($Cost_{Standard\ care}$).

Base case

An incremental analysis of the HPE rate and cost per practice in SMASH and standard care was performed. Incremental cost was calculated as the difference between the total cost of the intervention and the cost of standard care per practice. Incremental effectiveness was calculated as the number of HPEs avoided, generated by multiplying the absolute difference with the number of patients at risk. If SMASH was not dominant (incremental cost < 0; incremental effectiveness > 0) or dominated (incremental cost > 0; incremental

effectiveness < 0), an incremental cost-effectiveness ratio (ICER) was calculated (4). An ICER describes the additional cost of each unit of effect denoted as

Equation 1: $$ICER = \frac{cost_{interventionA} - cost_{interventionB}}{effectiveness_{interventionA} - effectiveness_{interventionB}}.$$

For SMASH, the cost per HPE avoided (ICER) was measured as described in Equation 2.

Equation 2: $$ICER = \frac{Cost_{SMASH} - Cost_{Standard\ care}}{HPE_{SMASH} - HPE_{Standard\ care}}.$$

Base case analysis

A probabilistic analysis was conducted using *TreeAge Pro Healthcare 2021* to account for the effect of parameter uncertainty on cost per HPE avoided (286). A random seed of 345 was used. Uncertain input parameters were characterised by probability distributions. For probabilities and utilities beta distributions and for costs gamma distributions were applied when necessary as suggested by the ISPOR taskforce report in 2012 (287). For the absolute difference, a normal distribution was chosen because the absolute difference can take on values above and below zero. To identify a distribution around the number of HPEs reviewed where individual data was available, multiple distributions were fitted and the best fit was identified following recommendations by Delignette-Muller et al. 2009 (288) using the R *fitdistrplus* command. Results for 10000 iterations sampled from uncertainty around the input parameters were plotted in a cost-effectiveness plane. Cost-effectiveness planes can be used to visualise the incremental costs and effectiveness pairs graphically (289).

3.2.5 Sensitivity analysis

Uncertainty associated with the analysis output can be investigated in deterministic or probabilistic sensitivity analysis (PSA). In PSA, different scenarios were tested changing assumptions compared with the base case analysis. In a deterministic one-way sensitivity analysis (OSA), specific input parameters are changed while all other input parameters stay constant. This allowed the analysis of the impact of specific input parameters or

assumptions on the results from the base case. Results of the OSA were presented graphically in a tornado diagram where each horizontal bar represents the changes in the ICER that resulted from varying the specific input parameter in the specified range.

3.2.5.1 Scenario analysis

Probabilistic sensitivity analyses were conducted to test robustness of assumptions made in the economic model or cost assessment. Details of these scenario analyses are summarised in Table 3.2 and compared with the base case analysis.

Impact of costs included

Economic evaluations of comparable interventions, for example, the study by Risor et al. (2017), only costed resources for maintenance and excluded set-up costs (290). In order to compare the results, scenario analysis 1 was conducted that did not incorporate set-up costs.

Impact of change in effect size over time

The rollout of SMASH showed a reduction of HPEs over time (48). The decreasing effect got smaller over time and HPE rates seemed to reach a ceiling effect after six months where no further reduction in HPEs was observed. To estimate cost of delivering SMASH for six months, the set-up costs stayed the same, but the maintenance costs were adjusted to six months. The number of HPEs avoided was calculated using the HPE rate at six months for SMASH and standard care using the same number of patients at risk as the base case analysis.

Table 3.2: Summary of parameters and assumptions in the probabilistic scenario analyses compared with the base case

Parameter	Base case	Scenario 1	Scenario 2
Time horizon	12 months	12 months	6 months
Cost components	Set-up and maintenance	Maintenance	Set-up and maintenance

3.2.5.2 Deterministic sensitivity analysis

Deterministic OSA were conducted to test robustness of assumptions made in the economic model or cost assessment (model uncertainty) and for uncertainties around the input parameters (parameter uncertainty). The OSA investigated how further changes in single parameters affected estimates of cost per HPE avoided. Parameters were selected for the OSA if there were alternative assumptions possible or where there was uncertainty around the input parameters. The ranges of parameters applied in the OSA are reported in Table 3.3.

Cost parameters

Server costs were based on the assumption that the lifetime of the server, would be the minimal observed time the server was running (model uncertainty). However, according to the IT expert [participant 3] working with the server, the server was assumed to run for much longer than four years. The effect of allocating the one-time server costs to one (minimum expected lifetime) and eight years (maximum expected lifetime) was analysed to test different scenarios on how long the intervention could be used.

The costs of managing HPEs depended on the number of HPEs reviewed and the percentage of HPEs that required patient contact. In the base case analysis, the uncertainty around the two estimates was accounted for by applying a probability distribution. The OSA was used to estimate the impact of them separately.

The base case assumed that 36 pharmacists were trained to use SMASH. Sensitivity analysis assumed that 29 (lower value) and 43 (upper value, equivalent to one pharmacist per practice) were trained. The number of practices between costs were shared was adjusted (20 and 600 practices). The maximum of 600 was chosen to represent the number of practices for a discussed rollout of SMASH to 600 more practices in Greater Manchester.

Outcome parameters

Uncertainty around costs of managing HPEs derived from the range in percentage of HPEs requiring patient contact and the number of HPEs (parameter uncertainty). Standard deviations of these two parameters were used in OSA. The uncertainty of the absolute difference in the HPE rate in SMASH and in standard care was applied in the OSA as the lower and upper bound of the 95% CI resulting from the ITSA and subsequent meta-analysis (parameter uncertainty).

Table 3.3: Parameter ranges applied in one-way sensitivity analysis

Parameter	Mean	Low range	High range	Assumption/source of range
Expected lifetime of server, years	4	1	8	Allocation of server costs over the expected lifetime
Pharmacists delivering SMASH	36	29	43	Number of trainings required to train all pharmacists
% of HPEs requiring patient contact	43.04%	22.75%	63.32%	SD (between pharmacist views and field notes)
Number of HPEs reviewed per practice	105.77	29.69	181.84	SD (between practices)
Absolute difference in HPE rates	0.96%	0.79%	1.12%	Confidence interval of meta-analysed absolute difference

HPE: hazardous prescribing event; SC: standard care; SD: standard deviation; SMASH: Safety Medication Dashboard

3.2.6 Ethical considerations

The University of Manchester Ethics Decision Tool (291) and the NHS Medical Research Council decision tool (292) were used to establish whether a formal ethical review of the interviews was required as data were collected from human participants. The Ethics Decision Tool confirmed that a formal ethical review was not required because data were collected about individuals acting in their own professional capacity. Further, the participants were not from a vulnerable group, were not at-risk of disclosing unprofessional conduct, and personal or sensitive data were not collected.

3.3 Results

3.3.1 Effectiveness of SMASH

The relative HPE reduction by SMASH was 50.68% (absolute reduction of 0.96%, 95% CI 0.79% to 1.12%) resulting in 10.5 (95% CI 8.6 to 12.3) HPEs avoided by SMASH per practice with 1097 patients at risk of a HPE [Table 3.4]. The authors of Peek et al (2020) report a substantial variation between practices [I^2 = 91.1%]. Table 3.4 reports the number of HPEs avoided for scenario (a), the difference between observed HPEs; or (b) based on ITSA result on predicted HPE rate for the effectiveness of SMASH compared with the counterfactual representing standard care.

Table 3.4: HPE rates with SMASH and standard care (Peek et al., 2020) (48)

Time point	HPE rate, in % with 95% CI		Absolute difference,	HPE avoided[a]
	SMASH	Standard care[c]	in % with 95% CI	per practice
Expected ITSA values, scenario (b)				
At 12 months	1.40 (0.95; 1.91) [b]	2.36 (2.07; 2.70)	**0.96 (0.79; 1.12)**	10.5 (8.6; 12.3)
At 6 months	1.77 (1.37; 2.22) [b]	2.46 (2.17; 2.79)	0.69 (0.57; 0.80)	7.5 (6.2; 8.8)
Observed values, scenario (a)				
At 12 months	**1.60 (1.49; 1.72)** [d]	2.36 (2.07; 2.70)	0.76 (0.58; 0.98)	8.4 (6.3; 10.7)
At 6 months	1.83 (1.71; 1.95) [d]	2.46 (2.17; 2.79)	0.63 (0.46; 0.84)	6.9 (5.1; 9.2)

[a]Number of HPE avoided generated by multiplying the number of patients at risk with the absolute difference in HPE rates between SMASH and standard care; [b]expected values from ITSA regressions of HPE rate for SMASH; [c]HPE rate extrapolated from pre-intervention trend; [d]observed values of HPE rates with SMASH; CI: confidence interval; HPE: hazardous prescribing event; ITSA: interrupted time series analysis; SMASH: Safety Medication Dashboard; in bold are estimates used in the base case analysis of the economic evaluation of SMASH vs. standard care

The number of HPEs avoided with SMASH was sensitive to the method used to estimate effectiveness. The estimated number of HPE avoided using the two methods and potential differences were graphically presented in Figure 3.2 as 'a)' and 'b)' and are reported in absolute numbers in Table 3.4 as scenario (a) and scenario (b).

At 12 months, the number of HPEs avoided was 25% larger when the ITSA results were used compared to when the observed values were used (10.5 vs. 8.4). At six months, the number of HPE avoided was 9% larger with ITSA results alone (7.5 vs. 6.9). While the expected

values regressed in the ITSA constantly decreased, the observed values decreased more rapidly, with a steeper slope in the beginning compared with the regression and showed a floor effect (48) as indicated by the hypothetical illustration in Figure 3.2 where the regression slope was still negative.

In the economic evaluation the effect of the intervention is described by the absolute difference as in scenario (b) that is applied to the observed HPE rate with SMASH [estimates presented in bold in Table 3.4]. Table 3.5 reports the final effectiveness estimates used in the decision-analytic model.

3.3.2 Resource use and unit costs

Resource use items relevant to quantify resource use for each cost component and relevant unit costs are reported in Table 3.5.

Table 3.5: Summary of input parameters (resource use, unit costs, effectiveness) with the distribution used in the probabilistic analysis

Resource use	Expected value	Source	Distribution
Server costs			
Installation on server	1	(40)	Fixed
No of practices sharing costs	43	(40)	Fixed
Expected lifetime of server	4 years	Assumption (current lifetime since SMASH-start 2016)	Fixed
Pharmacist training			
No of attendees per training	2.5	Interview [participant 1,2]	Fixed
Length of training	2 hours	Interview [participant 1,2]	Fixed
No. of pharmacists trained	36 pharmacists	(40)	Fixed
Salary band of trainer	NHS pay scale: band 8a	Interview [participant 1]	Fixed
Salary bands of attendees	NHS pay scale: 7 with band 8a and 29 with 7	Interview [participant 1]	Fixed
Room, travel expenses	0	Interview [participant 1,2]	Fixed
Initial meeting			
Attendee from practice staff	GP or practice manager	Interview [participant 1,2]	Fixed
Length of meeting	0.17 hours	Interview [participant 1,2]	Fixed
Salary band of pharmacist	NHS pay scale: 7 (band 8a), 29 (band 7)	Interview [participant 1,2]	Fixed

Salary band practice manager	NHS pay scale: band 5 or 6	(293)	Fixed
Room, travel expenses	0	Interview [participant 1,2]	
Management of HPE			
No. of HPEs reviewed, 12 months	18 (SD ±14)	Dashboard records, interview [participant 3]	Gamma
Percentage of HPE requiring patient contact	46% (SD ±20)	Interview [participant 1,2], field notes	Beta
Time pharmacist spend reviewing HPE requiring patient contact	0.33 hours	Interview [participant 1,2]	Fixed
Time GP spent with HPE requiring patient contact	0.08 hours	Interview [participant 1,2]	Fixed
Time pharmacists spend with HPE not requiring patient contact	0.17 hours	Interview [participant 1,2]	Fixed
IT services			
Weekly support in first 3 months	3.75 hours	Interview [participant 3]	Fixed
Weekly support after 3 months	1.88 hours	Interview [participant 3]	Fixed
Salary band of provider of IT services	University of Manchester staff: band 6	Interview [participant 3]	Fixed

Unit costs	Expected value	Source	Distribution
Server costs			
Installing software on the servers	£14082	Interview [participant 3]	Fixed
Accessing software in practices	£0	Interview [participant 3]	Fixed
NHS pay scale for scientific and professional staff:		PSSRU 2019	Fixed
Band 8a	£63 per hour		
Band 7	£53 per hour		
Band 6	£44 per hour		
Band 5	£34 per hour		
General practitioner	£110 per hour		
University of Manchester pay scale			
Band 6	£21 per hour[a]	Directorate of Human Resources (2019)	Fixed

Effectiveness	Expected value	Source	Distribution
HPE rate SMASH (observed)	1.60% (95% CI 1.49% to 1.72%)	(48)	Beta
Absolute difference in HPE rates (ITSA)	0.96% (95% CI 0.79% to 1.12%)	(48)	Normal
Patients at risk of HPE	47163	(48)	Fixed

[a]Midpoint of hourly salary in this band; CI: confidence interval; HPE: hazardous prescribing event; PSSRU: Personal Social Services Research Unit; SD: standard deviation; SMASH: Safety Medication Dashboard

The one-time cost for installing the dashboard on the servers was £14082 [participant 3]. The costs were allocated to each practice equally and distributed over four years (the minimum lifetime of SMASH). There were no further costs associated with the server. The total server cost per practice was £81.87 [Table 3.6].

The number of pharmacists trained to use the dashboard was 36 (40). No costs were generated for room bookings or expenses due to the close proximity of the practices to each other [participant 1 and 2]. The cost for training the 36 pharmacists that ran the intervention was shared between the 43 enrolled practices. Based on the length and number of trainings required, each practice got allocated 0.67 hours of trainer time and 1.67 hours of the attending pharmacists' time resulting in £137.53 per practice [Table 3.6].

Table 3.6: Quantifying resource use for key cost components clustered by set-up and maintenance costs on practice level

Cost component	Resource use item	Quantity of resource use		Unit cost, £	Cost, £
Set-up costs					
Server costs	Server cost all practices	0.02	per practice	3520.50 [c]	81.87
Pharmacist training	Trainer (pharmacist band 8a)	0.67	hours	65.00 [a]	43.53
	Attendees (pharmacist band 7/8a)	1.67	hours	56.14 [a]	94.00
	Total pharmacist training				137.53
Initial meeting	SMASH expert (pharmacist band 7/8a)	0.17	hours	56.14 [a]	9.36
	Staff attendee (practice manager)	0.08	hours	39.50 [a]	3.29
	Staff attendee (key GP)	0.08	hours	112 [a]	9.33
	Total initial meeting				21.98
Maintenance costs					
Managing HPE	Pharmacist	6.33	hours	54.94 [a]	1415.51
	GPs	0.95	hours	110 [a]	424.84
	Total managing HPE				1840.35
IT service	Researcher (university staff)	2.44	hours	20.89 [b]	51.06

[a]PSSRU (2019) (285); [b]University of Manchester Pay Scale 2019, [c]interview with participant 3

According to participants 1 and 2, the initial meeting on average took ten minutes and was attended by the pharmacist, as the expert introducing SMASH, and either a key GP or the

practice manager. It was assumed that 50% of initial meetings were with the key GP and 50% with the practice manager. As a result, ten minutes (0.17 hours) of pharmacist time per practice were required and the ten minutes of the attending staff member were split to 0.08 hours of either the key GP or the practice manager. Salaries of practice managers were assumed to be the mean of the hourly salary of band 5 and band 6 (293).

The pharmacists agreed that per HPE requiring patient contact 20 minutes (0.33 hours) were spent reviewing the HPE, contacting the GP and contacting the patient [participant 1 and 2]. For a HPE that did not require patient contact, the review required ten minutes (0.17 hours) on average. The contact with the GP required on average five minutes (0.08 hours). The field notes contained details on 358 HPEs of which 60.89% (218 out of 358) did not require patient contact. Estimates from participants 1 and 2 were 25% and 65% of HPE requiring patient contact, respectively. The mean of all three estimates was 43.04% with a standard deviation of 20.29%. After multiplying the number of HPEs and the staff time per HPE for HPEs requiring patient contact and those, not requiring patient contact, total costs of managing HPEs resulted in £1840.35 per practice [Table 3.6].

From participant 3, the resource use for IT services was elicited to be 0.1 and 0.05 full time equivalents of a band six researcher at the University of Manchester for hypothetical 50 practices in the first three months and the rest of follow-up, respectively. Assuming 37.5 hours per week as full time, the researcher spent 3.75 hours per week in the first three months and 1.88 hours per week in follow-up. In total, IT services required 122.21 hours of a researcher's time at 12 months for 50 practices and consequently 2.44 hours for one practice. Based on the mean salary of a band six researcher, this led to £51.06 per practice for 12 months of IT support [Table 3.6].

3.3.3 Cost of SMASH

The total cost of SMASH was £2133 per practice after 12 months (deterministic analysis). The costs of each resource component are reported in Table 3.7.

Table 3.7: Total cost of SMASH by cost component (deterministic analysis)

Cost item	Cost at 12 months, £	Cost at 6 months, £
Set-up costs		
Server costs	82	82
Pharmacist training	138	138
Initial meeting	22	22
Total set-up	241	241
Maintenance costs		
Managing HPE	1840	1242
IT service	51	27
Total maintenance	1891	1269
Total cost	2133	1510

HPE: hazardous prescribing event; IT: information technology; SMASH: Safety Medication Dashboard

3.3.4 Incremental economic analysis

In the base case analysis, the incremental costs were £2149 (2.5% to 97.5% credible interval £487 to £5790) at practice-level and the number of HPEs was reduced by 10.53 (2.5% to 97.5% credible interval 8.80 to 12.25) compared with a standard care practice [Table 3.8]. The expected incremental cost per HPE avoided for SMASH compared with standard care was £205 (2.5% to 97.5% credible interval £46 to £559). The deterministic analysis gave very similar results.

Table 3.8: Results of the base case and scenario analyses

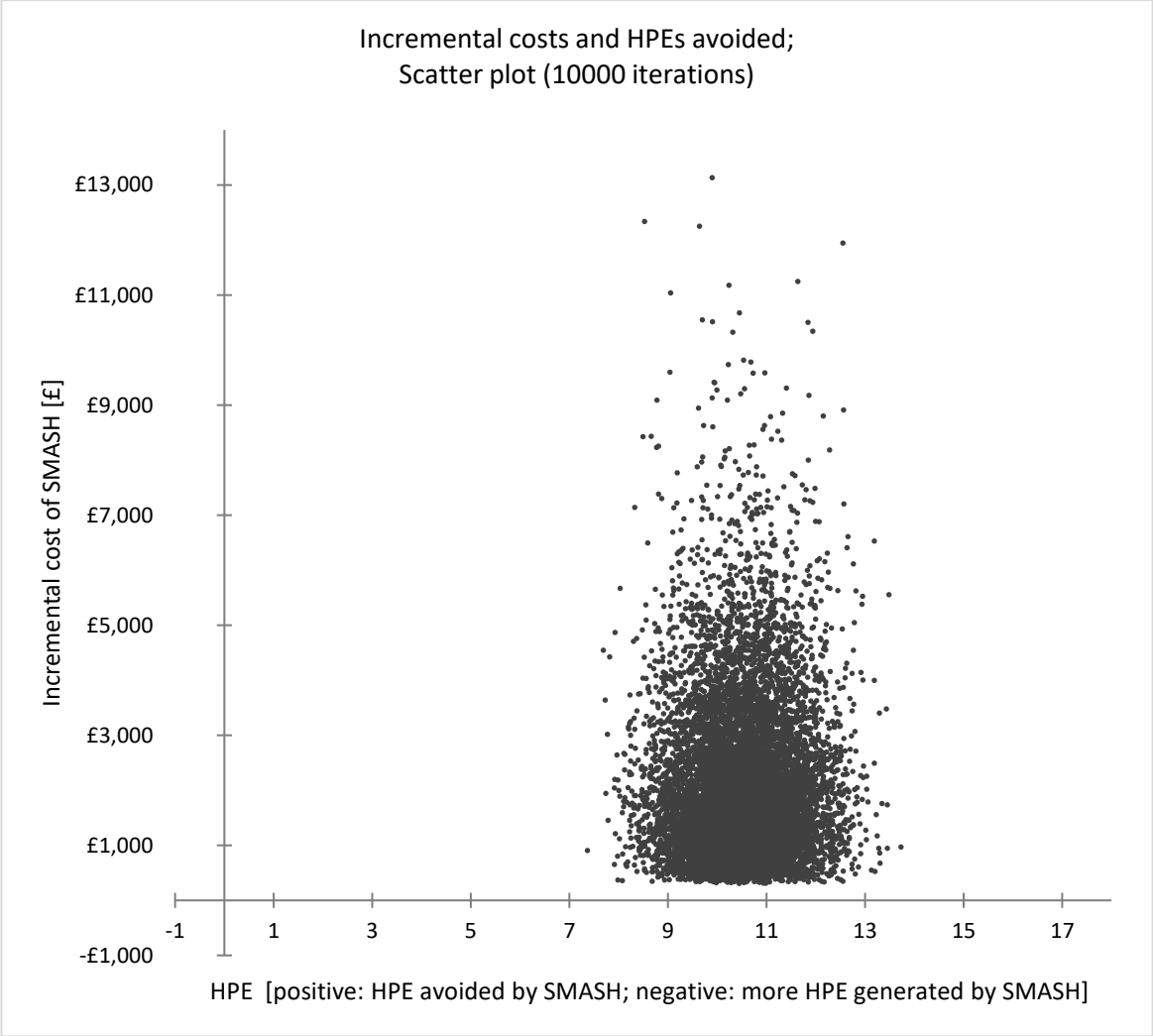
Scenario definition	Incremental costs, £ [b]	HPEs avoided [a,b]	Cost per HPE avoided, £[b]
Base case (probabilistic analysis)			
Total cost, 12 months	2149 (487; 5790)	10.53 (8.80; 12.25)	205 (46; 560)
Deterministic analysis			
Total cost, 12 months	2133	10.53	203
Scenario 1 (probabilistic)			
Only maintenance costs, 12 months	1907 (246; 5548)	10.53 (8.80; 12.25)	182 (23; 535)
Scenario 2 (probabilistic)			
Total cost, 6 months	1509 (377; 4179)	7.56 (6.30; 8.82)	201 (50; 561)

[a]Incremental effectiveness; [b]credible interval: 2.5% to 97.5% percentile; HPE: hazardous prescribing event; ITSA: interrupted time series analysis; SMASH: Safety Medication Dashboard

In the cost-effectiveness plane incremental costs and HPEs avoided from 10000 iterations sampled from parameter uncertainty were plotted [Figure 3.4]. A cost-effectiveness plane

is divided in four quadrants and the incremental effects are plotted on the x-axis and the incremental costs on the y-axis (294). Data points in the north-east quadrant represent interventions that are more costly and generate an additional health gain compared with the comparator. Dominant interventions that are more effective and less costly would generate results in the south-east quadrant. All data points left of the y-axis describe simulations where the intervention was less effective than the comparator.

Figure 3.4: Cost-effectiveness plane for SMASH vs. standard care (base case)



All observations were in the north-east quadrant, which indicated that SMASH is likely to be more effective but also more costly than standard care. The incremental costs were estimated to have a minimum value of £292 in the PSA. This was visible as a horizontal floor to the joint distribution of costs and HPE avoided in Figure 3.4. This was because all set-up costs and the IT service costs were fixed in the probabilistic analysis.

3.3.5 Sensitivity analysis

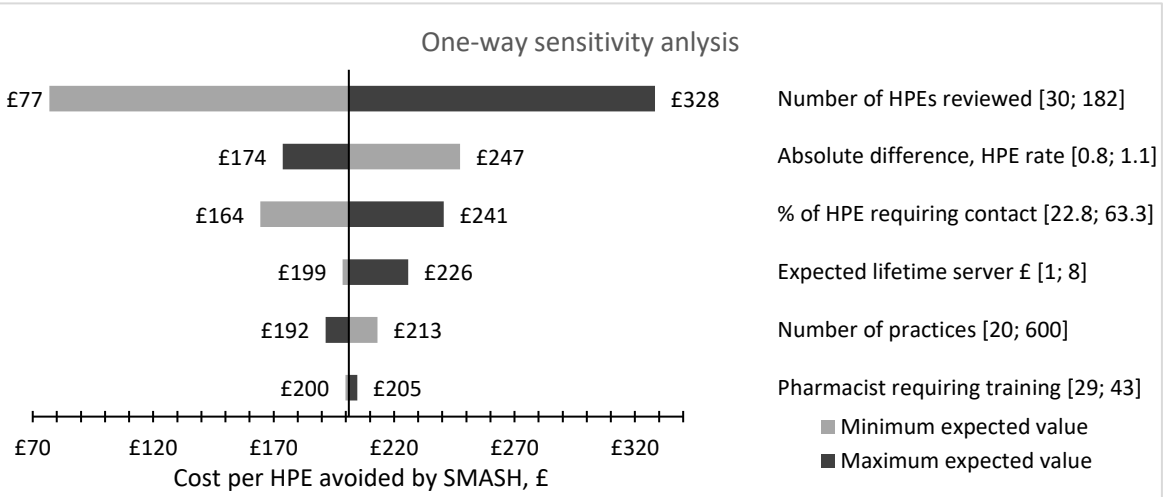
3.3.5.1 Scenario analysis

In scenario analysis 2, excluding set-up costs reduced the cost per HPE avoided by SMASH by £23 per practice [Table 3.8]. The incremental costs and effectiveness were smaller in the six months scenario compared with 12 months (£4 difference).

3.3.5.2 Deterministic sensitivity analysis

According to the results of the OSA reported in Figure 3.5, the cost per HPE avoided by SMASH was most sensitive to the parameter uncertainty around the number of HPEs reviewed. The cost per HPE avoided ranged from £77 to £328. The uncertainty around the absolute difference between HPE rates with SMASH and standard care changed the cost per HPE avoided by -14% (to £174) and +122% (to £247) compared with the mean expected cost per HPE avoided of £203for the minimum and maximum expected value, respectively. The OSA showed that of the individual cost components the uncertainty around managing HPE (number of HPEs reviewed and percentage of HPE requiring patient contact) had an impact on the cost per HPE avoided. The number of pharmacists that require training had the smallest impact on the total cost followed by the number of practices sharing the costs for server and the expected lifetime of the server.

Figure 3.5: Tornado diagram on impact of individual input parameters on the cost per HPE



Vertical axis cross: £203 per HPE avoided (deterministic); applied rangesin brackets (minimum expected value; maximum expected value); HPE: hazardous prescribing event; SMASH: Safety Medication Dashboard

3.4 Discussion

3.4.1 Principal findings

The study estimated the incremental costs per HPE avoided when using SMASH compared with standard practice in NHS England. Key findings were (i) a significant reduction of the relative number of HPEs by SMASH compared with standard care (10.53, 2.5% to 97.5% credible interval 8.80 to 12.25); (ii) incremental costs showed that this reduction came at a greater costs to NHS England (£2149, 2.5% to 97.5% credible interval £487 to £5790); (iii) main driver of the cost-effectiveness of SMASH were the costs of managing HPEs and the absolute difference in HPE rates between SMASH and standard care. The overall results of this study were robust to using any of the assumptions that were explored in the sensitivity analysis.

3.4.2 Comparison with findings from prior work

Not many interventions were found that identified cost per HPE or medication error avoided. The pharmacist-led information technology intervention (PINCER) used a similar design and setup as SMASH, aiming to reduce a similar set of HPEs [2.9.1]. PINCER was found to cost £80 per HPE avoided at 12 months (£79 at six months) in contrast to £205 in SMASH (227). The difference could be explained by the more regular reviews in SMASH (PINCER only reviewed patients with HPE three times in 12 months) or the fact that GP involvement was not included in the total cost of PINCER. The PINCER estimates were assessed for the 2012 cost year and older compared to the 2019 costs in SMASH.

A UK study testing the effectiveness of an intervention to reduce HPEs as part of ASPIRE in a trial setting did not report cost per HPE avoided but reported cost and effectiveness that can be compared to those of SMASH (152). ASPIRE aimed to reduce nine HPEs that were overlapping with the ten HPEs SMASH is aiming to reduce. The cost of ASPIRE for eleven months were £2439 per practice including intervention preparation, delivery and receipt. Not included in this estimate were costs associated with GPs and medication changes associated with actions to resolve HPEs. These were used in the subsequent cost-utility analysis and were dependent on the effect size. In SMASH, the total cost of £2149 included

costs for GP time spent with discussing actions to resolve HPEs with the pharmacist that costed £425 [Table 3.6]. Without GP resources SMASH costs were £1708, which is considerably less. ASPIRE was also less effective in reducing HPE rates. It reduced the HPE rates only to 4.94% compared with 5.99% in the control arm of the RCT with an OR of 1.03 (97.5% CI 0.89 to 1.18). The number of patients at risk in an average practice was not reported, hence, cost per HPE avoided could not be generated from the HPE rates and the cost estimates.

A Danish study implementing an automated medication system aiming to reduce administration errors costed €2.01 per administration error avoided at 6 months (290). The cost estimates only covered maintenance costs and did not include set-up or development costs. Cost of SMASH without set-up costs were £177 per HPE avoided at 12 months. The higher cost of SMASH could be explained by the time frame of 12 months compared with 6 months or the fact that the study aimed to automate the prescribing processes and not to add another layer of safety procedures to identify more HPEs, which comes at an additional cost for staff time.

3.4.3 Strengths and limitations of the cost analysis of SMASH

The assessment of costs of a complex DHI that involves multiple healthcare professionals required a complex costing approach. A strength of this analysis was the involvement of healthcare professionals in identifying resource use and validation of required assumptions. The healthcare professionals involved in the study rollout identified no further cost components that might have been missed and agreed that we used the limited available resources in the best way possible. The economic evaluation included costs from set-up to maintenance, which was often not done by other studies (290). Focussing solely on maintenance costs can underestimate the relevant costs that fall on budget constraints for healthcare (272). Additionally, this analysis was conceptually based on the real-world implementation of SMASH, and resource use was not estimated based on a hypothetical best-case scenario. Experiences of the real set-up of SMASH including potential challenges with pharmacists dropping out and meetings being adapted due to lack of time, which can impact the implementation, were used to increase external validity of the cost estimates.

The estimation of the specific cost generated in the 12 months roll-out of SMASH can introduce challenges with applicability of the costs to other settings. The cost for the server might vary between different providers or their expected lifetime might be different. This can affect allocation of the server cost and future discounting. The discounting strategy could impact the cost for the server or the cost for the IT support might differ, if not provided by the University of Manchester. In sensitivity analysis, however, the impact of varying these cost components on the overall results was small. Potential changes are therefore considered negligible.

The economic evaluation of the intervention was conducted retrospectively. Limitations of the intervention mainly derive from lack of quantitative information on resource use and the use of records, such as the field notes, that were not collected for costing purposes. Ideally, resource use would have been assessed prospectively to record purchased equipment, duration and participants of trainings or meetings, and time each staff member spent dealing with managing HPE or IT services. The use intensity of the dashboard was challenging to quantify. Use of the dashboard varied between pharmacists (246). While some pharmacists printed patient lists from the dashboard, others interacted with the various dashboard features to different extents. Recorded dashboard user interactions (247) could therefore not be used to estimate the real time pharmacists spent to review the HPEs. Results were sensitive to uncertainties around the time pharmacists and GPs spent with the intervention. The analysis relied on process indicators, such as the HPEs flagged by the intervention at specific time points and times they spent with each HPE on average, rather than records from the pharmacists how many HPEs were actually reviewed. Results from the sensitivity analysis supported the importance of these estimate for the cost per HPE avoided. To account for the uncertainty arising from the lack of recorded data, the mean of all new HPEs recorded and those possibly seen were used in the probabilistic analysis.

Because no information was available on the management of monitoring hazards detected by the dashboard, the cost assessment is limited to the cost of SMASH in managing HPEs. Monitoring hazards were responsible for about 7% of all hazards (monitoring and prescribing) identified at baseline. The effectiveness of SMASH in reducing monitoring

hazards and HPEs was analysed separately in Peek et al. (2020), and no combined effectiveness estimates were assessed. As a consequence, this analysis estimated the cost per HPE avoided, which potentially is higher than the cost per HPE and monitoring hazards avoided. The management of monitoring hazards is potentially less costly than managing HPEs, and the number of HPEs and monitoring hazards avoided is higher than the number of HPEs avoided alone.

A challenge encountered in this study was allocating costs between providers. SMASH was implemented across a group of providers, such that costs for server usage and IT services were shared. A systematic review of DHIs highlighted the lack of appropriate reporting of how costs were shared (277). In this study, the limitations of assumptions made to share costs between a group of practices was therefore reported in detail. Set-up costs per practice would decrease if costs were allocated to more practices. Practice sizes varied substantially from 970 to 15104 patients per practice in the SMASH rollout (48). The allocation of costs to practices is based on a 'typical' practice. No practice level data on HPE rates and practice size were reported in Peek et al. (2020) (48). Consequently, no sensitivity analysis was possible to estimate the effect of practice size on cost per HPE avoided. The sensitivity analysis of total cost showed that changes in the number of practices affected the total cost only minimally. Additionally, this study was robust against a change in the number of practices that shared the costs.

3.4.4 Strength and limitations of the cost-effectiveness analysis

The cost-effectiveness analysis in this chapter was based on the best available evidence and the findings present a real-world estimate of the cost per HPE avoided when SMASH is implemented. The probabilistic nature of the chosen analysis allowed to account for the uncertainty introduced by, for example, the substantial heterogeneity between practices with respect to effectiveness estimates.

So far, the cost per HPE avoided analysis only incorporates cost of SMASH. This neglects the fact that HPEs are associated with increased costs to the NHS and elevated levels of harm and death for patients (31). Walsh et al. (2017) found in a review on the economic

impact of HPEs that cost per general HPE ranged from €2.58 to €6432.16 (in 2015 €) (146). While SMASH directly effects HPE rates and it is sensible to measure these to indicate prescribing quality, this does not represent the potential cost-savings in the long term. The cost per HPE avoided reported in this study, thus, potentially overestimate the cost of SMASH. Additionally, incorporating clinical outcomes and quality of life into the CEA was recommended in the NICE framework for DHIs (12) and the CHEERS checklist [Appendix C]. Because no WTP threshold exists for reductions of HPE rates, the measure of cost per HPE avoided alone is not a useful statistic for decision makers.

The short time horizon contributes to these problems on the relevance of this CEA for decision makers. The CEA so far only generates estimates for a 12-months time horizon. Short time horizons were a common feature in CEAs of DHIs and have been criticised in a recent review (274). The effectiveness data in this study were only available for 12 months. No extrapolations were possible beyond this time frame because it was unclear how the HPE rates might change over time. Learning or decay effects were not explored in Peek et al. (2020)(48). Pharmacists working on the interventions suggested a flooring effect as indicated in the observed HPE rates but this could not be verified. In addition, no information on resource use after the 12 months was available.

Another limitation of this cost-effectiveness analysis was the lack of data on resource use in standard care. The analysis relied on the assumption that in standard care no resources are used that are not also used in the SMASH intervention. It cannot be completely ruled out that the medication review process in place before SMASH was introduced did not change during SMASH. For example, annual medication reviews were potentially less costly with SMASH because the number of potential HPEs that can be detected during these reviews is smaller with SMASH. However, there was no data available on how this might have changed and it will be difficult to quantify any such changes without a concurrent comparator.

3.4.5 Implications of quasi-experimental study design

A quasi-experimental study design was used to evaluate the effectiveness of SMASH. Deidda et al. (2019) and Kreif et al. (2012) provide checklists on how to assess the quality of the quasi-experimental evidence to be included in an economic evaluation. Both studies highlight the need to appropriately address potential biases and check if the main assumptions hold. As described in detail in Chapter Two [2.9.2], it is important to distinguish between different quasi-experimental analysis methods because they introduce different type of biases. This section describes the potential biases of the analysis method used in Peek et al. (2020) and how these were addressed. The second part focuses on the implications of the design on the economic evaluation in this study.

The method of analysis

SMASH was analysed using ITSA that is considered to be the strongest statistical analyses in quasi-experimental designs, where no comparator and only historical controls are available (13, 14, 264). Using a historical comparator, it was assumed that the population characteristics in each practice did not change significantly within the data collection period. This made control and intervention group comparable even without randomisation (minimises risk of time-invariant confounding). Also, the ITSA allowed for adjustments for any pre-intervention trends (295, 296). The HPE rates in SMASH were analysed using ITSA methods that allows for investigation of potential biases that might be present, for instance, secular trends, seasonality, random fluctuations and auto-correlation (297, 298). Testing for these biases, as performed for SMASH and as described in Appendix D, and accounting for them if required is pivotal to generate unbiased estimates (263, 264).

As described in section 2.9.2, ITSA designs are susceptible to history bias. For example, it cannot be ruled out that other factors or intercurrent events besides SMASH might have influenced the HPE rate in Salford. A concurrent intervention providing training for GPs to avoid HPEs could for example reduce HPE rates and explain away the observed effect. However, there were no indication of such interventions taking place concurrently. Compared with the RCTs conducted to assess the effectiveness of ASPIRE (152) and PINCER (45), the quasi-experimental evidence is more likely to be biased by intercurrent events if

no control data are available. By initiating SMASH at different time points between 2016 and 2017, the study design aimed to reduce this time-variant unmeasured confounding. For short-term effects, such as the 12-months in SMASH, bias due to potential external events affecting the outcome is low (299). Only with longer term interventions the risk of shocks to the effectiveness outcome is more likely to be a problem.

Another key assumption was that practices did not change over time to be representative comparators for the historic counterfactual. If the population at risk changes over time, this can introduce maturation bias. It is known that the number of patients at risk did not change a lot in the pre- and post-intervention period. In Peek et al. (2020), it was not reported how other characteristics of practices changed over time (48). The number of new GPs for example could have biased the results. If no GPs started post implementation, the learning effect over time could have been increased compared with a setting where GPs are replaced regularly. If there is imbalance between pre- and post-intervention characteristics in each practice, these could have been addressed by adjusted ITSA methods, such as a propensity score-based weighted method proposed by Linden and Adams (2011) (300). Detailed practice information was not available and so no adjustments were performed.

Another limitation of the methods applied in Peek et al. (2020) was the assumption of linearity in the ITSA. In this chapter, the difference between using regressed ITSA results and using the observed HPE rates was calculated. The available ITSA output was used to report the impact this had on the number of HPEs avoided. The number of HPEs avoided were sensitive to the regression model [Table 3.4]. At 12 months, applying the observed values for SMASH (not the regressed values) would have yielded a smaller incremental effectiveness. The difference between the two methods, however, was smaller at six months where the incremental effectiveness was more similar in SMASH and the counterfactual. This effect was indicated in Figure 3.2 that graphically represents the HPE rates observed and the regressions applied. The assumed linear trend is one of the challenges with this type of ITSA where linearity of the effect before and after the intervention is assumed. If the regressed values would predict the observed values appropriately, the two methods should produce the same effectiveness estimates. In

SMASH, there was a visible and explainable floor effect of the HPE rate after initiation of the intervention. This supports the findings in this study that the difference between expected values in the ITSA and the observed values increases with time. This was one of the reasons why the regressed values for the HPE rate with SMASH cannot be extrapolated beyond the 12 months where the fit of the regression was still reasonable. Future research could use a more flexible modelling approach to model the post-intervention HPE rates. Consequently, the results could not be extrapolated further and this CEA had a time horizon of 12 months equal to the observation time in the effectiveness study.

The random effects meta-analysis after the ITSA allowed the researchers to generate an estimate of heterogeneity between practices. Not accounting for heterogeneity was seen as a key limitation of many ITSA methods identified by the systematic review by Ewusie et al. (2020) (263). The authors criticized the use of aggregate data, which introduces aggregation bias that segmented linear regression does not account for (263). The systematic review by Ewusie et al. (2020) highlighted one study positively that introduced pooling methods using meta-analysis as was done for SMASH. HPE rates of individual indicators were often low with a mean number of three HPEs per practice. Performing ITSAs on these low numbers could weaken the segmented regression compared to pooled data that could provide a larger sample size. This could result in wider standard errors of the effectiveness estimates that further increases the uncertainty in the cost-effectiveness analysis. Future research could investigate a best practice method when pooling data from different practices and whether this should be conducted before or after the ITSA.

Quasi-experimental studies in economic evaluation

While guidelines exist on how to plan and conduct ITSA to measure effectiveness (265), the literature around applying these results to estimating cost-effectiveness was sparse. The available literature on guidelines how to use quasi-experimental study designs in economic evaluations focuses mainly on quasi-experimental designs in general (10) or controlled observational studies in particular (11). Deidda et al. (2019) provide a detailed checklist of requirements for the use of observational data in economic evaluations (10). The checklist by Deidda et al. (2019) is targeted at studies estimating not only effectiveness outcomes

but also costs prospectively. Peek et al. (2020) did not assess costs and only provided estimates of effectiveness. The majority of criteria were therefore not applicable to the retrospective cost assessment conducted in this study. The checklist items around the statistical methods, however, were applicable. The statistical analysis by Peek et al. (2020) fulfilled most of the requirements to estimate effectiveness described by Deidda et al. (2019). As described before, Peek et al. (2020) addressed the potential for history and selection bias by study design. Sufficient data points were available pre- and post-intervention start. Peek et al. (2020), however, did not address the checklist item to apply multiple statistical design methods. This would have been useful in sensitivity analysis of the economic evaluation and is a limitation of the provided data from Peek et al. (2020).

A literature review found only one model based economic evaluation that used results from a quasi-experimental study using ITSA (301). The lead author of this study was contacted to obtain insight into the detailed methods used. The study calculated a population average for the outcome and added an absolute difference based on the ITSA results for the comparator group. Only the absolute difference from the ITSA was used as the effect size, but the baseline probabilities were derived from a different population. In this Chapter, the absolute difference based on the expected values of the ITSA were also used in the base case analysis. Compared with the described study, the observed HPE rate with SMASH at 12 months was used as the baseline probability. This had the advantage that both estimates were from the same population. The OSA showed how sensitive the cost-effectiveness estimates were to the absolute difference of HPE rates in SMASH and standard care. The uncertainty derived from the ITSA results greatly affected the cost per HPE avoided. Future research could investigate the effectiveness of SMASH in a larger sample than the 43 practices in Salford, such as the rollout of SMASH to Greater Manchester, that could reduce the uncertainty around the effect size generated in Peek et al. (2020).

3.4.6 Implications of the use of process indicators

In SMASH, the economic evaluation depended on the use of HPE rates, a process indicator generated from routine data, rather than patient outcome data as recommended in the checklist by Deidda et al. (2019) (10). The HPE rate, as a process indicator, recorded by the

dashboard was difficult to interpret. HPEs could have been resolved automatically due to routine stopping of treatment and not due to SMASH. This stopping of treatment would still be shown as a resolved HPE in the dashboard records. HPE rates, therefore, introduce some form of measurement bias. However, the effectiveness data compared against extrapolations of historical rates for the last 24 months before use of SMASH. This accounted for prior trends for routine stopping, assuming routine stopping did not change over time. Additionally, SMASH did not resolve all HPEs. This can be explained by the fact that some HPEs can be prescribed intentionally. One example would be that the patient's condition requires the hazardous prescriptions because the risk of harm, e.g., stroke, without the hazardous prescription outweighs the risk of harm from ADEs with the hazardous prescription. Some HPEs were not resolved because patients refused a change in treatment or did not request follow-up prescriptions as reported in the field notes from pharmacists. In some cases, the GP refused to change prescriptions if the patient tolerated them well in the past despite the increased risk of ADEs. Future research could investigate what the optimal (possible) HPE rate would be acknowledging the measurement error and reasons why HPEs cannot be resolved. The optimal HPE rate could be used as a reference to evaluate the observed effectiveness.

3.4.7 Implications for thesis

This analysis was not able to estimate economic impact of harm, as only HPEs were recorded, not patient outcomes. Defining outcome measures was found to be one of the major challenges in evaluating DHIs (280). A review of DHIs found effectiveness studies not to report outcome measures that can be translated into health economic endpoints, such as QALYs (276). The review recommends measuring the effect on harm outcomes or hospitalisations. However, capturing the long-term effect of DHIs was reported to be challenging and rarely done (277). To analyse the relationship between HPEs and patient harm in a cohort creates methodological and ethical challenges. Methodological difficulties arise in prospective study designs with the small percentage of HPEs that actually lead to harm for the patient resulting in extremely large populations required to obtain sufficient power to detect differences in outcome events. The rollout of SMASH was limited to practices within Salford, and the population size was not expected to be sufficient to detect

a difference in serious harm outcomes associated with HPE prevalence. In addition, depending on the pathology of the HPE, there can be a delay between the occurrence of an HPE and the actual outcome, which could not have been detected in the 12 months follow-up. Ethical concerns impede the follow-up of patients where potentially harmful HPEs are identified with ethical obligations to intervene to avoid serious harm. In Chapter Two, the most common effectiveness measures in economic evaluations of interventions aiming to reduce HPE rates were either HPE rates or incidence of ADEs. Because the analysis of SMASH in Peek et al. (2020) was not powered to detect a difference between ADE rates, this latter method was not an option for this study (48).

The current CEA might underestimate the benefit of SMASH due to the short time horizon and the neglection of long term consequences of HPEs as described in 3.4.4. Under the assumption that higher HPE rates result in higher healthcare resource use and decrease quality of life, SMASH could generate fewer costs and higher quality of life in the long term. However, as described in Chapter Two [2.4 and 2.5], quantitative evidence on consequences of HPEs is sparse (31), and data used in other economic models were not appropriate for the setting and or population in this study. Future research can use data collected in routine practice to quantify the association between HPEs, harm to patients and direct costs to the healthcare system. This study can be extended to include estimates of long-term harm to estimate the cost per QALY gained. In the subsequent chapters, this dissertation demonstrates how these harm estimates can be generated using electronic health records [Chapter Four] and how harm estimates can be utilised to estimate the economic impact of HPEs [Chapter Five]. A state-transition model is used to extrapolate long-term costs and quality of life associated with the ADEs related to a specific HPE. The incremental costs and QALYs can then be compared to the cost-effectiveness threshold used by NICE to enable a better understanding of whether SMASH provides value for money to the NHS.

3.4.8 Implications for decision makers

The UK government aims to reduce hazardous prescribing and harm related to it (35). System level DHIs, such as SMASH, offer a solution to this. The additional resource required

probably increases costs. Evidence on long-term effectiveness and impact on health outcomes is uncertain. Hence, we need robust economic evidence to extrapolate the long-term impact of SMASH. In the meantime, a rollout of SMASH to about 600 new practices in Greater Manchester started in late 2020. This raises the question of transferability of intervention costs. Costs were found to be robust with regard to the number of practices recruited. For the implementation of SMASH in Salford, the Salford Royal Foundation Trust charged for accessing the server. For the rollout in Greater Manchester, different hosting costs might arise for the server itself and for annual server support. However, the total cost per HPE avoided were only slightly sensitive to these input parameters [Figure 3.5].

The majority of costs generated by SMASH resulted from managing HPEs. Participant 2 suggested that costs of running SMASH will reduce over time based on their experiences with SMASH after the follow-up phase of the rollout. There may be a substitution of labour inputs responsible for initially reviewing each HPE as relevant (pharmacist technicians instead of pharmacists) and for resolving identified HPEs (independent prescriber pharmacists instead of GPs). This evolving nature of the implementation of DHIs over time and difficulties with accounting for this in economic analysis has also been identified by other authors (272).

The results reported in this chapter are only generalizable to practices with no initiatives in place that aim to reduce HPEs. SMASH is part of the Neighbourhood Integrated Practice Pharmacists in Salford (NIPPS) service (302). Employed by the trust, pharmacists provide services to the practices in their neighbourhood, including reviewing HPEs flagged by SMASH. The NIPPS service aimed to involve pharmacists more in primary care to improve patient safety. If the intervention was to be implemented in other practices where pharmacist's services are already in place, this might reduce the additional costs for the pharmacists to review the flagged HPEs. If reviewing HPEs is already integrated in the practice routine, the dashboard could potentially make the review process more time efficient and consequently less costly. In such practices, the time pharmacists spend with dealing with HPEs would not be solely allocated to the intervention arm but would also be generating costs in the comparator. Depending on the pre-intervention state of practices with regard to practice pharmacists or pharmacist services, such as NIPPS, the cost for the

standard care arm might change. So far, the cost in the standard entailed no additional resource use because SMASH was seen as an addition to standard care not a substitution of services already in place. If pharmacist services are already standard practice, the resources for this service would be costed in the standard care arm. This would result in higher costs for standard care and therefore lower incremental costs for SMASH.

3.4.9 Conclusion

This study represents a first step towards assessing the economic impact of SMASH. While SMASH reduced the number of HPEs per practice, these now need to be translated into tangible health outcomes and costs that are associated with HPEs. Expressing health outcomes in terms of QALYs, instead of the proportion of HPEs avoided, enables the analysis of the full value for money of SMASH.

Chapter 4 - Examining the increased likelihood of adverse drug events associated with hazardous prescribing of NSAIDs

Chapter Four reports the methods and results of a cohort study using routinely collected health data to examine the increased likelihood of ADEs in the presence of NSAIDs in anticoagulated patients. The methods provide information on the study design, the data source and statistical analysis. Challenges encountered when using routinely collected data and how these were addressed in this study are described in the discussion of this chapter.

4.1 Introduction

ADEs affect almost one in ten hospitalised patients (303). The burden from six of these ADEs as a result of safety incidents was found to be higher than the burden of HIV and tuberculosis, or the burden of multiple sclerosis or cervical cancer measured as DALYs in English hospitals (304). Harm from ADEs was associated with 15% of the activity and expenditure of hospitals in developed countries in a recent review of the literature undertaken by the Organisation for Economic Co-operation and Development (OECD) (169). Almost half of ADEs are considered to be preventable, hence potentially due to a HPE (303). The WHO's third global patient safety challenge calls for actions to reduce harm from HPEs by 50% by 2021 (33). While the problem of HPEs was discussed in detail in the technical reports of the WHO safety challenge, it did not state how harm associated with HPEs can and should be measured. Estimating harm from HPEs is necessary to understand the burden of HPEs for the healthcare system. Various interventions have been developed aiming to reduce HPE rates, but studies evaluating these interventions often did not measure harm from HPEs as an outcome [2.8]. The studies assessing cost-effectiveness of interventions aiming to reduce HPEs relied on estimates of harm from the literature or estimates were elicited by experts, but there is a lack of high quality quantitative evidence on harm associated with HPEs (31). Quantitative evidence on the health and economic impact of HPEs is required to understand the relevance of HPE reductions by new interventions. Cost-effectiveness of interventions aiming to reduce HPE rates cannot be estimated appropriately without quantitative evidence on the link between HPEs and

harm. Methods used to study harm from HPEs have substantial limitations as described in Chapter Two [2.4]. The low incidence of HPEs and potentially long delays until onset of risk of harm require long follow-up times and large population sample sizes. Conducting a prospective study with long follow-up in a large population, such as an RCT, would be extremely costly. An RCT might also introduce ethical challenges, when estimating the natural history of harm from a HPE because patients known to have a HPE cannot be followed without ethical obligations to intervene. Overall, prospective study designs to estimate the harm from HPEs are not a method of choice due to financial, organisational and ethical reasons.

In the literature review in Chapter Two [2.4] on consequence of HPEs, two common methods were identified to estimate harm without the financial commitment of prospective follow-up of patients. In previous economic evaluations, potential harm from HPEs was either (i) estimated by experts or (ii) harm in the form of ADEs was measured and its preventability or association with HPEs was assessed. So, either healthcare professionals predicted potential future harm of identified HPEs, or real harm was tested for associations with HPEs. In the first method, the potential future risk of harm of an incident HPE is estimated by healthcare professionals. Often healthcare professionals used severity scales or algorithms to guide their judgement on harm. However, this method was found not to produce reproducible and reliable estimates. Harm estimates greatly differed between different scales or algorithms used, the healthcare professional's experience or the healthcare profession (144). Another limitation of this method is that it does not acknowledge that some HPEs are picked up and resolved before the harm occurs. This estimate of potential harm is therefore not a good measure to quantify harm from HPEs that actually reach the patient.

In the second method used in economic evaluations of interventions targeting HPEs, harm is measured directly as ADEs, and retrospectively healthcare professionals assess if these were caused by preventable HPEs or are a result of other medical conditions (117). Similar to estimating future risk of a HPEs, identifying causality between the observed ADEs and the HPEs is dependent on subjective assessment by healthcare professionals (54, 112, 113). Even though new instruments using an algorithm to judge on preventability aim to reduce

the input required by the healthcare professional to increase reliability of the results, the impact of the assessor on the results cannot be eliminated (107). Similar to the severity scales or algorithms used to estimate future harm from HPEs, the instruments used for the retrospective assessment of preventability and causality showed varying reliability and evidence on external validity of the instruments was often missing (111). In SMASH, incidences of ADEs were not measured and this method was therefore not an option to estimate harm from HPEs targeted by SMASH.

A third method identified in economic evaluations in Chapter Two [2.8] used harm estimates for specific HPEs from the literature. By narrowing down the possible ADEs to those with a clear association with the HPE, incidence of harm for a population with the HPE present and for those at risk of the HPE can be used from observational studies. The use of routinely collected data offers the analysis of a large cohort at a lower cost (305).

This chapter delineates how to conduct a retrospective cohort study using routinely collected data to investigate the association of HPEs with harm outcomes, using the example of NSAID use in anticoagulated patients. Linked primary and secondary care health record data were used in a UK setting. The primary care database used was the Clinical Practice Research Datalink (CPRD Gold) and the secondary data were used from Hospital Episodes Statistics (HES). The retrospective cohort study design used in this chapter focusses on specific harm outcomes that are potentially caused by the presence of a specific HPE. NSAID use is associated with various potential ADEs (306). The risk of GI bleeding events in anticoagulated patients, for example, is known to be particularly increased in situations where NSAIDs are co-prescribed with OACs (307-310). Under the assumption that the additional bleeding events identified in the patients exposed to the HPE (compared to those not exposed) are related to the HPE, further judgement on preventability or potential future harm is not required. This potentially allows for a more reliable estimation of the association between HPE and harm because subjective assessment methods are not needed.

The aim of this chapter is to estimate the association between NSAID use and related ADEs in patients receiving treatment with OACs in a UK ‘real world’ setting.

4.2 Background

OACs and NSAIDs are reported to be the drugs most commonly associated with preventable harm and preventable drug related hospital admissions (26, 311-314). The frequency of OAC prescriptions is increasing in the UK from 15.0 million doses per month in 2014 to 33.0 million doses per month in 2019 (315). With increasing prescription rates, the population at risk of this type of HPE is increasing.

4.2.1 The impact on GI bleeding events

OACs reduce cardiovascular risk by intervening in blood-clotting mechanisms and are used in the prevention of thrombotic cardiovascular events, such as stroke and systemic embolism. The same mechanism that reduces cardiovascular risk, also increases the risk of unwanted bleeding events of all kinds. While OACs can cause multiple types of bleeding, NSAIDs are specifically associated with GI side effects. NSAIDs alone are considered to increase the GI bleeding event risk (316-323). NSAIDs can cause damage to the GI system by various mechanisms (324). For example, NSAIDs can inhibit gastroprotective enzymes, e.g., prostaglandins, by blocking the enzyme cyclooxygenase. These enzymes are required to protect the GI system from peptic ulcers or bleeding events. NSAIDs also interact with phospholipids. The mucosa integrates NSAIDs instead of the phospholipids due to similarities in the chemical structure. As a result, the mucosa's sensitivity to acid, pepsin and other aggressors in the GI tract is increased. This amplifies the risk of GI bleeding events.

GI bleeding events are among the most common side effects of OACs and NSAIDs individually (318, 325-327). In general, GI bleeding events are rare but often have severe consequences (328). A review of the epidemiology of GI bleeding events reported GI bleed rates with NSAID use of 2.4%-12%, and with OACs between 0.7 to 1.4 per 1000 patient years (329). According to an observational study using the CPRD, 40% of GI bleeding events under antithrombotic treatment led to death within two years (330). Pirmohamed et al. (2004) found that more than 50% of deaths due to ADEs were caused by GI bleeding events (54). OAC and NSAIDs individually introduce a high risk of GI bleeding events, but evidence on their effect on GI bleeding events when prescribed concomitantly is sparse. Overall,

evidence suggests that the presence of NSAIDs in anticoagulated patients affects bleeding event risks and hospitalisations due to bleeding (307-309, 331, 332).

In Table 4.1, published studies on the increased risk of GI bleeding in anticoagulated patients with NSAIDs compared with no NSAIDs are reported. Evidence in the literature often describes non-UK populations (307, 330, 331, 333), or describes populations slightly different from the HPE definitions.

Table 4.1: Increased risk of GI bleeding events associated with NSAID use in anticoagulated patients from published studies

Reference	Risk ratio (95% CI)	Population	Dataset
Battistella (2005) (331)	Adj. OR 1.90 (1.40; 3.7)	Warfarin treated patients aged ≥66 years	Canadian administrative data
Lamberts (2014) (307)	Adj. HR 3.54 (3.29; 3.82)	Warfarin treated patients with AF	Danish registry
Schjerning (2019) (310)	Adj. HR 2.01 (1.40; 2.61)	DOAC treated patients with AF	Danish registry
Kent (2018) (308)	Adj. HR 1.68 (1.40; 2.02)	Rivaroxaban or warfarin treated patients with AF	Subgroup analysis of RE-LY RCT
Dalgaard (2020) (309)	Adj. HR 1.16 (0.88; 1.52)	Dabigatran or warfarin treated patients with AF	Subgroup analysis of ARISTOTLE RCT

Adj: adjusted risk ratio; AF: atrial fibrillation; ARISTOTLE: Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation; DOAC: direct oral anticoagulant; GI: gastro-intestinal; NSAID: non-steroidal anti-inflammatory drug; HR: hazard ratio; OR: odds ratio; RCT: randomised controlled trial; RE-LY: The Randomized Evaluation of Long-Term Anticoagulation Therapy

Differences can be a restriction to specific indications, drugs or exclusion of patient groups. For example, most studies were restricted to patients with atrial fibrillation (AF) (307-309, 333). OACs are initiated for AF in 60% and for VTE in 30% of the cases identified in the UK CPRD population (334). The RCTs in Table 4.1 for example describe subgroup analyses of clinical trials to compare the new DOACs with established warfarin treatment restricted to AF (308, 309, 335). The restriction to one of the indications does not represent the average OAC user, since this excludes 40% of the users (335). Other studies only investigate warfarin and not all OACs as defined for the HPE (307, 331, 333). These are usually studies conducted before the introduction of DOACs, for example, the Canadian study by Battistella et al.

(2005) that only included warfarin and other vitamin-K antagonists (331). DOACs already accounted for 74% of OAC prescriptions in the UK in 2019 and need to be considered because these might have a different risk of ADEs (315). The subgroup analyses of the RCTs often exclude high risk patient groups, e.g., patients with a high bleeding risk are excluded from the trial (308, 335). Patients might also withdraw from the trial as soon as they show any GI symptoms, so the numbers of serious events could be artificially small. This impedes generalisability of the results because the relative risk of bleeds might differ between high and low-risk patients.

Overall, the published literature did not provide any UK estimates on harm associated with NSAID use in anticoagulated patients. The studies all included restrictions limiting the generalisability to a UK setting and to the HPE type targeted by SMASH.

In response to the global safety challenge of the WHO, the UK government introduced the NHS Medication Safety Dashboard that links specific patients exposed to a specific HPE type with hospital admissions in England (336). The dashboard provides useful estimates on the number of patients exposed to the specific types of HPEs in England (HPE prevalence) and their risk of admissions due to specific ADEs. The dashboard, however, does not analyse the risk ratio of patients exposed and unexposed to the HPE.

The dashboard is accessible online and provides information on the same HPE types associated with an increased GI bleeding risk as used in SMASH and PINCER. Information on prescription items is derived from NHS Business Services Authority reimbursement records that only include dispensed drugs. Admission data are obtained from Hospital Episodes Admitted Patient Care, covering all admissions at NHS hospitals in England. Patient records and admissions were linked by NHS number, date of birth and gender. The dashboard's results include, for example, any hospital admission in a month due to GI bleeds in patients with prescriptions of an OAC and an NSAID in this month. In the fourth quarter of the financial year 2019/2020 (most recent available in November 2020), 29 out of 10000 patients with concomitant OAC and NSAID were admitted to hospital for GI bleeding events (336). The NHS Medication Safety Dashboard presents an estimate of harm from an English population in OAC users from any OAC indications without any restrictions

of the population, compared to the previously discussed published harm estimates. However, exposure status in the dashboard is estimated on a monthly basis. An anticoagulated patient that is, for example, hospitalised for a GI bleeding event in the beginning of the month and receives a prescription of an NSAID in the end of this month would be counted as an admission in a patient with the HPE even though the prescriptions were not really prescribed concomitantly before the admission. Consequently, the dashboard might overestimate the number of patients exposed to the HPE.

Overall, the dashboard does not provide the increased likelihood of ADEs of NSAID users compared with no NSAID users, but it provides an estimate of prevalence of the type of HPEs. The additional GI bleeding risk associated with NSAID use in all anticoagulated patients including DOACs has not been studied yet in a real-world population. This is also the first study to investigate the increased risk of NSAIDs in OAC patients in the UK.

4.2.2 The impact on cardiovascular events

Besides GI bleeding events, in recent years NSAID use has also been associated with an increased risk of cardiovascular events (337). Because patients treated with OACs are already at an increased risk of thrombotic cardiovascular events, the impact of NSAIDs in anticoagulated patients is of particular interest. Four studies were identified that investigated NSAID use in anticoagulated patients on cardiovascular ADEs, for example, stroke, systemic embolism, heart failure and myocardial infarction (307-309, 338). In a subgroup analysis of the Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) trial [Table 4.1] comparing dabigatran and warfarin, NSAIDs were significantly associated with stroke and systemic embolism (308). The combined stroke and systemic embolism outcome were also significantly increased with NSAID use in a Danish observational study [Table 4.1] (307). Contrary, a Japanese observational study found NSAID use not to be an independent risk factor of stroke/systemic embolism risk (338). In a subgroup analysis of the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial [Table 4.1] comparing apixaban and warfarin, NSAIDs were also not significantly associated with the combined stroke or systemic embolism outcome (309). Overall, these studies indicate a potentially increased risk of

stroke with NSAID use in an AF population treated with OACs. The risk of systemic embolism in both OAC, and OAC and NSAID users was considerably small (308, 309). The rarity of systemic embolism might have contributed to the often non-significant results on the relative risk increase with NSAIDs. Heart failure (309) and myocardial infarction (308) were not significantly associated with NSAID use in the respective trial subgroup analysis. Due to lack of evidence to suggest otherwise both events were not considered in this analysis. These findings do not completely rule out the chance that the risk of these ADEs is elevated in anticoagulated patients in the presence of NSAIDs. However, for this analysis potential ADEs were prioritised where data was available indicating an elevated risk.

In summary, evidence exists that the risk of stroke and systemic embolism in anticoagulated patients might be increased with concomitant NSAID use. Consequently, this study aims to investigate the increased likelihood of strokes and systemic embolism in anticoagulated patients in the presence of NSAIDs.

4.3 Methods

This section illustrates the methods used to estimate the association of NSAID use with ADEs in anticoagulated patients. The RECORD, RECORD-PE, STROBE reporting guidelines for studies using routinely collected health data were followed (339). The checklists are reported in Appendix F with references where in this study the items of the guidelines were addressed. The overall study design [4.3.1] and the data source used in this observational study are described [0]. Criteria to define the study population, exposure and identification of outcome measures are explained in sections 4.3.3 to 4.3.5. After the data were made available, it was cleaned and prepared for data analysis as described in section 4.3.6. The process how potential confounding variables were assessed and the method used to control for confounding are presented in section 4.3.7. Finally, this section introduces the statistical analysis and planned sensitivity analysis [4.3.8, 4.3.9].

4.3.1 Study design

The study estimated the incidence of HPEs and the association between the HPE type and clinical outcomes using an observational cohort design for a population from the linkage of CPRD GOLD, HES and ONS from 1 April 2007 to 31 December 2017. Temporal anchors relevant to the study design and the study population are listed in Table 4.2 as recommended by Patorno et al. (2020) to increase transparency (340).

Table 4.2: Temporal anchors for the observational cohort study

Term	Definition
Study period	1 April 2007- 31 December 2017
Cohort entry	First OAC prescription in study period in patients ≥18 years, at least 12 months of follow-up in an up to standard practice; cohort entry=index date
Outcome event date	Day of first serious GI event or stroke as recorded in hospital records (episode start date in HES inpatient data)
Washout window for denominator	Not applicable
Washout window for exposure	NSAID prescriptions prescribed 90 days before index date
Washout window for outcome	Not applicable; different assumptions tested in sensitivity analysis
Exclusion assessment window	At baseline, index date
Covariate assessment window	For comorbidities, all entries before index were considered. For concomitant drugs, prescriptions 6 months before index were assessed; BMI and blood pressure were identified from records 12 months before index
Exposure assessment window	Time varying exposure of NSAID use during continuous OAC use
Follow-up start	Start of follow-up at index date (first OAC prescription in study period)
Follow-up end	First occurrence of either: (i) Outcome event; (ii) transfer out date; (iii) last collection data in CPRD GOLD; (iv) death date (ONS/CPRD); calculated OAC stop date with a 30d grace period of last consecutive OAC prescription; or (v) end of study period.

BMI: body mass index; CPRD GOLD: Clinical Practice Research Datalink GOLD; GI: gastro-intestinal; HES: Hospital Episodes Statistics; NSAID: non-steroidal anti-inflammatory drugs; OAC: oral anticoagulant; ONS: Office for National Statistics

4.3.2 Data source

This study used the CPRD GOLD, HES and ONS linkage dataset, which offered the opportunity to analyse the incidence of HPEs from prescription and diagnosis records in primary care and to link these with mortality records and serious harm outcomes in

secondary care. For example, linkage to inpatient HES data is required to identify bleeding complications leading to hospital admission. Linkage to ONS mortality records is used as an additional information source for the identification of cause-specific death.

CPRD is a research service funded by the government that provides anonymised health record data from general practices in the UK since 1987. The CPRD Gold dataset, from here on referred to as CPRD, is one of the largest longitudinal datasets in primary care (341) and collects clinical and prescription data on patient level from around 600 practices in the UK covering 11.3 million patients (342). Around 4.4 million patients of these 11.3 million were currently registered on 2 July 2013, covering approximately 7% of the UK population. The dataset has been shown to be broadly representative of the UK population in terms of age, sex and ethnicity (342, 343), and validity of diagnostic coding is high (344, 345).

This study linked the primary care records from CPRD with secondary care records from HES Admitted Patient Care records. HES comprised admission details and clinical data for all secondary care attendances in England. 58% of primary care practices in CPRD have agreed to data linkage with HES (342). Patient level data where HES linkage is available were linked to the ONS mortality dataset. The ONS mortality records provide patient level death records including cause and date of death for patients in England and Wales (346). The data were also linked with the Index of Multiple Deprivation (IMD) at the patient postcode level. SMASH and PINCER were implemented in English practices, hence the linked records of English patients, was an optimal choice to identify HPE presence and patients at risk of HPEs.

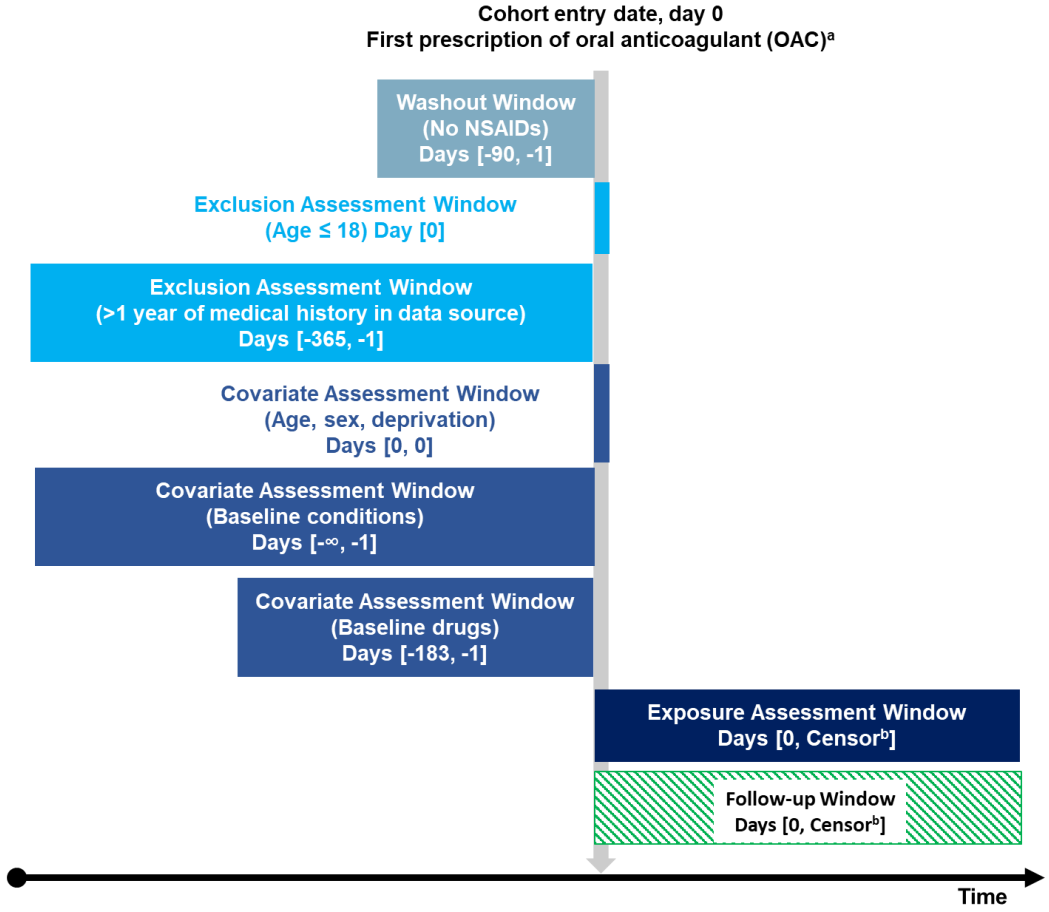
Access to the linked dataset is subject to protocol approval by the Independent Scientific Advisory Committee for MHRA database research (ISAC). ISAC is a non-statutory expert advisory body, established in 2006 to give advice on research-related requests to access data from the CPRD and linked datasets. The protocol is required to report feasibility, quality and public health value of the planned research, as well as detailed descriptions of the data analysis plan and limitations of the study design. The protocol (No 18_235) was written as part of this dissertation and was approved on 12 August 2018 with no revisions.

After access was granted, a data management plan was developed to guarantee the appropriate use, storage and deletion of the data once the project has ended (347).

4.3.3 Study population

The cohort was based on the denominator and numerator descriptions of the HPE [Appendix B]. Patients were at risk of the HPE (denominator) if they were at least 18 years of age and had any prescription of an OAC. OACs include the traditional VKAs, such as warfarin, and the DOACs that cover rivaroxaban, edoxaban, dabigatran and apixaban. Application of exclusion criteria is reported in Figure 4.1, which uses a reporting design recommended by best-practice guidelines by Patorno et al. (2020) (340).

Figure 4.1: Exclusion criteria and time anchors for cohort entry and follow-up of patients at risk of the HPE (patients with oral anticoagulants) (340)



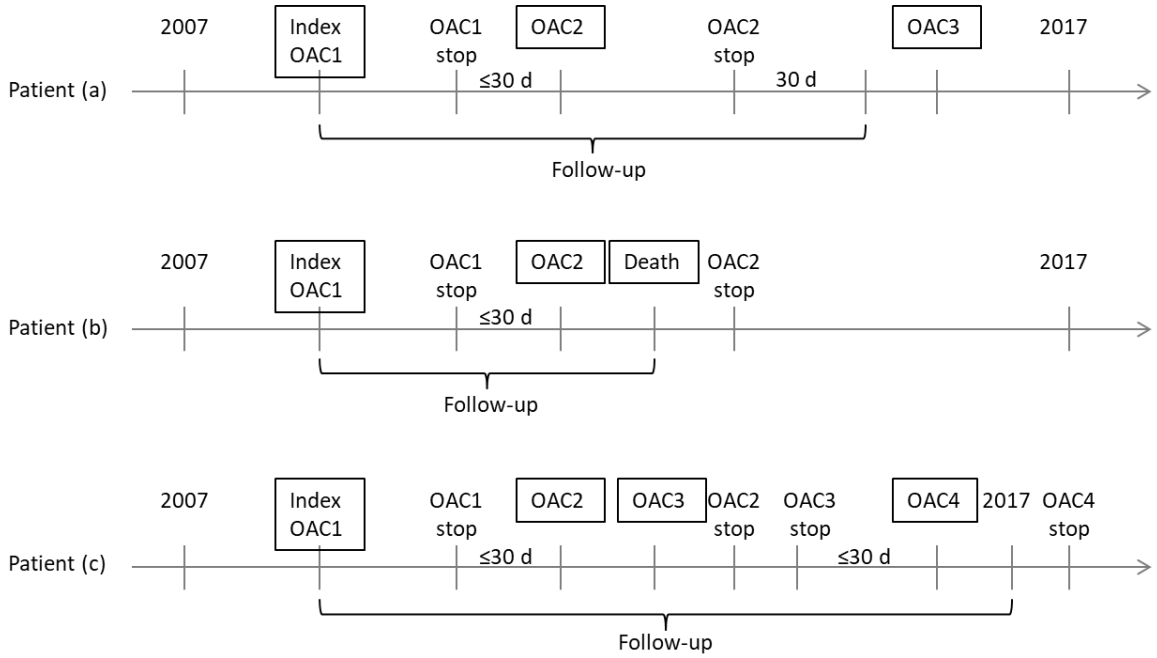
[a]Treatment episodes defined by prescription day and calculated stop date. Gaps of less than 30 days between stop date and next prescription were bridged. Thirty days were added to the last stop date of consecutive prescriptions; [b]earliest of: outcome of interest (serious GI event or stroke), discontinuation of OAC (including 30-day washout), death, last collection date of practice, transfer out date, end of the study period

Patients were excluded if they received a prescription of an NSAID 90 days before the index date, referred to as the 'new user' design, as proposed by Ray et al. (2003) (348). Of all NSAID prescriptions in anticoagulated patients in the study period, 99% of the prescriptions had a length of less than 60 days. Assuming a maximum prescription length of 60 days and adding a 30-day grace period resulted in an interval of 90 days. The 90 days were considered appropriate to exclude all prevalent users of NSAIDs. Including prevalent users can introduce various biases (348, 349). Firstly, the bleeding risk associated with NSAIDs is time-dependent with a higher absolute risk in the first three months of treatment (307). This can introduce bias due to left censoring because (i) patients who received an NSAID and had GI bleeding events in the high risk initiation phase before the index date are not included and (ii) patients who died before index date under NSAID treatment do not contribute any follow-up time (349). In the prevalent user design, it is assumed that no events can happen in the time before index date. Secondly, there may be a 'healthy user effect' if patients with prevalent NSAID treatment are more likely to continue treatment if no bleeding event had occurred. When a bleeding event occurred under NSAID treatment patients are more likely to discontinue and not to reinitiate NSAID treatment. Hence, the NSAID exposure group would include more 'healthy' patients that used NSAIDs without complications and include fewer patients who did not tolerate NSAIDs. Another advantage of the new user design is that baseline characteristics are automatically pre-treatment covariates. In the prevalent user design, it is difficult to separate pre- and post-treatment covariates. Conditioning on post-treatment covariates can lead to over-adjustment (349).

Patients from the study period were eligible for entry into the cohort if they were 18 years of age or older, were registered with a CPRD-participating practice (up-to-standard practices only) for at least 12 months prior to cohort entry and fulfilled the following requirements: The 'at-risk' population included all patients with a prescription of an OAC in the study period. Start of follow-up was defined as the first prescription of an OAC in the study period (index date). The follow-up was described as a continuous treatment episode of OACs. Treatment was considered to be continuous if the time between stopping and starting an OAC prescription was 30 days or less [Figure 4.2a]. A grace period of 30 days was considered to be an appropriate time to account for overlapping repeat prescriptions, where the patient has doses left at the end of treatment or potential hospital stays of

patients where no new prescriptions were issued from the GP because the hospital provides the medication during the stay. A US study found that 95% and 89% of gaps were <30 days, and 98% and 94% of gaps were <60 days, for prescriptions of dabigatran and warfarin, respectively (350). For the base case analysis, a 30-day grace period was considered appropriate, as this was around the median prescription length of OACs and NSAIDs in the CPRD dataset in the study period. A 60-day grace period was analysed in sensitivity analysis. Figure 4.2 illustrates how the 30-day grace period was applied using three hypothetical examples of patients who were prescribed OACs over time. Patients who switched between OAC treatments with a gap of less than 30 days were followed continuously.

Figure 4.2: Graphical presentation of follow-up times of three hypothetical patients



Patient (a): Follow-up ends after treatment gap of ≥30 days (calculated OAC stop date plus 30-day grace period)
Patient (b): Follow-up ends with death during continuous OAC treatment
Patient (c): Follow-up ends with study end during continuous OAC treatment

OAC: prescription date of oral anticoagulant from CPRD; OAC stop: calculated end of prescription; OAC1/2: numbered repeat prescription; Index: index date; Death: date of death from ONS; 2007-2017: study period; □: dates recorded in health records

Follow-up ended at the first occurrence of either: (i) stop of OAC for more than 30 days (Figure 4.2: Patient a); (ii) transfer out date when patient leaves the practice; (iii) last data collection date for the practice in CPRD; (iv) death date (Figure 4.2: Patient b); (v) end of

study period (Figure 4.2: Patient c); (vi) or the ADE outcome event, whichever occurred first [Table 4.2; Figure 4.1].

4.3.4 Exposure

A patient was considered to be exposed to hazardous prescribing when they received the hazardous prescription. The numerator consisted of the number of at-risk patients that received the hazardous prescription, here the NSAID. If a patient fulfilled the requirements for the numerator and the denominator, hence a prescription of an OAC and an NSAID, the patient is considered exposed to the HPE. If a patient met the requirements for the denominator but not the numerator, they belonged to the population at risk of but unexposed to the HPE. If the anticoagulated patient did not receive an NSAID during follow-up time, the whole follow-up period was seen as unexposed. NSAIDs are often repeatedly prescribed for short term use. Due to the time-varying nature of NSAID prescriptions, NSAID exposure was used as a time-dependent binary variable. The time periods where patients had a continuous prescription triggering the HPE were defined as exposure periods (including a 30-day grace period after the calculated prescription stop date). Exposure periods are assessed based on prescribing records of NSAIDs and do not include over the counter (OTC) purchases of NSAIDs.

4.3.5 Outcome measures

GI bl as described in the background sections 4.2.1 and 4.2.2. The primary serious harm outcome for this study was GI bleeding events, referred to as serious GI events. Serious GI events were identified in the HES records that excludes minor bleeding events recorded in primary care that did not lead to hospitalisation. Minor bleeding events were not considered relevant for this thesis, as they do not have a large impact on costs (351, 352) and quality of life (353). Serious GI events were identified using ICD-10 codes. The list of ICD-10 codes to identify the serious GI events are reported in Appendix J. Section 4.3.6.1 provides further detail about how this code list was developed. Events were extracted from HES if any of the ICD-10 codes was a primary diagnosis in a hospital episode as done previously in studies investigating bleeding risk in the linked CPRD/HES dataset (354-356). The primary diagnosis describes the main cause that contributed to the hospital stay and is

more likely to be correct compared with using secondary diagnoses (357, 358). From the ONS data, deaths with the primary cause coded as a serious GI event were also included as serious GI events. ONS death records are recovered from death certificates and considered most appropriate when assessing mortality and cause of death (359).

Four of the studies investigating the increased risk of serious GI events associated with NSAIDs in anticoagulated patients in Table 4.1 also estimated the increased risk of major bleeding (307-309, 332). While the impact of NSAIDs on the risk of serious GI events was higher than on major bleedings in anticoagulated patients in Kent et al. (2018) and Schjerning Olsen et al. (2015) and in warfarin users in Dalgaard et al. (2020) and Lamberts et al. (2014), this study investigates major bleeding as a secondary outcome for comparability reasons (307-309, 332). Major bleedings include ICD-10 codes for GI bleeding, intracranial haemorrhage (ICH), respiratory, urinary and rectal bleeding, haemoptysis, and other unspecified bleeding. The risk of stroke and systemic embolism as a potential ADE related to NSAID use were tested as a secondary outcome. OACs are primarily prescribed for stroke prevention (354) and NSAIDs have been associated with an increased cardiovascular risk (360). Death as a competing risk event was also assessed as a secondary outcome. Information on mortality were generated from ONS records.

4.3.6 Preparation for data analysis

Preparing the CPRD data to generate a survival time dataset that could be analysed, was a complex process. Various assumptions were made during the process that are presented in Appendix I. The CPRD, HES and ONS data were provided in separate files. Access to the CPRD data was not limited, but HES and ONS records were only available for patients eligible for linkage and with an OAC prescription in the study period. The process of identifying code lists and on identifying continuous medication use, are described here in more detail.

4.3.6.1 Identifying clinical code lists

Product code lists were required to identify the prescription dates in CPRD that define follow-up and exposure periods. Keywords for drugs were identified from the relevant BNF chapter covering the individual substance names (361) and applied in a search command

for *Stata* developed by Kontopantelis (362) that searches look up files for the keywords [*pcdsearch*]. The keywords are listed in Appendix I. In a second step, the Code browser application provided by the MHRA was used that searched the look up files for substance and product names. The derived code lists were compared with published code lists on the website 'ClinicalCodes.org', an online clinical codes repository introduced by the University of Manchester (363), or those available in other observational studies investigating bleeding event risk. A complete list of all codes identified in any of these lists was generated. The author of this dissertation used her clinical knowledge as a pharmacist to screen code lists and to exclude non-relevant codes.

ICD10-code lists were required for data extraction of diagnoses from HES records and were required to identify outcome events. Published code lists were searched using the same sources already used for the product codes for drugs. A similar search for published code lists on serious GI events was conducted by a different researcher to guarantee that all potential code lists were captured. The summary of all possible codes identified in the literature was discussed in the wider project team of PROTECT and with GPs to verify that the correct diagnoses were covered [Appendix I]. The final list of ICD-10 codes used to identify outcomes in HES records is reported in Appendix J.

Data about relevant covariates were mainly accessed from CPRD, which required Read code lists. Where available, code lists provided by PRIMIS, from the University of Nottingham, were used that were developed for the wider PROTECT project to programme the algorithm to identify patients with HPEs in PINCER. Where these were not available, code lists provided by a research team from the University of Nottingham that worked with a similar cohort in observational data (354) or other published code lists (364) were used. If no published code lists were available a keyword search was conducted in the MHRA code browser. These were screened for non-relevant codes based on the clinical knowledge of the author of this thesis.

4.3.6.2 Identifying periods of continuous medication use

Definitions for follow-up as well as for exposure were based on the assessment of continuous medication use [4.3.3]. Continuous medication use was described as consecutive prescriptions of the same drug class with a grace period of 30 days between calculated stop of the prescription and the start date of the next. Because the stop date was not recorded in CPRD, it was calculated from available records on quantity of units, e.g., tablets or capsules, and the daily dose prescribed. Pye et al. (2018) criticized the lack of guidance on how to process medication data and the authors developed an algorithm guiding how to report the processing of prescription data in CPRD to increase transparency (365). The first step to report was how missing and implausible records were cleaned. For DOACs and NSAIDs recorded daily doses were considered implausible if dose units were not in tablets but in milligrams or millilitres because these generated extreme durations and could not be used in the calculation of the prescription length. Daily doses of zero were also assumed to be implausible. All implausible daily doses were set to missing. Those daily doses missing or set to missing were replaced by the standard dose described in the British National Formulary (BNF) if this did not vary for different indications or formulations. For DOACs standard doses were identified from the BNF and replaced the missing daily doses. NSAIDs with many possible daily doses for the same strength were kept as missing.

Then, stop dates of each prescription were generated by adding the prescription length to the prescription date. The prescription length was calculated by dividing the prescribed quantity by the daily dose. If daily doses were recorded in the CPRD for more than 50% of the prescriptions of a drug group, the prescription length was calculated using the recorded daily dose and the recorded quantity. This was the case for NSAIDs with a missingness of all NSAID substances of 18% and for DOACs with missingness ranging from 22% to 27% depending on the substance.

The use of warfarin or other VKAs was different. In practice, daily doses are adapted according to International Normalized Ratio (INR) measurements (366). Patients can have prescriptions of warfarin with different strength tablets prescribed together, so that the daily dose can easily be adjusted at home without a new prescription being issued if the dose changed. The daily dose was therefore often not recorded. In about 85% of the

prescriptions of VKAs the daily dose was missing. The few daily doses recorded were not assumed to be consistent over the prescription length and were not used to calculate prescription lengths as done for available records for DOACs. For VKAs and NSAIDs missing prescription lengths were replaced by the sample's median prescription length, which was considered most representative of an average user. The median length was generated from time between prescription dates if missingness for daily dose records was ≥50%, as was the case for VKA prescriptions. If missingness of daily doses was <50%, the median of available calculated lengths was used, as it was the case for NSAIDs. Because missing daily doses for DOACs were already replaced by the standard BNF dose, this step was not required for DOACs.

For each prescription, the stop date was generated by adding the prescription length to the date the prescription was issued. Continuous prescription periods started with the issue of the first prescription and ended with the calculated stop date of the last prescription plus the 30-day grace period. The described approach to identify stop dates and then periods of continuous prescriptions was considered reasonable by pharmacists and epidemiologists of the PROTECT team. Further assumptions made were listed in Appendix I.

4.3.7 Confounding variables

This section reports (i) the process used to identify potential confounding variables, (ii) the relations between variables, outcomes and exposure, and (iii) the method used to adjust for covariates.

4.3.7.1 Identifying potential confounding variables

The risk of bleeding from using NSAIDs is highly dependent on a patient's specific comorbidities and concomitant medication (355). The choice to prescribe an NSAID might also be influenced by these comorbidities or medications. Neglecting the relationships between variables can introduce bias, for instance, confounder bias or collider bias (367, 368). A confounder is defined as a variable that has (direct or indirect) causal influence on both the exposure and the outcome. In contrast, a collider is defined as any variable that is causally influenced (direct or indirect) by the exposure and the outcome (368, 369).

Not conditioning on confounders introduces confounder bias and conversely, conditioning on colliders introduces collider bias. Both biases yield distorted estimates of the causal effect between exposure and outcome (367, 370).

A literature search of studies using observational datasets to assess bleeding risks with antithrombotic treatments was conducted to identify covariates used as controls for observable confounders. Details of the review are not reported, but the majority of studies controlled for a mix of cardiovascular risk factors and bleeding risk factors. The cardiovascular risk factors often represented variables of the CHA2DS2-VASc score. The score is calculated from the following parameters: (i) Congestive heart failure, (ii) uncontrolled hypertension, (iii) age over 65 years, (iv) diabetes mellitus, (v) stroke/transient ischaemic attack, (vi) vascular disease, e.g., myocardial infarction, and (vii) sex. The CHA2DS2-VASc score was developed to quantify the risk of stroke in patients with AF to measure the need for antithrombotic therapy and has been validated externally in numerous AF populations (371, 372). In recent years, it was also used to predict stroke risk in non-AF populations (373, 374), such as in acute coronary syndrome (ACS) (375), ACS with diabetes (376), heart failure (377) and interatrial block (378).

To identify bleeding risk, different scores have been developed: HAS-Bled score (379), ATRIA (380), ORBIT (381) and HEMORR$_2$HAGES (382). The HAS-Bled score has proven to be the most accurate to predict clinically relevant bleeding events (383-386). The score has been validated in patients with AF, venous thromboembolism (VTE) (387) and ACS (388). The score is calculated from the following parameters: (i) uncontrolled hypertension defined as a systolic blood pressure above 160 mmHg, (ii) renal disease defined as CKD stage 4 or worse, dialysis or transplant patients, (iii) severe chronic liver disease, (iv) history of stroke, (v) prior major bleeding or predisposition to bleeding, (vi) labile INR with a time in therapeutic range under 60%, (vii) age over 65 years, (viii) medication use predisposing bleeding (antiplatelets, aspirin, corticosteroids and specific antidepressants), and (ix) alcohol dependence. Together with the CHA2DS2-VASc score, the HAS-BLED score is recommended by NICE to make decisions on antithrombotic treatment in AF (41). In addition to treatment decisions regarding the need for OACs, the scores are also used in studies assessing stroke and bleeding events with different antithrombotic treatments

(330, 389-391). Due to the influence on treatment decisions of the two scores, covariates that are part of these scores were also considered potential confounders. A set of variables was created that included all potential confounders used in other observational studies from the review or that were part of the risk scores.

4.3.7.2 Understanding the relations between variables

For each identified variable, the literature was screened to identify if there was evidence for a causal relationship with the exposure or the outcomes. With three GPs, previously involved in this project, the variables from the list of potential confounders from the literature were discussed in a video conference facilitated by the author of this dissertation. The variables were grouped by consensus into categories according to whether each variable (i) was a confounder (affected NSAID exposure and GI bleeding outcome directly), (ii) was a collider (affected simultaneously by NSAID exposure and GI bleeding outcome), (iii) was a GI bleeding risk factor only (affected GI bleeding outcomes), or (iv) had no effecting on bleeding outcomes. The GPs also gave input in the video conference on potentially relevant variables missed in the literature review. Confounders were identified as variables that affect prescribing of NSAIDs and GI bleeding risk. NSAID prescribing can be negatively influenced if GPs are hesitant to prescribe an NSAID because of an existing condition predisposing bleeding. The GI risk factors were conditions indirectly associated with the outcome, e.g., when the condition itself affects bleeding risk through other mediators. The GI risk factors were assumed not to effect NSAID prescriptions but with the potential to increase GI bleeding risk. A list of variables identified in the review and by GPs grouped by their relation to the exposure and the outcome into confounder, collider, GI risk factors and variables with no assumed effect on the outcome or exposure is reported in Table 4.3.

The theoretical associations identified by the GPs were tested in univariate analysis in the dataset. All variables identified as risk factors, confounders or collider had a statistically significant ($p < 0.05$) association on the risk of serious GI events. Those variables with no assumed effect on the outcome did not have a significant effect in the dataset. This additional analysis supported the assessment of the GPs.

INR measurements are recorded in the CPRD but not consistently (392, 393). INRs tend to be measured more regularly for patients with labile INR and are not measured at all in anticoagulated patients treated with DOACs because their treatment dose is not adjusted to the INR measurement in contrast to warfarin doses. INR measurements are used to guide dosing regimens in patients treated with warfarin (41). They are particularly relevant after initiation of OAC treatment, hence after start of follow-up. Because of the inconsistent records, INR was not considered in baseline characteristics. Labile INR is a confounder, and not conditioning on a confounder potentially introduces confounder bias, but since it is only relevant for half of the patients using warfarin, relevance is questionable.

Table 4.3: Relation of potential confounders with NSAID use and serious GI events

Relation	Variables
Confounder	Patient demographics: age
	Comorbidities: alcohol dependence, severe chronic renal disease (CKD stage 4 or worse, dialysis, transplant patients), severe chronic liver disease, uncontrolled blood pressure (>160mmHg), bleeding event (primary care records: serious GI events including ulcer perforation and ulcer bleeding, rectal bleeds, intracranial haemorrhage; secondary care records: major bleeding events as defined in outcome measures), peptic ulcer (excluding haemorrhagic or perforated ulcers), oesophageal varices, anaemia, labile INR
	Medications: antiplatelets, aspirin, antidepressants (SSRI,TCA), corticosteroids
Collider	Gastroprotective agents (proton pump inhibitors, H2-receptor antagonists, misoprostol)
GI bleeding risk factor	Patient factors: gender, socio-economic status (index of multiple deprivation on patient level), ethnicity, smoking, high BMI (obese: BMI >30 kg/m^2)
	Comorbidities: coronary heart disease (heart failure, myocardial infarction, angina), cerebrovascular events (stroke, transient ischaemic attack), peripheral artery disease, venous thrombo-embolism (pulmonary embolism or deep vein thrombosis), valvular heart disease, hypertension (controlled), valvular heart disease, diabetes, COPD, cancer, adverse GI events (dyspepsia, heartburn), GI inflammation (gastritis, duodenitis and oesophagitis), helicobacter pylori infection, anti-epileptic drugs (phenytoin or carbamazepine)
No effect	Patient factors: diet
	Medications: statins, antibiotics (macrolides), nitrates

BMI: body mass index; COPD: chronic obstructive pulmonary disease; CKD: chronic kidney disease; GI: gastro-intestinal; GPs: general practitioner; INR: international normalized ratio; mmHg: millimetre of mercury; SSRI: selective serotonin reuptake inhibitor; TCA: tricyclic antidepressant

A causal diagram, a directed acyclic graph (DAG), was used to get an overview of relations between the variables, exposure and outcome (potential confounders and colliders). DAGs can guide the process to obtain an unbiased estimate on the causal inference between exposure and the outcome, by providing a formal mathematical framework to describe the relations between relevant variables (367, 368, 394, 395). Identifying these relations is paramount to understand how statistical conditioning on these variables would impact the effect estimates. The software DAGitty was used to draw the DAG that was used to guide discussions on relations between the variables with the three GPs (396). The DAG is presented in Appendix G [Figure G.1]. Due to the high number of confounders and the multiple relations between the variables and the exposure or outcome but also among the variables the DAG became very complex.

4.3.7.3 Covariate adjustment

The complexity of the DAG made it challenging to identify a sufficient set of covariates for covariate adjustment in a regression model without including mediators with the risk of over adjustment. Methods not requiring to specify a regression model to relate exposure and covariates to the outcome are propensity score based balancing methods (397). The choice of regression model can influence estimates using covariate adjustment, which contains the risk to be tempted to use a model with results closer to *a priori* expected results (398). Propensity score related methods are an alternative, when regression analysis is not appropriate, e.g., when the difference between distributions of the covariates varies substantially (397-400). Propensity score methods aim to achieve balance between exposed and unexposed group by stratification, inverse probability weighting (IPTW) or propensity score matching (PSM). These methods are described in detail in the literature, for example by Austin et al. (2011) (260), and is not described in this chapter.

Nearest neighbour PSM was chosen as the base case analysis. Greedy matching was used without a calliper. The matched cohort was used in the statistical analysis [4.3.8]. To predict the scalar propensity score from the covariate vector, a probit model was used, regressing exposure status on observed covariates.

Paramount for successful PSM is the selection of the correct variables to predict the propensity score (401-403). Selecting all variables associated with the outcome (risk factors) yielded best balancing scores compared with including only confounders (variables associated with outcome and exposure) or adding variables associated with the treatment alone (403). Consequently, all GI risk factors and confounders identified in Table 4.3 were included in the propensity score model. All variables are required to be baseline variables because only these are predictive of NSAID assignment.

Different methods are available to test if the matching process was successful and yields balanced cohorts for exposed and unexposed patient groups (404). Balance of the matched cohort in this chapter was tested firstly by comparing the standardised difference between treated and untreated groups. The standardised difference of the means of continuous or binary variables and the percentages in each category of non-binary categorical variables, was used as an indicator to identify variables not evenly distributed between treatment groups. The differences in means are reported as units of the pooled standard deviation. A variable with a standardised difference higher than 0.10 was considered not adequately distributed (260, 404). The `pstest` command in *Stata 2015* was used to generate the mean standardised differences (reported in %) and to visualise the variables with a standardised difference of more than 10% (402). The advantage of the standardised difference over t-test statistics was that it is not dependent on the sample size (260, 401, 405). In addition to comparing the standardised difference, Austin et al. (2011) recommends graphical methods. A density plot was used to identify the distribution of the propensity score among NSAID and never NSAID users (`twoway(histogram)` command in *Stata 2015*). The area of common support should include all patients in a well-balanced cohort.

4.3.8 Statistical analysis

Baseline characteristics were reported before and after matching (i) for the population exposed to NSAIDs at least once during follow-up (NSAID user), and (ii) for the population never exposed during follow-up (Never NSAID user). Details on assumptions made to extract covariates can be found in Appendix G [Table G.1]. Characteristics are reported as

mean with standard deviations for continuous measures, and as percentages for categorical measures.

After careful consideration of potential model approaches, the effect of NSAID use on the ADEs (serious GI events, stroke, systemic embolism, major bleeding) and death was estimated using a cause-specific Cox proportional hazard model. After the matching process created a balanced cohort and the balancing tests did not indicate otherwise, it was assumed that treatment assignment (NSAID exposure) was independent of the baseline characteristics. Consequently, no further adjustment of covariates was included in the Cox proportional hazard model. The median follow-up time of patients was 0.6 person years and it was assumed that baseline characteristics were unlikely to change in this short period of time.

This was applicable to all outcomes under the assumption that variables chosen included relevant variables for the non-primary outcomes as well. Multiple studies investigating major bleeding risk factors did not report additional risk factors that were not balanced in the matched cohort (406, 407). For stroke and systemic embolism, no risk factors were reported in various studies investigating risk factors in anticoagulated patients or in AF that were not considered for the primary outcome (408-410). However, the consulted GPs identified slightly different relationships of potential confounders with stroke compared with the primary outcome. Hypertension and statin use were not identified as direct risk factors for serious GI events [Appendix G], but they are considered direct risk factors for stroke and systemic embolism. The GPs considered hypertension and statin use to only indirectly impact the risk of serious GI events. The consulted GPs also considered hypertension and statin use not to be associated with NSAID use, hence they are not considered confounders for the estimation of stroke or serious GI event risks [Appendix G]. Because only direct risk factors for serious GI events, such as previous bleeding and uncontrolled blood pressure, were included in the propensity score assessment, the additional stroke risk factors were not included in the propensity score. However, the matching process generated cohorts that were balanced on these two risk factors as well.

The model was able to incorporate censored data if censoring occurred at random, hence independent of ADEs (411). Censoring due to loss of follow-up for other reasons, either the patients transferred out of the practice, data collection of the practice stopped or the end of the study period were considered non-informative. This assumption was used in other studies in the CPRD analysing bleeding events in OAC cohorts using Cox proportional hazard models (354, 355, 412). The Cox proportional hazard model also allows to encompass the time-varying nature of the internal covariate of NSAID exposure (413). Death and bleeding events were not considered to be independent. A patient who died cannot have the failure event in the future. The competing risk event of death therefore needed to be accounted for. By censoring the competing risk event, the model estimates the instantaneous risk of failure given that failure from any cause, competing or failure event, has not happened yet. The interpretation of these cause-specific hazards should always take into account the cause-specific hazard of the competing risk event (414). Other methods that account for competing risk events, such as the Fine and Gray method, could not encompass time-varying covariates and were therefore not considered (414, 415). The cause-specific hazard function denoted as (416)

$$\text{Equation 3: } \mathrm{H(t)} = \lim_{\Delta \mathrm{t} \to 0} \frac{Pr(t < T \leq t + \Delta t, K = k | T > t}{\Delta t}$$

where the instantaneous risk $\mathrm{H(t)}$ of the occurrence of event *K* (where k is one of the competing events here serious GI event or death) given that any *k*th event did not happen until time *t*. For each event *k* a separate proportional hazard model is used (414). This study was interested in GI bleeding risk and deaths in association with NSAID use. The cause-specific hazard for serious GI events at time *t* provided the instantaneous rate of serious GI events in patients with no serious GI event conditional on survival until time *t*. The cause-specific hazard for death at time *t* provided the instantaneous rate of death given that death or a serious GI event have not happened yet (416). The proportional hazard model for the cause-specific hazards for each event *k* (1= serious GI event; 2= death) denoted as

$$\text{Equation 4: } H(t|X(s), s \leq t) = H_{oj}(t) exp^{\beta_k X(t)}$$

was found most appropriate to estimate the regression coefficient β_k of the time-varying covariate effects of NSAID use. The proportional hazard assumption was tested using Schoenfeld residuals (417). The proportional hazard assumption was considered to hold if the null hypothesis of zero slope of the scaled Schoenfeld residuals regressed over time was not rejected. The proportional hazard assumption was also tested graphically. A log-log plot of survival versus analysis time was plotted that shows parallel lines if the proportional hazard assumption is not violated.

The proportional hazard model was not used to obtain cumulative incidence of serious GI events or deaths because the application of the exposure as a time-varying covariate and the competing risk of death did not allow to make inference from the cause-specific hazard function on the cumulative incidence function (414, 418, 419). The results of the Cox proportional hazard model were reported as HR with 95% CI. *Stata 2015* was used to perform the analysis.

4.3.9 Sensitivity analysis

Sensitivity analysis tested the assumptions that were made during the data preparation, with exclusion criteria, on conditioning or confounding. Primarily the sensitivity analysis aimed to test robustness of the increased risk estimated for the primary outcome of serious GI events. Because during the analysis of the secondary outcomes stroke risk was found to be substantially increased with NSAID exposure, robustness of this outcome was also tested.

Robustness of assumptions made during the data preparation and exclusion criteria

In sensitivity analysis, a 60-day grace period during continuous treatment use was explored. Additionally, various changes to the washout periods were investigated for the exposure and the serious GI event at baseline. A shortened washout period of 30 days for NSAID use was tested to explore if a smaller washout period influenced the results. 30 days was the median prescription length of NSAIDs. An extended washout of six months was also tested. A serious GI event washout window was also considered to explore if excluding patients with a hospitalisation for a serious GI event prior to the index date, would impact the HR.

Patients with a serious GI event ever before index were excluded. In the sensitivity analysis for stroke events, previous stroke events were used accordingly.

Robustness of assumptions on conditioning on confounding variables

For the base case analysis, all variables affecting the serious GI event outcome were considered in the propensity score model. However, it was suggested that including a collider in the propensity score model can introduce a small bias (420). The propensity score model was therefore rerun not including the collider of baseline GPA use.

Matching on the propensity score was chosen as the balancing method for the base case (PSM). In the sensitivity analysis, IPTW on the propensity score was investigated. The IPTW methods is of particular interest because it preserves the original sample size that could be useful in rare events, such as serious GI events. Because the probability of treatment assignment was low, resulting in low propensity score, the use of stabilised weights has been recommended (421, 422).

NSAID use was incorporated as a time-varying exposure in the Cox proportional hazard model. The exposure could be influenced not only by the baseline characteristics but also by changes in confounders during follow-up. In the base case analysis, it was assumed that baseline characteristics do not change over time and no further adjustment for confounders was required in the balanced population. In the sensitivity analysis, the proportional hazard model was adjusted for confounders identified in Table 4.3, ethnicity and deprivation level and the variables were included in the regression model as time-varying variables. For this analysis the covariates assessed at the index date were updated before each change in the exposure status. In the respective sensitivity analysis for stroke variables in the CHA2DS2-VASc stroke risk score [4.3.7.1], ethnicity and deprivation level were used.

For the variables smoking (1%), body mass index (BMI) (51%) and ethnicity (21%), not all patients had records. In the base case, missing records were considered as an additional category and the missing category was included in the matching process. Even though

these patients are potentially not missing at random, a sensitivity analysis was run dropping patients with missing smoking, BMI or ethnicity record.

E-value

Unmeasured confounding is a common problem in observational studies. While it is difficult to quantify the bias associated with unmeasured confounding, there are tools available to test if the observed effect is likely be explained away by unmeasured confounding (423, 424). In this study, the e-value was calculated. E-values are used to estimate the minimal strength of association an unmeasured confounder must have with the outcome and the exposure to be able to explain away the observed effect (424). For the mean HR and the lower range of the confidence interval for serious GI events and stroke, the e-value was calculated. For risk ratios (RR) greater than one, the e-value is calculated as follows (424)

$$\text{Equation 5: } E-value = RR + \sqrt{\left(RRx(RR-1)\right)}.$$

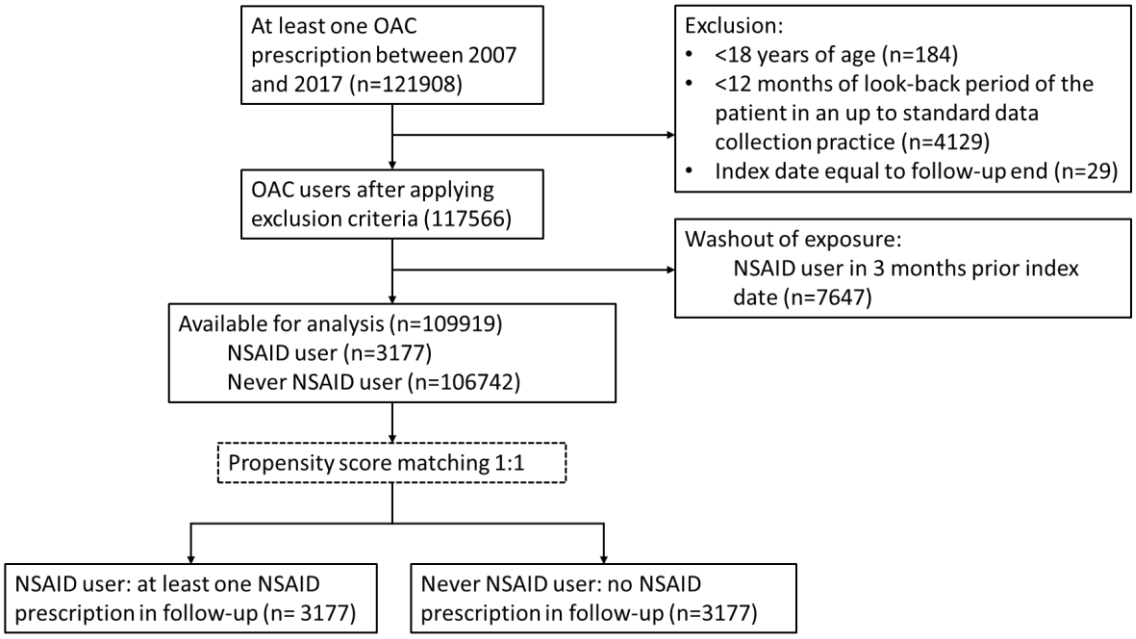
To be able to interpret the e-value, it was compared to the impact of known risk factors for the two outcomes as recommended in the published literature (423, 425). If known confounders conditional on exposure have a smaller impact than the estimated e-value, it is unlikely that an unmeasured confounder has a greater influence on the outcome (423). Hence, the risk that the observed effect can be nullified by an unmeasured confounder is low. The impact of variables on the risk of the outcome conditional on exposure was assessed either as a confounder of the effect of NSAIDs on serious GI events in Table 4.3 or for the stroke outcome variables in the CHA2DS2-VASc. The stroke risk score was described in section 4.3.7.1.

4.4 Results

4.4.1 Descriptive statistics

From CPRD, 121908 patients were identified with at least one OAC prescription during the study period (2007 to 2017) that were eligible for linkage. Figure 4.3 presents a flow diagram to illustrate how the final sample was derived. There were 117566 patients eligible for the cohort after the exclusion criteria were applied. The new user design requires exclusion of patients with NSAID use three months prior the index date. Afterwards 109861 patients were available for the statistical analysis. Of these patients with an OAC prescription and at risk of a hazardous prescription of a NSAID, 2.99% (n=3177) had at least one NSAID prescription.

Figure 4.3: Flow of the patients from the dataset included in the analysis of the primary outcome



Baseline characteristics are reported in Table 4.4 for the cohort never exposed to NSAIDs (labelled as 'Never NSAID user'), and the population exposed to NSAIDs at least once over the study period (labelled as 'NSAID user'). The unmatched cohorts had a mean standardised difference of 4.0% and proportions of individual characteristics were similar.

Table 4.4: Baseline characteristics of patients in the cohort before matching and in the propensity score matched cohort

Variable	Before matching				After matching			
	Never NSAID user (n=106742)[a]		NSAID user (n=3177)[b]		Never NSAID user (n=3177)[a]		NSAID user (n=3177)[b]	
Propensity score (mean, SD)	0.029	(±0.009)	0.031	(±0.009)	0.031	(±0.009)	0.031	(±0.009)
Standardised difference (mean)	4.0				2.0			
Age (mean and SD)	72	(14)	70	(14)	70	(14)	70	(14)
Women	48413	(45%)	1311	(41%)	1331	(42%)	1311	(41%)
Ethnicity								
White	82260	(77%)	2456	(77%)	2495	(79%)	2456	(77%)
Other	496	(0%)	15	(0%)	16	(1%)	15	(0%)
Asian	940	(1%)	33	(1%)	25	(1%)	33	(1%)
Black	711	(1%)	21	(1%)	12	(0%)	21	(1%)
Missing	22335	(21%)	652	(21%)	629	(20%)	652	(21%)
Deprivation index (IMD)								
1 (least deprived)	25332	(24%)	671	(21%)	658	(21%)	671	(21%)
2	24545	(23%)	725	(23%)	736	(23%)	725	(23%)
3	23140	(22%)	686	(22%)	678	(21%)	686	(22%)
4	18647	(17%)	590	(19%)	614	(19%)	590	(19%)
5 (most deprived)	15014	(14%)	503	(16%)	491	(15%)	503	(16%)
Missing	64	(0%)	N/A	(<0%)	0	(<0%)	N/A	(<0%)
Smoking status								
Current smoker	15379	(14%)	504	(16%)	540	(17%)	504	(16%)
Ex-smoker	54463	(51%)	1677	(53%)	1704	(54%)	1677	(53%)
Missing	941	(1%)	25	(1%)	29	(1%)	25	(1%)
Never smoker	35959	(34%)	971	(31%)	904	(28%)	971	(31%)
Blood pressure control								
Uncontrolled blood pressure	4529	(4%)	128	(4%)	118	(4%)	128	(4%)
Controlled blood pressure	87439	(82%)	2587	(81%)	2623	(83%)	2587	(81%)
Missing	14774	(14%)	462	(15%)	436	(14%)	462	(15%)
BMI (mean and SD)	29	(6)	30	(6)	30	(6)	30	(6)

Variable	Before matching				After matching			
	Never NSAID user (n=106742)[a]		NSAID user (n=3177)[b]		Never NSAID user (n=3177)[a]		NSAID user (n=3177)[b]	
BMI by category								
17.5-22.49	6342	(6%)	136	(4%)	152	(5%)	136	(4%)
22.5-24.9	7795	(7%)	178	(6%)	217	(7%)	178	(6%)
25.0-29.9	18959	(18%)	573	(18%)	538	(17%)	573	(18%)
30.0-34.9	11443	(11%)	437	(14%)	427	(13%)	437	(14%)
35.0-39.9	4736	(4%)	186	(6%)	178	(6%)	186	(6%)
≥40.0	2859	(3%)	121	(4%)	117	(4%)	121	(4%)
Missing BMI	54608	(51%)	1546	(49%)	1548	(49%)	1546	(49%)
Concomitant drugs (6 months before index date)								
Antiplatelet	7954	(7%)	188	(6%)	150	(5%)	188	(6%)
Aspirin	31482	(29%)	937	(29%)	882	(28%)	937	(29%)
Gastroprotective agent	34796	(33%)	1147	(36%)	1068	(34%)	1147	(36%)
Antidepressant	16280	(15%)	570	(18%)	574	(18%)	570	(18%)
Corticosteroids	9667	(9%)	314	(10%)	288	(9%)	314	(10%)
Anti-epileptic drug	1256	(<1%)	42	(<1%)	30	(<1%)	42	(<1%)
Statin	48811	(46%)	1433	(45%)	1427	(45%)	1433	(45%)
Comorbidities (before index date)								
Peptic ulcer	5452	(5%)	126	(4%)	131	(4%)	126	(4%)
Any adverse GI event	23351	(22%)	772	(24%)	716	(23%)	772	(24%)
GI inflammation	14719	(14%)	473	(15%)	467	(15%)	473	(15%)
Alcoholism	2286	(2%)	79	(2%)	86	(3%)	79	(2%)
Anaemia	13326	(12%)	352	(11%)	316	(10%)	352	(11%)
Cancer	12928	(12%)	365	(11%)	342	(11%)	365	(11%)
Any bleeding (primary care records)	13505	(13%)	426	(13%)	391	(12%)	426	(13%)
GI bleeding	9731	(9%)	304	(10%)	284	(9%)	304	(10%)
Intracranial bleeding	915	(1%)	19	(1%)	29	(1%)	19	(1%)
Rectal bleeding	2876	(3%)	104	(3%)	79	(3%)	104	(3%)
Coronary heart disease	26726	(25%)	836	(26%)	778	(24%)	836	(26%)

Variable	Before matching				After matching			
	Never NSAID user (n=106742)[a]		NSAID user (n=3177)[b]		Never NSAID user (n=3177)[a]		NSAID user (n=3177)[b]	
COPD	9060	(8%)	278	(9%)	244	(8%)	278	(9%)
Diabetes	17124	(16%)	516	(16%)	496	(16%)	516	(16%)
Renal disease (severe)	3455	(3%)	89	(3%)	87	(3%)	89	(3%)
Liver disease (severe)	645	(1%)	14	(0%)	20	(1%)	14	(0%)
Peripheral vascular disease	3200	(3%)	88	(3%)	79	(2%)	88	(3%)
Atrial fibrillation	57356	(54%)	1660	(52%)	1658	(52%)	1660	(52%)
Heart failure	14035	(13%)	396	(12%)	363	(11%)	396	(12%)
Hypertension	56139	(53%)	1674	(53%)	1673	(53%)	1674	(53%)
Stroke	17836	(17%)	451	(14%)	432	(14%)	451	(14%)
Valvular heart disease	11407	(11%)	257	(8%)	235	(7%)	257	(8%)
Venous thromboembolism	27531	(26%)	850	(27%)	852	(27%)	850	(27%)

Data are presented as mean (standard deviation) for continuous measures, and n (%) for categorical measures; GI varices are not reported because the percentage was 0 in all groups; [a]patients never exposed to NSAIDs and OACs at the same time (Never NSAID user); [b]patients exposed to NSAIDs at least once (NSAID user); BMI: body mass index; COPD: chronic obstructive pulmonary disease; GI: gastro-intestinal; IMD: index of multiple deprivation; NSAID: non-steroidal anti-inflammatory drug

The matched cohort consisted of 3177 NSAID users and 3177 never NSAID users [Table 4.4]. The tests conducted to test for balance between NSAID users and never NSAID users after matching indicated that the PSM was successful. The mean propensity score was 0.031 (SD ±0.09) in both groups. The histogram comparing the propensity score distribution in the two groups demonstrated a comparable distribution and had the same area of common support reported in Appendix H [Figure H.1]. The propensity score matched cohort in Table 4.4 showed a higher level of balance indicated by the lower mean standardised difference in the cohort after matching. The mean standardised difference was reduced from 4.0% to 2.0%. The mean standardised difference of individual estimates is reported in Appendix H [Figure H.2]. After matching, none of the observed variables had a standardised difference greater than 10% indicating balance between NSAID users and non NSAID users. The matched cohort had a mean age of 70 (SD±14) years of age, and the majority of patients was male. Missingness of deprivation indices, smoking status and blood pressure measurements was low, but half of BMI records were missing.

4.4.2 Base case analysis

The incidence of outcomes in the matched cohort for person time spent (in person years) with NSAID exposure (NSAID use) and without (no NSAID use) is reported in Table 4.5.

Table 4.5: Incidence rates, IRR and cause-specific HR of serious GI events, major bleeding, stroke, systemic embolism and all-cause death

Outcome	Person time with		Incidence rate [a]		IRR [b]	HR [b]
	NSAID	No NSAID	NSAID use	No NSAID use		
Serious GI event	635	7734	20.49 (11.90; 35.28)	6.98 (5.35; 9.12)	2.93 (1.47; 5.45)	2.96 (1.60; 5.46)
Major bleeding	621	7461	45.06 (31.11; 65.25)	16.08 (13.45; 19.24)	2.80 (1.79; 4.26)	2.86 (1.89; 4.33)
Stroke	634	7641	20.52 (11.92; 35.33)	8.38 (6.56; 10.70)	2.45 (1.24; 4.49)	2.48 (1.36; 4.53)
Systemic embolism	633	7644	3.16 (0.79; 12.64)	1.70 (0.99; 2.93)	1.86 (0.20; 8.21)	1.62 (0.36; 7.22)
All-cause death	626	7575	92.36 (71.24; 119.73)	26.76 (23.28; 30.76)	3.45 (2.52; 4.66)	3.40 (2.52; 4.58)

[a]Incidence rate per 1000 person years (95% confidence intervals); [b]unadjusted IRR and HR reported for NSAID users relative to no NSAID users with 95% confidence intervals; HR: hazard ratio; IRR: incidence risk ratio

The incidence rate for serious GI events was 6.98 (95% CI 5.35 to 9.12) per 1000 person years during person time without NSAID use and 20.49 (95% CI 11.90 to 35.28) per 1000 person years during person time with NSAID use. This resulted in an incidence risk ratio of 2.93 (95% CI 1.47 to 5.45) for serious GI events when exposed to NSAIDs during follow-up. The regression analysis identified a HR in the presence of the HPE of 2.96 (95% CI 1.60 to 5.49) for serious GI events. Testing the proportional hazard function showed no indication that the assumption might have been violated. Testing for non-zero-slope using Schoenfeld residuals was significant (p-value 0.67) and plotting the hazards of exposed and unexposed patients on a logarithmic scale showed no significant violation of the proportional hazard function [Appendix H: Figure H.3].

The incidence rates of major bleeds were higher compared to those of GI bleeds, with 45.06 (95% CI 31.11 to 65.25) and 16.08 (95% CI 13.45 to 19.24) per 1000 person years during NSAID use and no NSAID use, respectively. The HR comparing NSAID use and no NSAID use was slightly smaller for the risk of major bleeding compared with serious GI events. NSAID use was also associated with an increased risk of stroke. No significant association with NSAID use was found for the rarer event of systemic embolism. The tests for proportional hazards did not indicate any violations of the assumptions for the secondary outcomes [Appendix H].

4.4.3 Sensitivity analysis

Results of the sensitivity analysis for the primary outcome and the secondary stroke outcome are presented in Table 4.6. With extending the assessment window for NSAID washout as part of the new user design of the study, the sample size decreased. The number of patients in the total cohort for serious GI outcomes was 11448 patients with an assessment window of one month and decreased to 5298 patients with the six months washout window. The confidence intervals became larger with increasing the washout window. Excluding patients with serious GI events prior the index date resulted in a larger HR and a larger confidence interval compared to the base case without this exclusion criterion. Excluding patients with missing records for BMI, ethnicity or smoking status reduced the cohort size by more than half and resulted in a smaller and non-significant HR. The same effects were observed for the stroke outcome.

Table 4.6 Results of the sensitivity analysis testing the impact of assumptions on the number of events, the follow-up time and HRs for serious GI events and stroke

Sensitivity analysis (n=number of patients)	Number of events		Person time, years		HR (95% CI)[a,b]
	NSAID use	No NSAID use	NSAID use	No NSAID use	base no NSAID use
Serious GI events					
Base case (n=6354)	13	54	635	7734	2.96 (1.60; 5.46)
Assumptions on data preparation					
60-day grace period (n=9824)	26	135	1509	18323	2.53 (1.66; 3.86)
NSAID washout assessment window 6 months (n=5308)	13	46	496	6755	3.81 (2.04; 7.09)
NSAID washout assessment window 1 months (n=11460)	24	91	1190	11632	2.40 (1.52; 3.78)
Excluding patients with serious GI events (n=6180)	13	49	620	7645	3.37 (1.81; 6.25)
Assumptions around confounding variables					
Stabilised IPTW instead of PSM (n=109894)	16	1067	626	103975	2.47 (1.40; 4.34)
GPA not used as variable in PSM (n=6354)	13	50	635	7761	3.29 (1.77; 6.11)
Cox regression with time varying confounder (n=6354)	13	54	635	7734	3.00 (1.62; 5.59)[c]
Excluding patients with missing records (n=2610)	4	20	247	2854	2.05 (0.70; 6.00)

Sensitivity analysis (n=number of patients)	Number of events		Person time, years		HR (95% CI)[a,b]
	NSAID use	No NSAID use	NSAID use	No NSAID use	base no NSAID use
Stroke					
Base case (n=6342)	13	64	634	7641	2.48 (1.36; 4.53)
Assumptions on data preparation					
60-day grace period (n=9796)	21	157	1509	18337	1.79 (1.13; 2.84)
NSAID washout assessment window 6 months (n=5298)	11	49	495	6739	3.11 (1.61; 6.01)
NSAID washout assessment window 1 months (n=11448)	18	106	1192	11708	1.67 (1.00; 2.76)
Excluding patients with stroke (n=5450)	6	34	539	6378	2.33 (0.97; 5.59)
Assumptions around confounding variables					
Stabilised IPTW instead of PSM (n=109894)	16	1570	627	103779	1.71 (0.95; 3.08)
GPA not used as variable in PSM (n=6342)	13	60	634	7704	2.61 (1.42; 4.78)
Cox regression with time varying confounder (n=6342)	13	64	634	7641	2.75 (1.50; 5.02)[d]
Excluding patients with missing records (n=2610)	2	23	247	3002	1.20 (0.28; 5.10)

[a]Hazard ratio (HR) reported for NSAID users relative to no NSAID users; [b]tests indicated no violation of the proportional hazard assumption. Tests for non-zero slope were not significant (p-value>0.05) and the log-log plots were parallel [Appendix H]. HRs unadjusted for all sensitivity analyses but the time-varying confounder model; [c]adjusted for age, antiplatelet, aspirin, antidepressant and cortisone use, gender, race, deprivation, uncontrolled blood pressure, alcohol dependence, bleeding events, peptic ulcer disease, GI varices, renal and liver disease, anaemia; [d]adjusted for heart failure, stroke/transient ischaemic attack, hypertension, coronary artery disease, diabetes, female gender; PSM: propensity score matching; GPA: gastroprotective agent; IPTW: inverse probability of treatment weighting; NSAID: non-steroidal anti-inflammatory drug

E-values

The calculated e-values and the results on the impact of other variables on the risk of the outcome conditional on NSAID exposure are reported in Appendix H [Table H.1]. For serious GI events an e-value of 5.25 (lower bound: 2.58) was calculated. The observed risk ratio could be explained away by an unmeasured confounder that was associated with both the NSAID prescription and serious GI events by a risk ratio of 5.25-fold each, but weaker confounding could not do so (424). In comparison, the maximal impact a measured confounder had on the outcome conditional on NSAID exposure was 2.25, for the HR for peptic ulcer [Table H.1]. Hence, all measured confounders had a smaller association with NSAID use and serious GI events compared with the mean e-value and its lower bound. For stroke the observed risk ratio could be nullified by an unmeasured confounder that was associated with both the NSAID prescription and the stroke by a risk ratio of 4.23 (lower bound, 2.06). The stroke risk factors tested for their association with NSAID use and the outcome from the CHA2DS2-VASc generated smaller HRs than the mean e-value. The HRs for previous stroke or transient ischaemic attack (HR 3.81), hypertension (HR 2.77), peripheral artery disease (HR 2.24) and coronary heart disease (HR 2.46) had an association with the outcome conditional on NSAID use that was larger than the lower bound of the e-value.

4.5 Discussion

4.5.1 Principal findings

This is the first study presenting methods to identify the link between harm from ADEs associated with NSAID use in anticoagulated patients in the UK. Concomitant NSAID and OAC use was associated with an increased risk of serious GI events (HR 2.96, 95% CI 1.60 to 5.46), major bleeding (HR 2.86, 95% CI 1.89 to 4.33), stroke (HR 2.48, 95% CI 1.36 to 4.53) in a population of anticoagulated patients with NSAIDs compared to anticoagulated patients without concomitant NSAID treatment. The association between NSAID use and an increase risk of systemic embolism was not significant (HR 1.62, 95% CI 0.36 to 7.22). The analysis showed that the methods were robust to various changes in assumptions

tested in the sensitivity analysis. The HPE of NSAID use in anticoagulated patients was found to have substantial impact on bleeding events and other ADEs.

4.5.2 Comparison with published literature

In the background section of this chapter, studies investigating the impact of NSAIDs on ADE risks were briefly introduced to show the novelty of the cohort study in this dissertation [4.2]. This section aims to compare not only the investigated cohorts as done in section 4.2 but also the methods and results of these studies.

Three studies investigated the impact of NSAIDs on both the risk of serious GI events, and stroke/systemic embolism (307-309). The study by Dalgaard et al. (2020) and by Kent et al. (2018) were subgroup analyses of RCTs comparing DOACs with warfarin, and the study by Lamberts et al. (2014) used routinely collected data from a Danish registry. Dalgaard et al. (2020) analysed a subgroup of the ARISTOTLE trial that compared warfarin and apixaban use in AF, while Kent et al. (2018) looked at a subgroup of the RE-LY trial that investigated warfarin and two doses of dabigatran. Both were multinational trials with similar population groups and population size. Both studies used Cox proportional hazard models with time-varying NSAID exposure as was done in this chapter. In the ARISTOTLE and RE-LY trial, 2185 (13.2%) and 2279 (12.6%) patients on OAC treatment were NSAID users, respectively. While the RE-LY trial comparing dabigatran and warfarin users identified a significant increase in serious GI events (HR 1.81, 95% CI 1.35 to 2.43) in the subgroup analysis of NSAID users during the trial and non NSAID users (308), the ARISTOTLE trial comparing apixaban and warfarin did not identify a significant increase in the risk of serious GI events (HR 1.08, 95% CI 0.64 to 1.82) (309). Stroke events in the RE-LY trial were also significantly increased with NSAID use (HR 1.55, 95% CI 1.11 to 2.16) as was systemic embolism (HR 2.43, 95% CI 1.08 to 5.46). Dalgaard et al. (2020) investigated the combined risk increase associated with NSAIDs of stroke/systemic embolism but did not report significant results.

The results of our study with regards to serious GI event and stroke risks are in line with the estimates from the RE-LY trial with the difference that the analysis conducted in this

study included all patients with an OAC and applied less exclusion criteria. The ARISTOTLE trial subgroup analysis did not identify a significant increase in bleeding event risk or stroke/systemic embolism risk and also the point estimates showed only a small increase in the hazard rate for patients not reporting a concomitant prescription of an NSAID. The analysis by Dalgaard et al. (2020) also used a new user design and excluded NSAID users at baseline as was done in this chapter. Person years of NSAID users with actual NSAID exposure was 1898 person years. Hence, almost three times more than in this study even though the number of patients with at least one NSAID during the trial was smaller. However, no information was available from trial related publications (309, 326) or from the trial registration website 'clinicaltrials.gov' on how the concomitant medication use was assessed during the trial. The trial protocol only planned assessment of co-medication at baseline not during the trial (426). An unprecise definition of exposure to NSAIDs could have diluted the effect of the NSAID in ARISTOTLE. The exposure time with NSAIDs was much higher than in this study. While the NSAID prescribing rate could have been much higher in the ARISTOTLE trial, compared with the CPRD cohort, it is more likely that the exposure time in Dalgaard et al. (2020) overestimated the NSAID exposure.

Kent et al. (2018) was the only study reporting a significant association of NSAIDs and the increased risk of systemic embolism. Despite the slightly larger cohort of NSAID users in this chapter compared with Kent et al. (2018), the analysis did not yield significant results. There was no information reported on person years with actual NSAID exposure among the patients with at least one NSAID during follow-up in Kent et al. (2018). This could have been larger despite the smaller cohort size. The authors reported that NSAID use was assessed at regular patient meetings. 'Follow-up visits occurred 14 days after randomization, at 1 and 3 months, every 3 months thereafter in the first year, and then every 4 months until the study ended' (308). There was no evidence, however, on what exactly was asked at these meetings or what questionnaires were filled in. The study also did not exclude patients with baseline NSAID use, which could have resulted in selection bias. Differences among the study results could be a result of the vague definitions of how NSAID exposure was identified in the subgroup analyses of the trials.

The third study was a retrospective cohort study by Lamberts et al. (2014) that used the Danish registry dataset to identify the effect of NSAID use on serious GI events, major bleeding and stroke/systemic embolism in an AF population treated with warfarin or phenprocoumon (307). NSAID exposure was identified as prescriptions recorded in the dataset for NSAIDS and coxibes with 4897 patient years with NSAID use. Person time with NSAID exposure was much lower in this chapter with 635 person years. Lamberts et al. (2014) reported an adjusted HR of 3.54 (95% CI 3.29 to 3.82) for serious GI events. This estimate is slightly higher compared with the HR of 2.96 (95% CI 1.60 to 5.46) identified in this study. For major bleedings a HR of 2.96 (95% CI 2.64 to 3.31) was identified in Lamberts et al. (2014) (307).

The authors also used NSAID use as a time-varying exposure in a proportional hazard model but excluded patients with NSAID use only 30 days before baseline. When patients with NSAID use only one month before the index date, instead of the three months in the base case analysis in this chapter, were excluded from the analysis, the HR was smaller (HR 2.4, 95% CI 1.52 to 3.78). In Lamberts et al. (2014), the covariates included covered a whole range of variables that we identified as mediators in the DAG, such as hypertension, chronic renal failure, liver failure, previous stroke, history of alcohol misuse, previous bleeding event, heart failure, diabetes, previous embolism or vascular disease. This could have resulted in overfitting of the regression model, which could contribute to the different results. The authors did not adjust for gastroprotective agents. Gastroprotective agents are a collider because both NSAID use and serious GI events are a cause for gastroprotective agent prescriptions. Adjusting for colliders can distort the observed relationship because it opens the collider path that is otherwise blocked. Not adjusting for the collider is therefore the correct thing to do, but this can be problematic if the cohorts are systematically different in the proportion of patients with this variable at baseline. The baseline characteristics in Lamberts et al. (2014) cohorts with and without NSAIDs differed by more than 5% in the proportion of patients with gastroprotective agents. Standardised differences were not reported. This can result in selection bias. Patients with NSAIDs are more likely to be prescribed the gastroprotective agents that reduce the risk of serious GI events. In the matched cohort in this chapter, NSAID users and never users only differed by 2%, reducing the selection bias caused by differences in baseline characteristics.

The other outcome assessed in Lamberts et al. (2014) was the impact of NSAIDs on stroke/systemic embolism events. NSAID use was significantly associated with a risk increase (HR 1.67, 95% CI 1.41 to 1.98). The risk increase of systemic embolism associated with NSAIDs was much smaller (HR 1.62, 95% CI 0.36 to 7.22) than that of stroke in the study in this chapter (HR 2.48, 95% CI 1.36 to 4.53). A combined outcome would potentially result in similar results.

Overall, this study adds to the literature on NSAID use in OAC populations and supports the assumptions that NSAIDs increase the risk of not only serious GI events but also of stroke in anticoagulated patients. The impact of NSAIDs on systemic embolism was often only estimated as a combined outcome with stroke and did not yield significant results similar to the cohort study reported in this chapter. This study was the first UK study investigating the increased risk of NSAID use on ADEs in anticoagulated patients and it can be argued that the study design is more robust than that of studies in the existing literature. The exposure periods are well defined and based on prescribing records and not on patient self-report as in the trial subgroup analyses by Kent et al. (2018) and Dalgaard et al. (2020). The detailed discussions with clinicians around confounders and colliders and the resulting propensity score matched cohort enabled an analysis of two cohorts that did not differ systematically in the observed baseline characteristics. The study design in this study was therefore considered stronger than the regression adjustment for mediators and confounders in Lamberts et al. (2014).

4.5.3 Implications of the study design

A key advantage of this study was the use of a clear definition of the HPE type. In Chapter Two [2.1], the challenges of identifying HPEs without clear definitions were discussed. In this study, objective criteria could be used that did not require any subjective assessment what situations qualify as an HPE. The complicated and time-consuming process of identifying whether the ADE was actually caused by the HPE (causality assessment) as well as assessing whether the ADE would not have occurred without the NSAID prescription (preventability assessment) was not feasible as part of this programme of work or within the dataset. Challenges of these methods were described in detail in Chapter Two [2.4].

The data, available from the datasets, would potentially not be sufficient to test causality or preventability because individual medical record charts from the hospital were not available for review. It would also not be feasible to do these assessments for the number of events identified in the datasets. It was assumed that by limiting the harm outcomes to those potentially associated with the HPE, the risk difference identified between patients with and without the HPE would be attributable to the presence of the HPE. Every combination of OAC and NSAIDs was considered an HPE. There might still be patients, where the combination of an OAC with the NSAID was indicated and it cannot be identified from the data if the hazardous prescription was justified or was prescribed by error, but the association with the bleeding outcomes is the same independent of the intention of the prescriber. The associated harm from the HPE is the same, but it might not always be considered preventable.

4.5.4 Implications of the use of routinely collected data

The use of routinely collected data, such as the electronic health records from the CPRD and HES, yielded a large sample size that enabled the analysis not just of HPEs as a composite but to estimate the specific risk of a single type of HPE. The dataset offered results from a 'real-world' setting. Compared to clinical trial settings, where regular follow-up visits or reminders for GPs and patients could have resulted in HPEs being resolved earlier or patients being reminded to ask for repeat prescriptions, where they might not have done so in a real-world setting. Results from observational data are therefore considered to have a higher external validity (427). However, this increased external validity comes at the cost of an often decreased internal validity due to confounding. In observational studies, one cannot exclude that there was unmeasured or residual confounding.

Unmeasured confounding in routinely collected data

Unmeasured confounding is a common challenge in routinely collected data that were not collected for research purposes (428). While unmeasured confounding cannot be ruled out completely, this study took different measures to get an understanding of potential unmeasured confounding. First, clinicians were consulted discuss if they are aware of any

unmeasured confounders not recorded in the dataset. Second, e-values were calculated and compared to known risk factors and confounders.

In this study, close collaboration with GPs aimed to generate a comprehensive understanding of potential confounders and relations of variables with the exposure and outcome. The discussions did not suggest any unmeasured factors that affected an outcome (serious GI events, major bleeding, stroke, systemic embolism) and the NSAID exposure simultaneously. Even if this suggests a low risk of unmeasured confounding, there might be other biases present resulting from unmeasured variables. While the clinicians were not aware of any unmeasured confounders, there were a few risk factors for the outcomes not recorded in the datasets. For serious GI events diet was identified as an indirect risk factor and is not recorded in the CPRD [Appendix G]. For stroke or systemic embolism, multiple risk factors exist that were not recorded in the CPRD dataset, e.g., psychological stress, physical activity and diet (429). The clinicians consulted did not consider these risk factors as confounders because they were not thought to be associated with NSAID use, but they could still introduce other biases. It could, for example, not be checked if these unmeasured variables were balanced in the base line cohorts that could have introduced selection bias.

The other approach to understand the risk of unmeasured confounding in this study was the use of e-values. E-values describe the association of an unmeasured confounder with the outcome and the exposure required to explain away the observed effect (424). The e-value is a useful tool to understand the risk of unmeasured confounding without any assumptions on the nature of the unmeasured confounder, but they can only be interpreted in the context of the study. It has been criticised that e-values presented on their own, can be interpreted in different ways (425), and are only useful when compared with the impact of other confounders (369, 423).

For serious GI events, the e-value indicated that an unmeasured confounder would be required to be associated with both the serious GI events and NSAID exposure by a risk ratio of 5.25 (lower bound 2.58). The measured risk factors recorded in the CPRD and conditional on the exposure had a maximum association with the outcome of a HR of 2.25 for the risk factors of history of peptic ulcer [Appendix H: Table H.1]. The second highest

association was found with prior bleedings. The association of the strongest predictor identified within the dataset was much lower than the identified mean e-value and lower than the lower bound of that e-value. This is in line with other published studies investigating the risk factors of bleeding events with anticoagulants. A Swedish study found the strongest risk factor for serious bleedings in anticoagulated patients to be previous bleedings with and adjusted HR of 1.85 (95% CI 1.74 to 1.97) (407). All other risk factors had a smaller impact. A prospective registry dataset for patients receiving rivaroxaban identified heavy alcohol use of more than 80g alcohol per day compared with no alcohol consumptions as the risk factors with the strongest association with major bleeding (HR 2.37, 95% CI 1.24 to 4.53) (406). Overall, risk factors for serious GI events are in general less associated with serious GI events than the estimate of the e-value or its lower bound. Unmeasured confounders are less likely to have a stronger association than the key known risk factors in our study and in the literature. The risk of unmeasured confounding for serious GI events was therefore considered low.

For stroke, the mean e-value was 4.23 (lower bound 2.06). The risk factor with the strongest association with stroke conditional on NSAID use was prior stroke or transient ischaemic attack with a HR of 3.81 followed by hypertension (HR 2.77) and coronary heart disease (HR 2.46) [Appendix H: Table H.1]. The association of the strongest predictor identified within the dataset was lower than the identified mean e-value but not lower than the lower bound of that e-value. In line with these findings, a case-control study in 22 countries by O'Donnell et al. (2010) found the first ever stroke to be associated with hypertension by an OR of 2.64, followed by coronary heart disease with an OR of 2.38 (429). However, these key risk factors were all measured in the CPRD and were accounted for in this chapter. More interesting were risk factors that were not measured in the CPRD. The association of these in this chapter unmeasured risk factors were all comparably lower in the case control study by O'Donnell et al. (2010). The diet risk score (OR 1.35, for highest *vs* lowest tertile), regular physical activity (OR 0.69), psychosocial stress (OR 1.30), and ratio of apolipoproteins B to A1 (OR 1.89 for highest *vs* lowest tertile) were not associated with the outcome by a risk ratio greater than the lower bound of the e-value. Overall, risk factors for stroke that are not measured in the dataset in this chapter had a weaker association with stroke than the e-value and its lower bound. Unmeasured confounders are less likely

to have a stronger association conditional on NSAID treatment than the risk ratios by O'Donnell et al. (2010) predicted. The risk of unmeasured confounding for stroke was therefore considered low.

Residual confounding in routinely collected data

Residual confounding is any distortion of the results present after controlling for confounders by design or analysis. This can be a result of the aforementioned unmeasured confounders or of measurement error. In routinely collected data, measurement errors, such as misclassified diagnoses or misclassified ICD-codes used to screen for outcomes, cannot be ruled out. A UK study found that the quality, completeness or correctness of recordings of serious GI events or systemic embolism in the HES dataset was better than in primary care (430). Primary care data alone, for example, were found to only record about 20% of the bleeding events in AF patients recorded in secondary care (431). In the US, routinely collected health records based on ICD-10 codes were found to have a high predictive value for bleeding events (432). For stroke the positive predictive value of stroke diagnosis in linked CPRD/HES data was high with 79% and a negative predictive value of 100% (433). To minimise the uncertainty around the outcome events, the code lists used to identify the events were based on extensive literature searches by the author of this dissertation and a second researcher. The code lists went through a rigorous consensus process with GPs and pharmacists to identify a comprehensive set of codes for the relevant outcomes.

Misspecification of treatment episodes was also possible as a result of the process to generate prescription stop dates that are not recorded in the CPRD. In the process of generating consecutive treatment episodes, various assumptions had to be made that potentially affected the outcome. If a treatment episode with NSAIDs was assumed to last longer than it actually did, an outcome event occurring in this misspecified period could be counted as an event in the wrong exposure group. To test robustness of the assumptions made to identify treatment episodes, the impact of different grace periods in sensitivity analysis was analysed. The larger grace period of 60 days resulted in lower HRs for serious GI events and stroke events as expected due to a less precise NSAID exposure definition.

With larger grace periods, the often short periods of continuous NSAID exposure are overestimated, diluting the increased risk estimate.

Other misspecifications can be a result of unmeasured variables. NSAIDs, for instance, are a drug often sold over the counter (OTC) in pharmacies. These OTC prescriptions were not recorded in the CPRD. If patients, not exposed to NSAIDs according to the CPRD records, acquired OTC NSAIDs, the measured effect of NSAID exposure on harm outcomes would have been underestimated. However, this analysis was conducted to estimate the harm between NSAID users and non-users to predict the consequences of SMASH in reducing NSAID prescribing in patients receiving OACs in primary care. SMASH is implemented in primary care practices and has no effect on OTC NSAID use. Patients in practices with SMASH can therefore still buy OTC NSAIDs, as it is possible in the cohort study in this chapter. Therefore, the increased likelihood of ADEs with NSAID use observed in this chapter is representative of the patients targeted by SMASH with regards to potential OTC NSAID use.

Adherence to treatment is also not reported in the dataset. From the CPRD data alone, it was not known if the patients collected the prescribed drug at the pharmacy and took the drugs as prescribed. The NSAID exposure time measured in the cohort study in this chapter, might not result in the patient actually taking the medicine. The assumption was made that patients who received a prescription of a drug, picked it up at the pharmacy and were adherent to the GP treatment. If patients considered exposed in this study did not actually take the prescribed drug, this could have had a diluting effect on the analysis. The results report a significant association of NSAID use with the ADEs despite this diluting effect of potential non-adherence.

Another challenge that can introduce residual bias, was missingness. While missingness for smoking status, ethnicity and blood pressure measurements were low, about half of patients had no BMI recorded in the 12 months before the index date. It is unclear how many of these missing patients potentially had a high BMI that was reported as a risk factor for serious GI events. If patients exposed to the NSAID were less likely to have a missing BMI record than unexposed patients, this could have reduced the risk increase associated

with the NSAID exposure. However, high BMI was only an indirect GI risk factor and not a confounder and it was assumed the missingness was tolerable. In the sensitivity analysis, NSAID use was still associated with an increased risk of serious GI events and stroke, when patients with missing BMI, smoking or ethnicity records were excluded. The HRs estimated for this sensitivity analysis are smaller than those for the respective base case analysis. Other authors speculated that patients with missing BMI, smoking and ethnicity records are those with only minimal contact with primary care (434). Patients with missing records are therefore likely to be systematically different from those without missing records and cannot be considered missing at random. This is one of the key assumptions required to do multiple imputation (434). Multiple imputation techniques were not used in this study because the data were not considered missing at random. The potential to distort the observed effect was considered low because the variables were not considered confounders by the GPs.

Another variable with inconsistent reporting was INR. Time in therapeutic range from INR measurements is rarely available and not recorded systematically in CPRD. In the discussions with GPs, uncontrolled INR was considered a potential confounder. While records exist for warfarin users, INRs are not measured routinely for DOAC users. This has therapeutic reasons but makes the systematic assessment as a covariate difficult. It is also known that not all practices receive INR records from anticoagulation clinics, which adds to the inconsistency of the records (56). According to the GPs involved in this study, they would be hesitant to prescribe a NSAID to a patient with uncontrolled INR. This could have led to a higher proportion of patients with controlled INRs in the exposed group, which could have had a diluting effect on the impact of NSAIDs.

4.5.5 Implications of the analysis

The challenges of electronic health records that derive from unmeasured and residual confounding were explored in this discussion. By using a robust study design, this study aimed to minimize the effects of such confounding. PSM aimed to resolve imbalance between observable baseline characteristics of NSAID users and non NSAID users and the new user design aimed to mitigate selection bias, such as healthy user bias.

The propensity score

There are several advantages to the propensity score matched analysis chosen for this study. In PSM, the analysis is separate from the study design, similar to a clinical trial, and only one model is needed to estimate the propensity score. Another advantage of propensity score method is their transparency. To test if the propensity score balances the distribution of base line characteristics in the propensity score matched samples, various diagnostic measures are available (404). Testing for balance of covariates in standard regression methods is a 'black box' (435) and specification of adequacy of the applied model is easier in propensity scores analysis compared to regression adjustment (260). Matching on the propensity score was found to be an effective propensity score method to balance out the difference between treatment groups compared with IPTW or stratification on the propensity score (401). This was tested in an empirical case study and Monte Carlo simulations were used to identify the methods eliminating a greater degree of the systematic difference between exposed and unexposed individuals using within-quintile standardised differences (401).

The new user design

The study only included new NSAID users to increase internal validity of the study. New user designs are used to reduce bias due to healthy user effect or time-dependent event risks and to replicate a trial like design more appropriately. This had also the advantage that the baseline characteristics were all pre-treatment variables. For estimating the propensity score, it is important not to include any post-treatment covariates, as it is supposed to represent the probability of treatment assignment. In a prevalent user design, propensity score methods were considered less effective in adjusting for confounding (436) because the propensity score does not estimate the probability of treatment but a mix of the probability of continuing treatment and initiating treatment (349). This could be a result of including post-treatment covariates that were potential mediators. Severe renal disease pre-treatment, for example, is included in the propensity score as a confounder because it effects the bleeding risk and NSAID prescribing. Including the pre-treatment variable in the propensity score model adjusted for the decreased likelihood of GPs to prescribe an NSAID to this patient group because of the already elevated bleeding risk. After treatment

assignment, severe renal disease is not a confounder anymore but a mediator. NSAIDs can have severe renal side effects and renal insufficiency is associated with an increased bleeding risk. Conditioning on these outcome mediators, such as severe renal disease, can result in a reduction of the total causal effect [4.3.7].

The increased internal validity through a reduction of prevalent user bias came at the cost of a reduced sample size. The three-months washout period excluded more than half (5074 out of 8251) of patients with at least one NSAID prescription during follow-up. The washout period did potentially not cover all prevalent NSAID users. Extending the washout period would have led to the loss of even more patients. The washout period was based on the maximum observed prescription length (99% percentile) plus a grace period of 30 days. The few prescriptions potentially overlapping with the index date were considered negligible. Longer washout periods were considered to include more previous NSAID users and only a small number of additional prevalent users. The new user design did not include all previous NSAID users, but it aimed to remove those with continuous or prevalent use of NSAIDs as identified as sufficient by Ray et al. (2003) (348). Any residual effect NSAID use had on pre-treatment characteristics is supposed to be captured by the baseline characteristics that were balanced between treatment groups and therefor considered negligible. As a result of the reduced sample size, precision of the results was reduced in sensitivity analysis looking at changes in washout windows for NSAIDs. The confidence intervals got larger with extending the washout period. The three-months washout period was considered to be appropriate to remove the most prevalent user bias and still yield significant results.

Another trade-off existed between the increased internal validity and a reduced generalisability. The new user design gives a less biased estimate, but results are not generalisable to all OAC users anymore. The association measured compared NSAID users with patients with similar characteristics to NSAID users, not with all OAC users in the general population. If the aim of this study would have been to identify the difference in OAC users from the general UK population, this would be a clear limitation. However, the aim of this study was to utilise these estimates in the overall economic evaluation of SMASH. Patients identified in SMASH were taken off the NSAID if the intervention was

successful and were then grouped as patients without the HPE. Hence, the characteristics of the patients with and without the NSAID should be similar as they were in this chapter.

Statistical limitations

This study was not able to estimate the cumulative incidence of serious GI events in the exposure groups. The presence of competing risk events and time-varying exposure as described in section 4.3.8 did not allow inference from the cause-specific hazard function on the cumulative incidence function (414, 418, 419).

4.5.6 Implications for thesis

In order to estimate the economic impact of HPEs, their effect on harm needs to be quantified. Subsequently, consequences in form of costs and quality of life associated with harm (ADEs) can be assessed. This study helped to understand the ADEs affected by NSAID use in anticoagulated patients. Serious GI events and stroke were significantly associated with NSAID use in anticoagulated patients. Assessment of the economic impact of this specific HPE should entail the consequences associated with these two ADEs. Incidence of systemic embolism was very low with fewer than two events in 1000 person years in non-NSAID users [Table 4.5]. The non-significant HR (95% CI) of the increased risk could be due to the fact that the analysis was not powered to find a difference. However, the low incidence rate suggests that this ADE has only a small effect on the economic impact of this HPE type compared with the effect of the more frequent serious GI events and stroke.

This study focused on events recorded in secondary care that led to hospital admission. Not all ADEs associated with NSAIDs lead to hospitalisation. Serious GI events and strokes are predominantly managed in secondary care, but other ADEs common to NSAIDs, such as dyspepsia, would usually be managed in primary care (306). The subsequent chapters need to assess how NSAIDs impact ADEs that occur outside of the hospital as well.

Going forward, different aspects of this study can be used to assess the economic impact of NSAID use in anticoagulated patients. The HRs reported here are used to project long-term consequences of the HPE. This study also provides a detailed description of

characteristics of patients at risk of this HPE type. Quality of life, for example, is dependent on age (437), and information on the mean age of an anticoagulated patient in England is necessary to project impact of ADEs on quality of life.

4.5.7 Considerations for future work

As described in Section 4.5.6, the subsequent chapters extrapolate the long term consequences of exposure to the hazardous prescription by estimating the cost and quality of life associated with the ADEs. Future post-doctoral research will extent this analysis to estimate the total healthcare cost and the differences of exposed and unexposed patients from the dataset in this study as outlined in the ISAC protocol (No 18_235). The CPRD provides information on resource use in primary care on GP consultations, diagnostic tests and prescriptions. For secondary care resource use, information about outpatient visits, hospital day care, admissions and A&E attendances can be identified from different HES datasets. Unit costs will be taken from the available sources for UK health care. Primary care unit cost will be obtained from the PSSRU publication (438) and secondary care unit cost from the national reference cost schedule (439). Healthcare resource group (HRG) will be linked to each HES record using the latest HRG reference Costs Grouper software.

Future research could also consider exploring other potential ADEs not included in this chapter. While non-significant, there is an indication that the hazardous prescription also affects the risk of other cardiovascular events such as MI (308) and heart failure (309). The increased risk assessed could be due to chance or findings are non-significant because the study was not powered to identify a difference. As described before both studies do not transparently report how NSAID exposure is identified and rely on patient recall. As a result of mis-specification of exposure periods, the effect could have been diluted which makes finding a significant difference more difficult. The availability of more precise data on NSAID prescriptions from the CPRD records might reduce these diluting effects. Hence, the possibility to find a significant difference might be higher.

Chapter 5 - Economic impact of NSAID use in anticoagulated patients

Chapter Five reports on a study estimating the economic impact of NSAID use in anticoagulated patients using a state-transition model. The methods describe how the state-transition model was conceptualised, and how it was populated. The results illustrate the impact of hazardous prescribing of NSAIDs on incremental costs and QALYs. In the discussion, the relevance of this projection of harm and healthcare costs associated with this type of HPE are described.

5.1 Introduction

The health burden of ADEs that were potentially caused by HPEs in the UK was estimated to be almost £100 million per year (31). In Chapter Four, the increased risk of ADEs associated with the presence of a specific HPE type were quantified. This chapter takes this analysis further to project costs and health outcomes associated with NSAID use in anticoagulated patients.

Evidence on the estimated economic impact of HPEs varies between countries, care settings and time of the study (30, 146). No UK estimates on the economic burden of HPEs in general were identified in two systematic reviews of cost of HPEs (146, 153). Because evidence on cost (in 2015 Euros) of HPEs varies substantially by HPE type from €68 per hazardous inhaler prescription to €6.9 million for litigation claims associated with anaesthetic error (146), it was important to estimate the economic impact of each HPE type separately. For the economic impact of NSAID use in anticoagulated patients, no evidence was found in the literature. In order to generate precise and specific estimates, the economic impact of NSAID use in anticoagulated patients is assessed in this chapter. QALYs are used as a patient outcome measure to estimate quality and quantity of life as recommended by NICE (3). Instead of attaching a fixed cost and QALY estimate to the ADEs identified in the cohort study in Chapter Four, this study used a modelling approach. Modelling approaches, often used in economic evaluations, represent a simplified version of reality and allow generation of estimates of healthcare outcomes and costs not captured

in the primary analysis (440). Models can present a realistic pathway or sequence of events that patients with and without HPE would experience. The modelling approach also allows inclusion of primary and secondary care events as well as associated costs. Solely costing secondary care events was criticised in earlier studies by reviews on the economic impact of HPEs because it does not account for long-term consequences on costs and quality of life (146, 441).

The overall aim of this chapter was to estimate the economic impact of NSAID use in anticoagulated patients. This was achieved through the following: (i) identifying the type and probability of ADEs associated with NSAID use, (ii) constructing a state-transition model representing the treatment pathways of anticoagulated patients with and without a prescription of an NSAID, (iii) populating the model with UK relevant transition probabilities, utilities and resource use data, and (iv) generating an estimate of the impact of NSAID use on patient outcomes (measured as QALYs) and healthcare costs to NHS/PSS.

5.2 Methods

The methods describe the development, input parameters and analysis of the decision-analytic model. The Assessment of the Validation Status of Health-Economic decision models (AdViSHE) criteria were followed and are reported in Appendix L (442).

5.2.1 Developing the model

The economic impact of hazardous prescribing of NSAIDs on anticoagulated patients was estimated via construction of a cohort-level state-transition model (Semi-Markov model). This approach was selected over a decision tree approach because state-transition models are more flexible and useful when a decision problem involves a risk that is not constant over time, such as mortality. More advanced approaches, for instance, using patient-level discrete event simulation, were not used, as it was expected that the higher data needs of these approaches would not be met. Health states included were specific to the HPE type under investigation representing key ADEs associated with NSAID use in anticoagulated patients and the probability of these events in the presence and absence of the NSAID.

Literature searches

The literature was searched for existing model structures and ADEs associated with the HPE type, transition probabilities, health related quality of life and resource use data. Literature searches were conducted through the electronic databases Medline, Embase and the HTA database using HPE and treatment pathway-specific search terms. Search terms for the literature searches conducted are reported in Appendix K [Table K.1]. Search terms describing OAC and NSAID users were based on all medications relevant for these drug groups. The search terms for NSAIDs included general terms for NSAIDs and specific terms for the most common seven NSAIDs according to the 2009 NHS NSAID safety audit (443) and mesh terms/subheadings for these if available in Medline or Embase. Search terms for OACs were based on the drug substances available in the UK and the main diagnoses of OACs. To identify input parameters associated with the chosen health states, the literature was searched for serious GI events, GI discomfort, symptomatic ulcer and stroke. Justifications for the choice of health states are provided later. Search terms for adverse GI events were developed using published search criteria of functional dyspepsia from Agah et al. (2020) (444) and the diagnoses defining GI discomfort in the MUCOSA trial that were diarrhoea, abdominal pain, dyspepsia, nausea, vomiting and flatulence. Because, for symptomatic ulcer, various subheadings were available in Medline and Embase, the search was mainly based on those with the addition of keywords on typical locations in the upper GI tract with ulcerations, such as gastric ulcer. To identify studies looking at mortality, search terms were used from a recent review by Tian et al. (2020) (445). Search terms to identify existing model structures, resource use and health related quality of life were used in line with search terms used to inform state-transition models developed for other HPE types in PROTECT.

References in English or German and limited to humans were included. After excluding duplicate records, references that remained for further evaluation were selected on title and abstract. Studies were included if they reported relevant economic models or examined issues on the incidence and/or prevalence, treatment health related quality of life or resource use of the consequences of NSAID use in anticoagulated patients. Subsequently, full texts of the retrieved references of the previous section were evaluated. Finally, reference lists of the retrieved papers were reviewed. The results of the literature

searches are reported in Appendix K [Table K.2] with the number of studies available for abstract screening and a list of the relevant studies identified. If multiple studies provided potentially relevant records used to inform model structure or identify potential model parameters, studies were considered preferable if they were UK-based/relevant, large and recent (i.e., published since 2010).

Model structures reported in the literature

The literature search identified no previous state-transition models comparing patients with concomitant OAC and NSAID treatment with patients treated with OACs only. The search terms are presented in Appendix K [Table K.1] for the search 'state-transition models comparing NSAID use in anticoagulated patients'. An additional literature review was performed to identify state-transition models that incorporated health states following treatment with NSAIDs for patients with any health condition. The results of this review informed the health states for the state-transition model in this thesis by identifying relevant outcomes of patients who were either exposed or not exposed to NSAIDs [Appendix K: Table K.2].

There were 70 published state-transition models that compared either (i) different NSAIDs or (ii) an NSAID strategy to a non-NSAID strategy [Appendix K: Table K.2]. Among the identified models on NSAIDs, nine examined the impact of gastroprotection on GI events in people prescribed NSAIDs (227, 446-453). Others investigated different treatment strategies including NSAIDs for lower back pain, rheumatoid or osteo-arthritis (454-459). Only two models were identified that assessed the economic impact of a type of HPE (149, 151). These models had the advantage that they described in detail, how and when actions to resolve the HPE type were taken. Moriarty et al. (2019) compared patients over 65 years of age with the hazardous prescription of an NSAID and without the NSAID. The original economic evaluation of the PINCER trial by Elliott et al. (2014) compared patients with a history of peptic ulcer with a concomitant prescription of an NSAID or without. The latter was used as a starting point for developing the new model (227). This model was chosen because forgoing literature searches, methods and results were reported more transparently and, in more detail, than in Moriarty et al. (2019). The other NSAID models

identified were searched for any developments in more recent models in this area and more recent data to inform model parameters so that the resulting model reflects current evidence.

While NSAIDs have always been associated with GI adverse events, their cardiovascular risk was under-recognized for a long time (337). The following sections describe how and if GI adverse events and cardiovascular ADEs were included in published state-transition models: first, in models around the exposure to NSAIDs and, second, models in the denominator population of anticoagulated patients.

GI adverse events reported in NSAID models

The *de novo* state-transition model in this thesis aims to represent key GI adverse events associated with NSAID use as health states. The published state-transition models comparing NSAID treatment strategies, referred to as NSAID models, had in common that they included 'no GI complications', 'GI discomfort', 'serious GI event' and 'dead' as health states. In addition to these health states, symptomatic ulcer was included in the *de novo* state-transition model in this thesis. While most health states were consistent between models, the models differed with regards to how and if peptic ulcer was incorporated. The model structures based on de Groots' state-transition model (151, 452, 453) incorporated peptic ulcers that caused bleeding or perforation within a broader serious GI event health state. Most other models included a separate mutually exclusive health state for symptomatic ulcer (an ulcer diagnosed after the patient reported GI symptoms). Asymptomatic ulcers refer to ulcers identified in endoscopies that did not cause any symptoms (460). The published models based on the state-transition model developed by Latimer et al. (2009) for clinical guidelines on osteo-arthritis from NICE (450, 454, 456-458, 461) and the model used in the PINCER evaluation included symptomatic ulcer as a separate health state. A single centre study in South Korea reported NSAID use to be an independent risk factor for symptomatic ulcers, but the study did not find asymptomatic ulcers to be a risk factor (460). Asymptomatic ulcers are because of their asymptomatic nature, not assumed to impact health related quality of life. Asymptomatic ulcers are not reported to GPs by the patient and are not likely to result in hospital admissions if no

symptoms are present. Symptomatic ulcers, on the other hand, are thought to impact health related quality of life and are associated with resource use for managing the condition. Consequently, only symptomatic ulcers were relevant in the *de novo* state-transition model in this thesis.

GI adverse events in decision-analytic models evaluating OACs

To get an understanding of how serious GI events were modelled, not only in NSAID models but also in the denominator population, a literature search on models comparing different OAC treatments in AF was conducted [Appendix K]. The main published models identified that were frequently used by other researchers were (i) a state-transition model by Sorensen et al. (2009) (462) used in a health technology appraisal for NICE (TA249) (463) and other studies (351, 464-471), (ii) the model by Dorian et al. (2014) used in TA275 (472) and multiple other studies (473-480), and (iii) a structure by Gage et al. (1995) (481) that was adapted by Lee et al. (2012) (482-484). Five additional model structures were identified that used a different model structure from the three main published models (485-488).

The review of the published models focused on how serious GI events were included in the model structure. Of the GI adverse events described in the NSAID models, only serious GI events, such as GI bleeding, were modelled in the published models on OACs. Serious GI events were mainly modelled as extracranial haemorrhages of which the majority of the events are serious GI events. Long-term impact of serious GI events was not incorporated in the structures of the three main models (462, 472, 481). Serious GI events were included as temporary events or transient states with no effect on future transition probabilities but with a case-fatality, a one-time cost and disutility in the three main published models. Only three of the five models using a different structure form the three main published models predicted a long-term effect of serious GI events in subsequent cycles. Ademi et al. (2015 and 2016) (487, 489) included a separate health state for the time after a serious GI event to account for a prolonged reduced quality of life and Sing et al. (2013) (486) and Lopez-Lopez et al. (2017) (485) expected an increased mortality in subsequent cycles. In the published models, only acute costs for the serious GI event were included. For the *de novo*

state-transition model in this chapter, it was assumed that a serious GI event does not impact subsequent cycles with regards to healthcare costs.

Inclusion of cardiovascular adverse events

In the original PINCER model, cardiovascular events associated with NSAID use were not included as a health state, as this evidence was not mature at this stage (227). More recently, evidence has suggested that NSAIDs could be associated with an increased risk of cardiovascular events (337). For anticoagulated patients, the effect of NSAIDs on thrombotic cardiovascular adverse events, such as stroke, was of specific interest because anticoagulated patients already have an increased risk of these events. In the literature review evidence on an increased risk of stroke, systemic embolism, heart failure and MI was investigated [Appendix K].

The mixed results on the effect of NSAIDs on stroke and systemic embolism were discussed in detail in Chapter Four [4.2.2, 4.5.2]. The results of the cohort study in Chapter Four indicate a significant increase of the stroke risk in the presence of concomitant NSAID therapy. This confirmed the results from Kent et al. (2018) in a post-hoc analysis of the RE-LY trial data in anticoagulated patients. Consequently, stroke was included as a cardiovascular ADE in the *de novo* state-transition model developed in this thesis. The incidence of systemic embolism in both OAC, and OAC and NSAID users was low and the difference non-significant [4.4.2]. Heart failure (309) and MI (308) were also not significantly associated with NSAID use. To avoid introducing more uncertainty into the model by including systemic embolism, heart failure and MI that were not significantly associated with NSAID use, these were not included as ADEs. Because potential harm from the HPE related to these ADEs is not included in the model, this could underestimate the incremental cost and QALYs. Not including these events was, therefore, a conservative assumption. It was assumed that these ADEs, however, are rare and only have a small effect on the incremental cost and QALYs associated with the HPE type.

Expert consultation

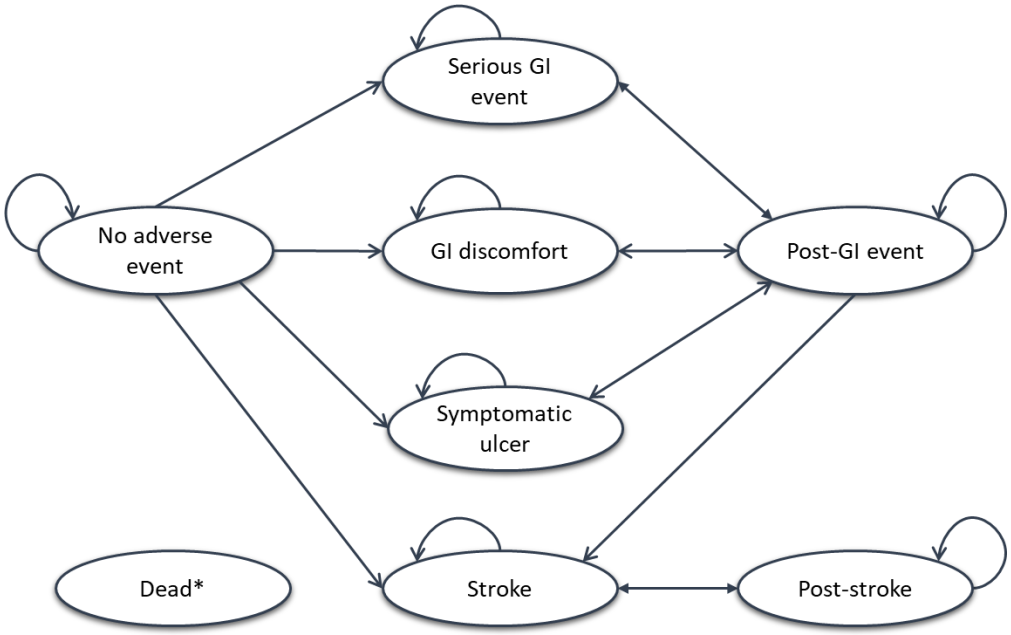
Different experts were involved in the design of the *de novo* state-transition model and to assess face validity of the assumptions made during the model conceptualisation. The model structure and key assumptions were discussed with the multi-disciplinary members (including GPs, pharmacists, epidemiologists, health economists and lay collaborators) of the PROTECT study team. A summary of key assumptions made were sent to two GPs from the PROTECT study team, to confirm these were reasonable. Questions with feedback are reported in Appendix M. The final model was also presented to a patient representative in a video conference call for face validation. The patient is diagnosed with AF and is experienced in NSAID use and NSAID related ADEs. The face validation included a sense check of input parameters, and a discussion on comprehensibility of the model structure and assumptions made. The *de novo* state-transition model also underwent an internal validity check. A health economist, experienced in conceptualising models around NSAID use, checked for errors in the working models and written report following a checklist developed within the PROTECT study team that is reported in Appendix N.

5.2.2 Final model structure

In this section the final model structure is described with the model specifications and the cohort characteristics. The health states in the state-transition model for this study were based on the original NSAID model used in the PINCER evaluation (227), newer models identified through literature searches and expert input.

Figure 5.1 illustrates a simplified structure of the state-transition model. A list of assumptions with regards to the model structure is provided in Appendix M.

Figure 5.1: Final model structure



→ Arrows describe possible transitions between health states;
* From all health states patients can transition to the dead state

Two cohorts are followed through the same state-transition model. The non-HPE cohort consisted of patients with OAC treatment and the HPE cohort consisted of patients with OAC and concomitant NSAID treatment. The two cohorts only differ in specific transition probabilities as described later. All patients are assumed to start in the no adverse event state. ADEs potentially attributable to NSAID use were stroke and GI adverse events. The GI adverse events were serious GI events (GI bleeding events, ulcer perforation and ulcer bleeding), GI discomfort (diarrhoea, abdominal pain, dyspepsia, nausea, vomiting, flatulence) and symptomatic ulcer (excluding ulcers identified during endoscopy that did not cause symptoms). From all health states patients could transition to the dead state.

After entering an ADE health state (serious GI event, GI discomfort, symptomatic ulcer or stroke), the model assumes that the HPE is resolved and the NSAID is removed in the subsequent cycle. The model structure assumes that there is no impact of the NSAID on the risk of having a subsequent ADE. Once any initial ADE is resolved, the patient either experiences a recurrent event or transitions to a post-event state. In the two post-event states, it is assumed that there is no increased risk of any adverse event compared with the non-HPE cohort (who received no NSAID treatment). The post-GI event state is equivalent

to the no adverse event state in the non-HPE cohort with regards to transition probabilities, resource use and health related quality of life. This structural assumption was supported by the two GPs and two pharmacists consulted. All four experts agreed with the assumptions that the HPE was preferably resolved by removing the NSAID and that following any GI adverse event, removing the NSAID would be the first choice. For stroke, the GPs indicated that they would remove the NSAID in only 50% of the cases. This was not reflected in the model because data to populate this were not available. There was no evidence on how NSAIDs increase the risk of stroke recurrences. Assuming the recurrence risk is the same, the only difference would be the cost of the NSAID, which is negligibly small. Therefore, it was considered justifiable to assume that patients moving to the post-stroke state stop the NSAID treatment if they were exposed to the HPE.

From the post-stroke state, patients could either die, experience a second stroke or remain in the health state. Stroke was considered the most severe ADE, and no GI adverse events were possible afterwards. This approach to use stroke as a semi-absorbing state is widely used in the decision models comparing OAC treatments (473-480). In contrast, after entering the post-GI event state transitions to all other ADE states are possible.

There are no transitions possible between the ADE states. The ADEs are not precursor events but alarming events. Patients go to the state of the worst ADE that happens and then the HPE is resolved. All patients have a risk of death according to the health state they are in. Death can be a result of the ADEs or death as a consequence of age. ‘Dead’ is an absorbing state.

The model specifications are summarised in Table 5.1. The costs and outcomes were discounted at the recommended rate of 3.5% per annum and half cycle correction was applied (3). The cost year used was 2019. Costs were assessed from the perspective of the NHS/PSS.

Table 5.1: Model specifications

Specification	***Details***	***Justification***
Time horizon	Lifetime (100 years of age)	As per NICE reference case (3)
Cycle length	3 months, half cycle correction	Appropriate time period to observe outcome
Perspective	NHS and social care	As per NICE reference case (3)
Discounting	3.5% costs and benefits	As per NICE reference case (3)
Health states	***Description of health state***	
No adverse event	This is the point of model entry for either the HPE cohort (OAC+NSAID) or the non-HPE cohort (OAC only)	
GI discomfort	People experience GI discomfort (diarrhoea, abdominal pain, dyspepsia, nausea, vomiting, flatulence), requiring treatment by a general practitioner	
Symptomatic ulcer	People have a clinical diagnosis of a peptic ulcer following adverse GI symptoms, which may include endoscopic confirmation	
Serious GI event	People experience a serious GI event (GI bleeding, peptic ulcer bleeding or GI ulcer perforation) requiring treatment in hospital	
Post-GI event	Following an alarming GI event, people return to a health state where the risk of subsequent events is the same as in the non HPE cohort; from here, they can experience further events	
Stroke	People experience stroke requiring hospitalisation.	
Post-stroke	Following an alarming stroke event, people return to a health state where the risk of subsequent events is the same as in the non HPE cohort; from here, they can experience only further stroke or die	
Dead	There is a risk of death from all health states. Mortality rates depend on age and sex and can be increased when ADEs occur	

ADE: adverse drug event; GI: gastro-intestinal; HPE: hazardous prescribing event; NICE: National Institute for Health and Care Excellence; NSAID: non-steroidal anti-inflammatory drug; OAC: oral anticoagulant

Where it was necessary to inflate costs, the Hospital & Community Health Services (HCHS) index was used to inflate costs up to 2014/15 and then the newer NHS Cost Inflation Index (NHSCII) was used to inflate from 2014/15 to 2018/19 (285). The inflation indices are summarized in Table 5.2 below.

Table 5.2: Inflation indices for different cost years since 2009/2010

Price year	HCHS Pay and Prices Index (1987/8=100)	NHSCII Pay and Prices Index (2014/15=100)
2009/10	268.6	
2010/11	276.7	
2011/12	282.5	
2012/13	287.3	
2013/14	290.5	
2014/15	293.1	100.0
2015/16		100.4
2016/17		102.5
2017/18		103.7
2018/19		106.1

HCHS: Hospital & Community Health Services; NHSCII: NHS Cost Inflation Index

To extrapolate the impact a change in HPE rates had on health outcomes and costs, the cost-utility analysis followed the HPE and non-HPE cohort for a life time. The time horizon should capture all health effects and costs relevant to answer the decision problem as recommended by the ISPOR good modelling task force (490). Following the recommendations in the NICE reference case, a lifetime horizon was used that followed patients up to a 100 years of age (3). A cycle length of three months was considered appropriate to represent the frequency of clinical events and interventions. Three month cycles were consistently used among OAC models in AF (TA249 (463) and TA275 (491)) and NSAID models, such as the osteoarthritis model used for national clinical guidelines (492). Clinicians expressed agreement with the choice of the cycle length [Appendix M].

5.2.3 Cohort characteristics

The non-HPE and the HPE cohort consisted of anticoagulated patients. Cohort characteristics were selected to reflect the target population in the UK as closely as possible. Results from the epidemiological study conducted as part of this dissertation on patient characteristics were used to define the cohort characteristics in this model. This study was described in detail in Chapter Four and is referred to as the cohort study conducted as part of this dissertation. The baseline characteristics of the cohort were representative of patients who were prescribed an OAC in primary care and were reported in Table 4.4. The mean age of 70 years in the CPRD/HES/ONS cohort study was used as the age at model entry. The proportion of men (55.11%) and women (44.89%) was also used from the primary care records.

5.2.4 Input parameters

This section describes the sources of evidence for input parameters (transition probabilities, health state resource use and costs, as well as health related quality of life) used to inform the *de novo* state-transition model. Summaries of values for each input parameter type are provided in the individual sections.

5.2.4.1 Transition probabilities

In this section the transition probabilities used in the model are reported. Ordered by ADE type, the evidence on transition probabilities in the non-HPE cohort from (i) the no adverse event state to the ADE health states, (ii) the probability of death from the ADE (fatality of events) and (iii) the probability of a recurrence of the ADE are described. The transition probabilities for the HPE cohort only differ from the non-HPE cohort in the transition probabilities from the no adverse event state to the individual ADE states. The increased likelihood of the ADEs in the presence of NSAIDs is described at the end of this section. A summary table of the values for transition probabilities used in the state-transition model is reported in Table 5.5.

Probability of GI discomfort in the absence of an NSAID

In the phase 3 trials of DOACs compared with warfarin, ARISTOTLE, ROCKET-AF and Engage AF TIMI 48, the risk of GI discomfort was not increased in patients with the DOAC under investigation compared with warfarin, and therefore the incidence of GI discomfort was not reported. The only DOAC that was found to have an impact on dyspepsia in particular was dabigatran (326). Due to this increased risk, results on the effect on GI discomfort were reported in more detail for the RE-LY trial comparing dabigatran and warfarin. A combined outcome including upper abdominal pain, abdominal pain, abdominal discomfort and dyspepsia was reported for the dabigatran and warfarin cohort. Because dabigatran is rarely used in the UK (315) and the other RCTs in AF found no difference in the probability of GI discomfort between warfarin and the DOAC, the probability for warfarin was assumed to be representative of the OAC cohort. The mean age in the RE-LY trial was 71, which is very close to the age at cohort entry in this state-transition model. The probability of GI discomfort was 5.8% (348 out of 6022 patients on warfarin) in a mean duration of warfarin treatment of 23 months. This probability was used to derive a three-months probability for GI discomfort. Risk of GI discomfort was assumed to be constant over time for warfarin and dabigatran users as reported in a Danish observational study (493).

Probability of death from GI discomfort

Following consultation with the GPs, it was assumed that subsequent to a primary-care managed episode of GI discomfort there would be no increase in the risk of death [Appendix M]. Therefore, age and sex adjusted general population mortality rates, as per Office for National Statistics (494), were applied to this health state. Other decision-analytic models, such as by Dorian et al. (2014) (472) or Lip et al. (2012 and 2014) (477, 495), also used a constant probability of dyspepsia with no fatal events.

Probability of recurring GI discomfort in subsequent cycle

No literature was identified regarding the short-term recurrence/resolution of GI discomfort. Following consultation with the GPs, it was assumed that once somebody entered the GI discomfort state, their risk of recurring dyspepsia for the following cycle is the same as the risk of recurring serious GI events for which there is some observational data [see subsection below on 'Probability of recurring serious GI events']. After a maximum of two cycles in the GI discomfort state, it was assumed that the GI discomfort was resolved, and the person entered the post-GI event state. Consultation with the GPs confirmed this was a reasonable approach [Appendix M].

Probability of symptomatic ulcer in the absence of an NSAID

No evidence was found in epidemiological studies that OACs were associated with an increased risk of symptomatic ulcers compared with the general population (460, 496, 497). Clinically, the mucosa is not affected by OACs and the RCTs did not report an increased risk of peptic ulcer. The probability of reported peptic ulcer was low in the ARISTOTLE trial with 0.01% of warfarin and 0.01% of apixaban patients and in the ROCKET-AF trial with 0.03% for rivaroxaban and 0.00% for warfarin according to results presented on the website 'clinicaltrials.gov' on reported ADEs during the trials. The ADEs reported in the RCTs were not seen as representative, as it was not clear if these self-reported events were complete. An observational cohort study using The Health Improvement Network (THIN) database analysed trends in symptomatic ulcer incidence until 2005 for a UK primary care setting. High risk patients with bleeding or peptic ulcer history were excluded. The risk of

uncomplicated peptic ulcer per 1000 person years was 1.4 (95% CI 1.3 to 1.5) and 1.1 (95% CI 1.0 to 1.2) for men and women, respectively (497). This rate for the age group 65-74 years of age by Cai et al. (2009) was used to calculate the three-months peptic ulcer risk in the OAC cohort weighted by the proportion of female and male patients in the model population.

A similar incidence rate was reported in a systematic review of peptic ulcer incidence including studies until February 2009. Incidence of uncomplicated peptic ulcer was 0.9 (95% CI 0.78 to 1.04) per 1000 person years with most studies dated before 2000 and only one from the UK (498).

Probability of death from symptomatic ulcer

An observational study in Denmark collected data on 4421 patients with incident uncomplicated ulcer (i.e., non-bleeding) between 1993 and 2002 (499). The study reported a standardised mortality ratio of 11.6 (95% CI 9.6 to 13.9) in the first month following the diagnosis (499). A Finnish observational study by Malmi et al. (2016) collected data on 4154 patients with uncomplicated peptic ulcers between 2000 and 2008 (500). The reported standardised mortality ratio was 2.16 (95% CI 2.05 to 2.29) at one year. The Finnish study was chosen for the model because data were more recent and the estimate seemed more reasonable compared with mortality related to other ADEs. The standardised mortality ratio represents a factor by which the expected number of deaths in the absence of an ulcer in the normal population is multiplied to generate the number of deaths in patients with an ulcer. The factor was multiplied by the death rate in the no adverse event state and converted into probabilities.

Probability of recurring symptomatic ulcer

No study was found that assessed the risk of symptomatic ulcer recurrence. A Korean and a Finnish study were available that looked at any ulcer recurrence reporting a cumulative incidence over five and one year, respectively (501, 502). The Finnish study by Malmi et al. (2014) included all types of hospitalised peptic ulcers, e.g., perforated, bleeding, asymptomatic and symptomatic ulcers. The observational study collected data from

hospitals in a district in Finland and reported a cumulative incidence of recurrent ulcers of 13.1% (95% CI 12.4% to 13.9%) in the first year (502, 503). The Korean observational study by Yoon et al. (2013) investigated recurrences of NSAID induced peptic ulcers and reported a five-year cumulative incidence of 10.9% (95% CI 2.6% to 19.2%). However, sample sizes were small with only 57 NSAID related ulcers. Both studies did not report a three-months probability of the recurrence. Assuming a constant increase of the recurrence risk over time, a three-months probability could be calculated, but Malmi et al. (2014) showed that the recurrence risk is highest in the first months. They also did not measure recurrence of symptomatic ulcer but measured ulcers in hospital, including bleeding events that were considered serious GI events in this model. Instead of using these estimates that were reported for a general population, the meta-analysed estimates in anticoagulated patients used for serious GI events were considered more appropriate (504). Tapaskar et al. (2020) reported a three-months probability of 10.1% that did not need to be transformed further. The two GPs involved in the clinical validation considered 10.1% a reasonable probability of recurrence for symptomatic ulcer.

Probability of serious GI events in the absence of an NSAID

Incidence rates for serious GI events in anticoagulated patients from Chapter Four were used to derive the three-months probability of serious GI events. The mean age of anticoagulated patients was 72 years, and the incidence rate was 7.78 (95% CI 6.03 to 10.04) serious GI events per 1000 person years for anticoagulated patients with no concurrent NSAID use. To adjust for the increasing risk of serious GI events with age, a risk adjustment factor was previously used in a NICE technology appraisal (491) and other models (473-480). The risk adjustment factor was based on a systematic review on intracranial haemorrhage (505). However, more recent data from the The AnTicoagulation and Risk factors In Atrial fibrillation (ATRIA) study, a cohort study using routinely collected health data in the US, reported a two-fold increase of the bleeding risk for warfarin patients between 70 and 79 years of age compared to patients aged under 60 years. The ATRIA study, however, only reported an adjusted RR of 2.1 (95% CI 0.7 to 6.4) for patients ≥80 years compared with patients aged under 60. The entry age of this cohort is already after

the risk levelled off and, therefore, no increased risk of serious GI bleeding events over time was applied.

Probability of death from serious GI events

No UK study was found that reported fatality of serious GI events in anticoagulated patients. Serious GI event rates reported in Chapter Four were too small to show robust incidence rates of death subsequent to these.

An observational study using Swedish routinely collected health data assessed 90-day mortality of anticoagulated patients after 4291 hospitalisations for GI bleeding events (506). Of 652 DOAC patients with a serious GI event, 71 (10.9%) died within 90 days. For the 1293 warfarin users, the 90-day mortality was 11.4% (n=147) and 10.9% for the 652 DOAC users. The mean age of the cohort was 78 years for warfarin and DOAC patients experiencing GI bleeding events. The incidence estimates for warfarin and DOACs reported in the study were combined (weighted by the number of patients in each group) and used as three-months probability of death following the serious event in the age group 75-79 years. After validation from GPs, for other age groups the estimate was adjusted using the relative relationship to the respective mortality in the general population (494).

Probability of recurring serious GI events

A systematic review and meta-analysis of studies published up to February 2019 assessing the risk of GI bleeding recurrence, thromboembolism and mortality in patients with a GI bleeding event conducted a meta-analysis of the 90 day re-bleeding probability (504). The identified studies investigated OAC users from the incident bleeding event taking the OAC for AF, deep vein thrombosis or pulmonary embolism. The mean age in the studies was 75 years. For 192 patients that continue OAC therapy after the incident bleeding event as was assumed in this model, a 90-day probability of recurrence of 10.1% was reported. A Danish observational study reported almost no change in the cumulative incidence of recurrent GI bleeding after three months (330). Therefore, people were permitted to remain in the serious GI event state for a maximum of one additional cycle, after which they move to the post-GI event state.

Probability of stroke in the absence of an NSAID

Incidence rates for stroke in anticoagulated patients reported in Chapter Four [4.4.2] were used to derive the three-months probability of stroke. The incidence rate of strokes that lead to hospital admission in anticoagulated patients without NSAID treatment was 7.12 (95% CI 5.46 to 9.30) per 1000 person years for anticoagulated patients with no concurrent NSAID use and a mean age of 72 years. From this incidence rate, a three-months probability of 0.18% was generated and applied in the model. This estimate was considered reasonable because it was in line with estimates from the ARISTOTLE trial (325). From the one-year probability in ARISTOTLE a three-months probability of stroke of 0.24% for apixaban and 0.26 for warfarin was calculated.

The risk of stroke increased with age in patients with and without OACs and was adjusted for the age groups over 79 years (507). An analysis of the Atrial Fibrillation Investigators database with patient level data from 12 RCTs comparing treatments in AF reported a 1.45 (95% CI 1.26 to 1.66) increase in stroke risk with every ten year increase in age for patients on OAC, antiplatelet or placebo (508). The data suggested that the HRs of the effect of age were similar between OAC, antiplatelet and placebo treatment. Lip et al. (2015) also demonstrated similar trends with age in stroke and thromboembolism Kaplan-Meier curves for patients with and without OACs (507). As a result, it was assumed that estimates from an AF population including patients with OACs, antiplatelets and non OAC users were appropriate to estimate the risk increase due to age in a population on OAC treatment.

Probability of death from stroke

A study by Komen et al. (2019), using Swedish routinely collected health data, investigated the 90-day mortality after 6017 strokes. Of those patients with warfarin or DOAC treatment 296 patients died resulting in a 90-day mortality of 17.6% (506). Compared with trial estimates from ARISTOTLE, this estimate seemed reasonable. The one-year probability of death after a stroke was found to be 30.1%, with over half of these deaths within 30 days after the event (325).

Probability of recurring stroke

The probability of experiencing a second stroke in the subsequent cycle after a stroke event was highest in the first three months after the incident event (509, 510). Data from the South London Stroke Registry, a register of all first stroke incidences in a defined population in London, reported cumulative mortality rates from Kaplan-Meier estimates. The three-months probability of a recurrent stroke was 2.1% (CI 95% 1.3% to 3.4%) (511).

A multivariate analysis found an increased risk of stroke recurrences with every ten years of age (HR 1.16, 95% CI 0.98 to 1.37) (509). This HR was applied to the probability of recurring stroke for the age groups over 79 years.

Probability of stroke in the post-stroke state

The probability of experiencing a stroke event in secondary prevention of stroke, hence the probability of transitioning from the post-stroke state to the stroke state was investigated in subgroup analyses of the RCTs comparing DOACs and warfarin according to a systematic review (512). The three most common OACs in the UK are apixaban, rivaroxaban and warfarin that were compared in the ROCKET-AF and ARISTOTLE trial. The subgroup analyses of patients with a history of stroke or transient ischaemic attack identified a three-months probability of stroke of 0.69% for rivaroxaban and warfarin (513), or 0.67% for apixaban and warfarin (514). A weighted average weighted by the number of patients with a previous stroke in each study was used to generate the three-months probability of experiencing a stroke in the post-stroke health state.

Probability of death in the post-stroke state

A subgroup analysis of the ARISTOTLE trial investigated the impact on death of the existence of a previous stroke or transient ischaemic attack in patients with warfarin and apixaban compared with patients without a previous stroke or transient ischaemic attack (514). Previous stroke or transient ischaemic attack was associated with an increased risk of death with a HR of 1.27 (95% CI 1.11 to 1.45). This study was seen as representative of the post-stroke population because it represented two of the three most common OACs in the UK (315).

Probability of death from the no adverse event or post-GI event state

OAC use is associated with a reduced mortality rate compared with non-use in an AF population (515). A meta-analysis of AF trials identified a mortality rate of 46.3 (CI 95% 39.9 to 53.2) per 1000 person years (516). Age varied with a mean or median age of 70 to 73 years between the trials. Age and a predominance of men were similar to the patient characteristics in Chapter Four chosen for the *de novo* state-transition model.

Table 5.3: Age-specific three-months probability of death by health state

Health state	Age group (years)				Source of base value[a]
	70–74	75–79	80-84	85-89	
No adverse event/ Post-GI event	**1.15%**	2.03%	3.59%	6.69%	Gomez-Outes (2016) (516)
Serious GI event	6.37%	**11.23%**	19.85%	37.01%	Komen (2019) (506)
GI discomfort	**1.15%**	2.03%	3.59%	6.69%	Gomez-Outes (2016) (516)
Symptomatic ulcer	2.47%	4.33%	7.59%	13.89%	Malmi (2016) (500)
Stroke	5.64%	9.96%	**17.60%**	32.81%	Komen (2019) (506)
Post-stroke	**1.70%**	3.01%	5.31%	9.90%	Easton 2012 (514)

[a]Base value is the value identified in the literature for the respective age group. The base value is in bold letters. The probability of death for other age groups was estimated from this value based on the relative risk of death by age in general population as per Office for National Statistics

Because patients age with time while they move through the cohort model, the probability of death was adjusted with age. The probability of death for other age groups was estimated from the identified base value according to the relative risk of death by age in the general population. The death probabilities in the general population by age group that were used as reference were derived from the ONS (494). The probability of death dependent on age groups in the model is summarised in Table 5.3.

Impact of absence/presence of NSAIDs on the probability of adverse events

This section reports the probabilities affected by the presence of NSAIDs in patients with OACs. The probability of these ADEs in the absence of NSAIDs was outlined before. The key feature of the HPE and the non-HPE cohort was that the transition probabilities only differed in the no adverse event health state. All other transition probabilities were the same. Hence, the probabilities affected by the presence of the NSAID are those from no

adverse event to any of the ADEs (stroke, serious GI events, symptomatic ulcer, GI discomfort). Risk ratios were used to reflect the effect of NSAIDs on the likelihood of the ADEs in the model and are summarised in Table 5.4.

Table 5.4: Summary of impact of the presence of the NSAID on transition probabilities

Description	Risk ratio (95% CI)	Distribution[a]	Reference
Increased likelihood of serious GI event	HR 2.96 (1.60; 5.46)	Lognormal	Chapter Four
Increased likelihood of GI discomfort	OR 2.12 (1.73; 2.58)	Lognormal	(517)
Increased likelihood of symptomatic ulcer	OR 1.70 (1.49;1.94)	Lognormal	(518)
Increased likelihood of stroke	HR 2.48 (1.36; 4.53)	Lognormal	Chapter Four

[a]Distribution for probabilistic analysis; GI: gastro-intestinal; HR: hazard ratio; NSAID: non-steroidal anti-inflammatory drug; OR: odds ratio

The impact of NSAIDs on death was assumed to be captured by the fatal events in the ADE health states. As a result, the probability of transitioning from the no adverse event state to the dead state was the same in the HPE and non-HPE cohort. From the cohort study in Chapter Four, hazard ratios were derived for the increased risk of serious GI events and stroke events. The cohort study only investigated events recorded in secondary care. Risk ratios for the increased likelihood of symptomatic ulcer and GI discomfort were defined as ADEs managed in primary care in the model specifications [Table 5.1]. Consequently, the risk ratios need to be derived from other sources. The results of the conducted literature search are reported in the Appendix K [Table K.2].

The impact of NSAIDs on the risk of symptomatic ulcer was investigated among other risk factors in a case-control study using the UK THIN database (518) that is broadly representative of the UK population (519). For 3914 cases of uncomplicated peptic ulcer disease and 9969 controls, an OR of 1.70 (95% CI 1.49 to 1.94) was reported for current NSAID use. The increased risk of GI discomfort with NSAIDs was investigated in a case-control study using health records from the UK QRESEARCH database (517). In the study period from 2000 to 2004, uncomplicated GI adverse events were more likely in patients with NSAID use. The risk of GI discomfort with different NSAIDs varies. The risk ratio of the

most common NSAID, with 59% of prescriptions in the UK, was used to represent the general population in the model (443). The OR of GI discomfort for naproxen compared to no naproxen use was 2.12 (95% CI 1.73 to 2.58) (517) and was used in the base case analysis. This study represented a UK population and has been used for the transition probabilities to GI discomfort/dyspepsia in a previous economic model that informed national guidelines in the management of osteoarthritis (492).

The increased transition probabilities in the HPE cohort are applied for a maximum of four cycles Table 5.5. At 12 months, it is assumed that the HPE is detected and resolved even if no ADE occurred. Annual reviews of medications are performed in GP practices that are thought to detect the hazardous prescription. After 12 months, the transitions probabilities in the HPE cohort are the same as in the non-HPE. This assumption was supported by the GPs [Appendix M].

Table 5.5: Summary of transition probabilities in the non-HPE and the HPE cohort

From ⬇	To ⬇	Three months probability (95% CI)		Distribution	Reference
		Non-HPE	HPE[b]		
No adverse event	Well	1—AOP	1—AOP	N/A	N/A
	Serious GI event	0.26% (0.24%; 0.27%)	0.76% (0.41%; 1.39%)	Dirichlet	Chapter Four
	GI discomfort	0.79% (0.59%; 0.99%)[a]	1.66% (1.36%; 2.01%)	Dirichlet	(326), risk ratio (517)
	Symptomatic ulcer	0.03% (0.03%; 0.04%)[#]	0.06% (0.05%; 0.07%)[c]	Dirichlet	(497), risk ratio (518)
	Stroke[c]	0.18% (0.13%; 0.23%)	0.44% (0.24%; 0.80%)	Dirichlet	Chapter Four
	Dead[c]	1.15% (0.99%; 1.32%)	1.15% (0.99%; 1.32%)	Dirichlet	(516)
Serious GI event	Serious GI event	10.10% (7.58%; 12.63%)[a,d]		Dirichlet	(504)
	Post-event	1—AOP		N/A	N/A
	Dead[c]	6.37% (4.77%; 7.96%)[a]		Dirichlet	(506)
GI discomfort	GI discomfort	10.10% (7.58%; 12.63%)[a,d]		Dirichlet	(504)
	Post-event	1—AOP		N/A	N/A
	Dead[c]	1.15% (0.99%; 1.32%)		Dirichlet	(516)
Symptomatic ulcer	Symptomatic ulcer	10.10% (7.58%; 12.63%)[a,d]		Dirichlet	(504)
	Post-event	1—AOP		N/A	N/A
	Dead[c]	2.47% (1.85%; 3.09%)[a]		Dirichlet	(500)
Post-GI event	Post-event	1—AOP		N/A	N/A
	Serious GI event	0.26% (0.24%; 0.27%)		Dirichlet	CPRD
	GI discomfort	0.79% (0.59%; 0.99%)[a]		Dirichlet	(326)
	Symptomatic ulcer	0.03% (0.02%; 0.04%)[c]		Dirichlet	(497)
	Stroke[c]	0.18% (0.13%; 0.23%)		Dirichlet	CPRD
	Dead[c]	1.15% (0.99%; 1.32%)		Dirichlet	(516)
Stroke	Stroke[c]	2.10% (1.30%; 3.40%)		Dirichlet	(511)
	Post-stroke	1—AOP		N/A	N/A
	Dead[c]	5.64% (4.23%; 7.05%)[a]		Dirichlet	(506)
Post-stroke	Post-stroke	1—AOP		N/A	N/A
	Stroke[c]	0.67% (0.50%; 0.83%)[a]		Dirichlet	(513, 514)
	Dead[c]	1.70% (1.49%; 1.94%)		Dirichlet	(514)
Dead	Dead	1		Fixed	-

[a]No measure of variance reported (ranges are based on +/-25% of the mean); [b]calculated from risk ratios reported in Table 5.4; [c]age dependent probability reported for age group 70-74; [d]applied for first cycle after event only, then assumes people move to post-GI event; AOP: all other probabilities

5.2.4.2 Health state resource use and costs

This section summarises the resource use and unit costs by health state. It was assumed that people in all health states (other than dead) in the HPE and the non-HPE cohort would receive their standard OAC treatment, which was not included in the health state costs. Temporary discontinuations of OACs after serious GI or stroke events were considered negligible and were not included. At the end of this section, a table reports resource use, unit costs and the total health state costs [Table 5.6].

No adverse event (HPE and non-HPE)

In the no adverse event state and the post-GI event state, no cost other than the resource use required for the hazardous prescription or the alternative treatment are generated. After consultation with the GPs, and in accordance with other published state-transition models (151, 450), paracetamol was used as the alternative non-hazardous prescription [Appendix M]. The cost of paracetamol for 90 days is included in the no adverse event state in the non-HPE cohort and in the post-GI event state. The cost of NSAIDs for 90 days is included in the no adverse event health state in the HPE cohort. NSAIDs were costed in the HPE cohort in the no adverse event health state until the HPE is resolved. After a maximum of 12 months, equivalent to four cycles in the no adverse event state of the state-transition model, it was assumed that the HPE would be detected and resolved and patients are switched to paracetamol. Hence, once the HPE is resolved the same resource use applies in the HPE and the non-HPE cohort in the no-adverse event state.

The resource use of treatment with the two drugs was generated as follows: The most commonly prescribed NSAID in England was naproxen with 59% of all oral NSAID prescriptions (443) and most frequently dispensed as 500mg tablets (96). The defined daily dose was 500mg, resulting in one naproxen tablet per day (520). Paracetamol was most commonly dispensed as 500 mg tablets (96) and its defined daily dose is 3000mg. The daily dose paracetamol was set to six tablets per day.

Resource use associated with GI discomfort

Previous work by Elliott et al. (2006) provided relevant disaggregated resource use data for GI discomfort (521). This paper was from 2006 and therefore expert opinion was sought from two GPs to confirm whether this still reflects current practice [Appendix M]. Resource use included costs of treatment according to UK guidelines and costs of one GP visit. The UK guideline recommendations for uninvestigated dyspepsia include a full dose proton-pump inhibitor (PPI) for four weeks once per day (522). Of the full dose PPIs listed in Appendix A of the guideline, omeprazole 20mg was the most frequently dispensed PPI in England as of June 2020 (96).

Resource use associated with symptomatic ulcer

For symptomatic ulcer, Elliott et al. (2006) reported resource use in the NHS (521). To reflect current practice, GPs were consulted to adjust the reported resource use to the current care practice [Appendix M]. The updated resource use included one diagnostic endoscopy, two GP visits, one outpatient visit, one H. pylori test and prescriptions of PPIs. For peptic ulcer, UK treatment guidelines recommend treatment of a full dose PPI for eight weeks (522).

Resource use associated with serious GI events

Healthcare costs in the serious GI event state were derived from a multi-centre, randomised trial assessing resource use of GI bleeding events in six UK hospitals (523). The study reported resource use of the initial admission to hospital and 28 days after discharge. It was the only study that reported post-discharge costs in addition to in-patient hospital resource use and included a comprehensive list of resource use items. The in-hospital admission incorporated resource use for intravenous fluids, lab tests, medication, fluids, blood component transfusion, endoscopies, surgery or radiological interventions to control bleeding, and ADEs. Resource use for post-discharge care for 28 days included re-admission, admission to nursing homes/residential care, A&E visits, outpatient clinic visits, GP visits etc. Unit costs were attached to resource use from the cost year 2012/2013 using

UK sources for reference costs. Inflated to 2018/2019 cost a serious GI event generated £3137.44 (SD 757.29).

Other published models calculated a mean cost of serious GI events based on the HRG unit costs (472, 524). This approach only accounts for resources related to the in-hospital episode and was therefore not used in this model. Unit costs for HRG codes are available from the NHS reference costs (525). The weighted average, weighted by activity of each HRG codes including 'gastrointestinal bleed' in their description as done by Dorian et al. (2014) (472), resulted in a mean cost of £1204.58 (SD 820.13).

Resource use associated with stroke and post-stroke

The majority of the patients in the UK prescribed an OAC have a diagnosis of AF (354). Stroke severity and costs are higher in patients with AF than in non AF populations (526). Resource use of stroke was therefore applied from an AF population. Luengo-Fernandez et al. (2015) reported the cost of stroke events by event severity from patients in the Oxford Vascular study (OXVASC), a prospective cohort study of all vascular events in Oxfordshire, UK (527). The reported acute and long-term cost estimates were previously used in two NICE technology appraisals (491, 524). Resource use for the acute event was assessed for the in-hospital episode (diagnostic tests) and post discharge care for 90 days after stroke event (re-admission, emergency transport, A&E visits, outpatient clinic visits, admission to nursing homes/residential care, GP/nurse visits etc). Resource use for long-term care after stroke include resources used for re-admission, inpatient diagnostic tests, emergency transport, A&E visits, outpatient clinic visits, admission to nursing homes/residential care and GP/nurse visits. Costs were reported for the first 90 days of each event (acute event costs) and long-term costs were calculated as the mean annual excess costs of five years after the event. The reported acute event costs were used to generate costs for the stroke health state and the excess long-term annual costs were used to generate costs in the post-stroke health state of this model. The annual costs were divided by four to generate three-months costs. Costs were reported by stroke severity in Luengo-Fernandez et al. (2015). To generate input parameters for this study, a weighted average based on the severity

distribution of strokes in patients with apixaban and warfarin in the ARISTOTLE trial was (477). The costs (cost year 2008/2009) were inflated to the cost year 2018/2019.

Unit costs

Unit costs associated with the non-drug related resources, for instance, GP visits or outpatient visits, were used from the PSSRU: Unit costs of health and social care (285), and NHS reference costs database (525). At the time of analysis, the most up to date unit costs were used (2018/2019) from the PSSRU: Unit costs of health and social care (285), and NHS reference costs database (525). Unit costs of drug treatment were used from the NHS electronic drug tariff (528).

Table 5.6: Summary of resource use, unit costs and total health state cost

Resource use				
Health state	Resource item	Amount	Distribution	Source
No adverse event HPE model	One naproxen 500mg tablet daily, 90 days	1	Fixed	(96, 443, 520)
No adverse event non-HPE model	Six paracetamol 500mg tablets daily, 90 days	1	Fixed	(96, 520)
GI discomfort	One GP visit	1	Fixed	(521), expert opinion
	Omeprazole 20mg/day, 28 days	1	Fixed	(96, 522)
Symptomatic ulcer	Endoscopy	1	Fixed	(521), expert opinion
	GP visit	2	Fixed	(521), expert opinion
	Omeprazole 20mg/day, 56 days	1	Fixed	(521), expert opinion
	Out-patient visit	1	Fixed	(521), expert opinion
	H. pylori test	1	Fixed	(521), expert opinion
Serious GI event	In-hospital episode and 29 days post discharge	1	Fixed	(523)
Stroke	In-hospital episode and 90 days post discharge care	1	Fixed	(477, 527)
Post-GI event	No resource use	N/A	N/A	Assumption
Post-stroke	Total annual excess healthcare resource use	1	Fixed	(477, 527)
Unit Costs				
Unit cost item		Cost	Distribution	Source
28 naproxen 500mg tablets		£2.93	Fixed	NHS drug tariff[e]
100 paracetamol 500mg tablets		£3.53	Fixed	NHS drug tariff[e]
28 omeprazole 20mg tablets		£1.18	Fixed	NHS drug tariff[e]
GP visit		£33.00	Fixed	PSSRU 2019[c]
Outpatient visit (gastro-enterology)		£141.00	Fixed	NHS reference costs[d]

H. pylori test	£8.00	Fixed	NHS reference costs[d]
Diagnostic endoscopic upper gastrointestinal tract procedures (outpatient procedure)	£354.00	Fixed	NHS reference costs[d]
In-hospital episode - serious GI event	£3,137.44 (SD 247.61)	Gamma	(523)
In-hospital episode - stroke	£10645.86 (SD 12048.60)	Gamma	(527)
Excess healthcare post-stroke	235.20 (SD 133.26)	Gamma	(527)
Total health state cost			
Health state	Cost	Distribution	Source
No adverse event (HPE model)	£7.68	Gamma[b]	NHS drug tariff[e]
No adverse event (Non-HPE model)	£19.06	Gamma[b]	NHS drug tariff[e]
GI discomfort	£34.18	Gamma[b]	Elliott 2006; PSSRU[c]; NHS reference costs[d]; NHS drug tariff[e]
Symptomatic ulcer	£523.52	Gamma[b]	Elliott 2006; PSSRU[c]; NHS reference costs[d]; NHS drug tariff[e]
Serious GI event	£3137.44 (SD 757.29)	Gamma	(523)
Post-event	£19.06	Gamma[b]	NHS drug tariff[e]
Stroke	£10645.86 (SD 12048.60)	Gamma	(477, 527)
Post-stroke	£235.20 (SD 133.26)	Gamma	(477, 527)
Dead	£0.0001[a]	N/A	Assumption

[a]A cost of £0.0001 was used in TreeAge Pro Healthcare 2021 because an input cost of £0 is not possible; [b]when no standard deviation was available, the mean itself was used in the Gamma distribution; [c]Personal Social Services Research Unit 2019 (285); [d]NHS reference costs database 2018/2019 (525); [e]NHS electronic drug tariff (528); GI: gastro-intestinal; GP: general practitioner; NHS: National Health Service

5.2.4.3 Health related quality of life

We used published estimates of health related quality of life for each health status in the state-transition model. Preferentially, we used utility estimates derived from UK populations using EQ-5D-3L (529) or EQ-5D-5L (530) questionnaires, with UK tariffs. The EQ-5D-3L were used in preference to 5L when presented with a choice as recommended in the NICE reference case (531). If the utility estimates were derived from non-UK populations, UK tariffs were used where possible. The exact estimates applied in the model adjusted to the three-months cycle length are reported in Table 5.7.

Utilities associated with no adverse event/post-GI event

The Euro heart survey collected data on quality of life in patients with AF, the most common diagnosis associated with the use of OACs, in 35 European countries from 2003 to 2004 (532). The study used the EQ-5D-3L questionnaire and results were translated into utilities using UK tariffs for the different dimensions. The Euro heart survey estimates have been used in other peer-reviewed models (485, 533) and in a NICE technology appraisal (524, 534). Mean age of participants in the survey was slightly lower than the start age of this model with a mean age of 66 years. Percentages of men (60%) and women (40%) were very similar to the proportions in the population in the *de novo* state-transition model. The majority of patients (68%) were prescribed OACs. A mean utility of 0.779 (SD 0.253) was reported. This utility was used to calculate QALYs generated in the no adverse event health state in the HPE and non-HPE model and in the post-GI event state. To account for quality of life decreasing with age, the utility was multiplied by the ratio of the utility for a given age range relative to a reference age (66–70 years), based on general population utilities estimate. Estimate for the general population were derived from Ara et al. (2011) that pooled EQ-5D-3L data from four consecutive Health Surveys for England from 2003 to 2006. The Health Survey for England is conducted annually and in a random sample of the population living in private households. The method to account for the decreasing utility with age, was also used in a HTA (524). The pooled dataset included 41174 respondents with completed EQ-5D-3L questionnaires (437). The relative relationship between the general population estimate from the for the age group 66-70 years in Ara et al. (2011) (437), and the utility estimate from Berg et al. (2010) (532), was used to calculate age dependent utilities.

Utilities associated with GI discomfort

The reduction in health related quality of life due to GI discomfort was assumed to be a constant absolute decrement relative to having no GI discomfort. A UK-based RCT recruited people aged 18 to 65 presenting in primary care with dyspepsia (535). This study reported a baseline utility of 0.74 for 679 people based on the EQ-5D-3L questionnaire. The mean age of the sample is not reported in the paper, however based on the general population utility values for the 55-60 age group (0.8222) (437), this would be a decrement of 0.08. A

small study in Malaysia asked people with headaches to complete the EQ-5D-3L and then compared the derived utility values by whether or not people also had dyspepsia (536). The 24 people with both conditions had a mean utility value of 0.82 (SD 0.18), and the 69 people, who only had headaches, had a mean utility value of 0.90 (SD 0.16). The utility decrement for dyspepsia from this study is also 0.08. A 2001 study by Groeneveld et al. (2001) included a time trade-off study and reported that the QALY loss associated with dyspeptic symptoms was 0.09 (537). The estimates from all three of these sources are similar. Because two studies reported a utility decrement of 0.08, this value was used in the model.

Utilities associated with symptomatic ulcer

The reduction in health related quality of life due to symptomatic ulcer was assumed to be a constant absolute decrement relative to having no symptomatic ulcer. No study directly measured a utility estimate for symptomatic ulcer in a UK population. Three studies were found that reported potential utility estimates (437, 538, 539). A study from 2001 by Groeneveld et al. assessed utilities for 73 patients with a diagnosis of peptic ulcer and dyspeptic symptoms in the US (537). Groeneveld et al. (2001) conducted interviews with patients to generate utilities for dyspepsia and peptic ulcer using time trade-off methods. However, the methods for the utility assessment were not described transparently. They reported a QALY decrement associated with a peptic ulcer, including peptic ulcer bleeding, of 0.11 (range from 0.05 to 0.19). In the second study by Maetzel et al. (2003), quality of life associated with symptomatic ulcer in arthritis patients was assessed for a Canadian cost-effectiveness analysis of NSAIDs (538). A survey of 60 randomly selected participants was conducted. Rating scale and standard gamble methods were used to identify utility weights for symptomatic ulcer compared with arthritis without ADEs. However, the exact methods were not reported transparently. The raw data, any participant information or calculations were not presented. For symptomatic ulcer, a utility multiplier of 0.5523 was found with a range from 0.4651 to 0.6511 for patients with arthritis and symptomatic ulcer relative to arthritis patients without the event over three months. Ranges were not provided as standard deviation or confidence intervals, and it was unclear what the ranges described. Nevertheless, the utility weights were used in the economic model informing the National Guidelines for osteoarthritis in the UK (492) and other published models (450,

451). The third study by Ara et al. (2011), used EQ-5D-3L data from the Health Survey for England (437). The large survey collects information on different conditions and links it with the EQ-5D-3L. The survey is based on self-report of patients and their perception of potential conditions they have. The study does not report symptomatic ulcer as one of the conditions but, following a discussion with the patient representative, symptoms from GI discomfort would most likely be similar to bowel/colon symptoms reported in the survey.

In this model, the utility reported in the third study was used. It was the only English study, was the most recent and used the preferred EQ-5D-3L questionnaire. This was also the only study transparently reporting the methods how the utility estimates were derived and contained the largest sample size. Despite the strong assumption that utilities from symptomatic ulcer are similar to those with colon/bowel symptoms, this was considered the best available evidence. For the 70-75 age group, the general population utility value was 0.7790 whereas for the group with bowel complaints it was 0.6455 – a decrement of 0.1335 (437). The fact that this decrement is comparable to the three-months decrement identified by Groenveld et al. (2001) (0.11) and the derivative of this study used in the economic evaluation of the PINCER trial (0.13), showed that the utility estimate for colon/bowel conditions was comparable to that of earlier studies.

Utilities associated with serious GI events

The reduction in health related quality of life due to serious GI events was assumed to be a constant absolute decrement relative to having no serious GI event. A recent systematic literature review of utility decrements associated with bleeding events in people taking dual antiplatelet therapy reported that for gastrointestinal bleeds the decrements ranged from 0.005 to 0.016 (540). A paper published in 2015 by Campbell et al. (2015) reported utility values derived from EQ-5D-3L for a cohort of 936 participants who were admitted to one of six UK university hospitals with an upper GI bleed (541). Another study investigating the utility of GI bleeding in an anticoagulated population reported utilities from the ENGAGE-AF TIMI RCT that compares edoxaban and warfarin treatment (542). The EQ-5D-3L was used and utilities assessed every three months at routine check-up visits. US tariffs were applied.

In the state transition model in this thesis, the same utility decrement as used in the previous PINCER model (0.18) was used (227). The mean utility was calculated over three months based on the temporary utilities in Spiegel et al. (2005) for inpatient treatment for ulcer haemorrhage (0.46 for ten days) (543) and the dyspepsia after the inpatient stay (0.87 for 80 days). This decrement of the three-months utility was found to better represent the utility of serious GI events over three months, than other studies. In Campbell et al. (2015), participants completed the EQ-5D-3L 28 days after they were admitted. Because of this delay in reporting, this does not capture the utility associated with the serious GI event itself (541). The utility values from the RCT were not used in the base case because they used US tariffs and utility was only assessed every three months at routine check-up visits not specifically after the event (542).

Utilities associated with stroke

The reduction in health related quality of life due to stroke was assumed to be a constant absolute decrement relative to having no stroke. The OXVASC study collected utilities of all stroke events (544). From 2002-2007, EQ-5D-3L utilities were available for 445 patients with stroke and 381 controls, with a mean age of 75 years. Other OAC models used in HTAs submitted to NICE (524, 534) used utilities from a time trade-off study published in 2001. From 57 patients, utilities for mild and severe stroke were assessed showing highly skewed utilities. The state-transition model by Sterne et al. (2015) used the utility for severe stroke for all severities of strokes identified in their model (524). This overestimates the disutility of stroke neglecting the impact of the more common minor and moderate strokes (544). Edwards et al. (2011) applied the utility for the different severity levels in their model submitted as part of an HTA (534). In the *de novo* state-transition model in this thesis, the utility from the OXVASC study was used (544). This was the largest and most recent study and the only English study that used the preferred utility elicitation method. Both HTAs assumed that the disutility from Robinson et al. (2001) persisted over the cycle length of three months. The more recent data from the OXVASC study, however, clearly showed that the utility after one months was much higher than this (544). The reported utility at one month after the stroke were assumed to capture the average utility over the first three month after the event. The utility difference between stroke compared with the controls

was of 0.22 (95% CI 0.18 to 0.26). This difference was used as the three-months decrement applied to the age dependent utility.

Utilities associated with the post-stroke state

The reduction in health related quality of life due to long-term consequences of stroke was assumed to be a constant absolute decrement relative to no prior stroke. The OXVASC study reported not only utilities at one month after the stroke event but also long-term utilities (544). The study reported a utility difference between patients with a previous stroke and patients in the control group of 0.18 (95% CI 0.13 to 0.23, $p < 0.001$).

The only comparable study, assessing the long-term utilities impact of stroke, was a prospective cohort study by Haacke et al. (2006) in Germany (545). The study investigated long-term quality of life in patients that experienced a stroke . The utility estimates were used in an earlier HTA of rivaroxaban (534). Overall, 77 patients completed the EQ-5D-3L questionnaire and utility estimates were generated using German tariffs. A utility of 0.68 (SD 0.34) and 0.64 (SD 0.33) was reported for patients aged 65-75 and over 75 years, respectively.

As described earlier, utility estimates were preferably used from UK studies. Therefore, the more recent English utility difference estimated in the OXVASC study was used as the three-months utility decrement for post-stroke.

Disutility of correcting HPE

Correction of this HPE involved the discontinuation of the NSAID and a switch to paracetamol therapy. According to the GPs and pharmacists, this was the preferred action to resolve the HPE in practice [Appendix M: Table M.2]. It was assumed that this switch did not affect utilities. Alternative treatments were considered to have the same treatment effect (i.e., the same pain relief as NSAIDs). The same assumption was previously made in a HPE model involving NSAID use in the elderly (151). The patient representative highlighted that this might not be the case because paracetamol has a slightly smaller effect on pain relief. This could, however, not be included in the model because no quantitative

data exist on the disutility of paracetamol treatment compared with NSAID treatment. As a result, the same utilities were used in the HPE and non-HPE cohort in the no adverse event state.

Table 5.7: Utility estimate of the reference group health state (no adverse drug event) and utility decrements associated with the adverse drug events

Health state	Utility score	Source/assumption	Distribution
Reference group health state[a]			
No adverse event	0.78 (SD 0.25)	EQ-5D-3L data from Euro heart survey of patients with atrial fibrillation (532)	Beta
Post-GI event	0.78 (SD 0.25)	Same as for no adverse event	Beta
Adverse drug event related utility decrements applied to reference group[b]			
GI discomfort	-0.08 (± 25%)[c]	(535, 536)	Gamma[e]
Symptomatic ulcer	-0.14 (-0.21; -0.07)[d]	Estimate for colon/bowel symptoms from the Household Survey for England (437)	Gamma
Serious GI event	-0.18 (±25%)[c]	Decrement based on 10 days in hospital and 80 days post discharge (149)	Gamma[e]
Stroke	-0.22 (-0.26; -0.18)[d]	(544)	Gamma
Post-stroke	-0.18 (-0.23; -0.13)[d]	(544)	Gamma
Dead	0	Utility of 0 for dead	N/A

[a]Age dependent utility; the reported estimate is from the reference age group of 66-70 years; adjustment to age with the relative relation to estimates from the reference age group in the Household Survey for England (437); [b]decrements were assumed to last for three months; [c]no uncertainty level provided and an arbitrary range of ±25% was assumed appropriate to reflect uncertainty; [d]95% confidence intervals; [e]when no standard deviation was available, the mean itself was used in the Gamma distribution

5.2.5 Analysis

The state-transition model was built in *TreeAge Pro Healthcare 2021* (546). The model was populated with probability, cost and health status data [5.2.4] to allow the generation of the point estimates and distributions of discounted outcomes (QALYs) and NHS costs in a cohort exposed to the HPE, and a cohort not exposed. The input parameters identified in section 5.2.4 were fit uncertainty distributions. Distributions appropriate for the different input parameters were chosen based on recommendations by Briggs et al. (2006) (547). A gamma distribution was used for costs and for absolute utility decrements, a beta distribution for utility values and either a beta distribution or Dirichlet distribution was used for transition probabilities. Hazard ratios were sampled from a lognormal distribution. Total cost of the health states per cycle were fitted a gamma distribution because it reflected the

natural skewness of healthcare costs. Single resource use and unit cost estimates were considered fixed values. The beta distribution is bound by zero and one and was therefore used for probabilities. However, beta distributions were only appropriate if there was one pathway leaving the health state (4). Hence, if the probability described a binary event or no event situation, the beta distribution was used. If multiple pathways were possible in the model, such as the probabilities leaving the no adverse event state, the beta distribution was not appropriate. A multivariate generalisation of the beta distribution that can incorporate multiple pathways, the Dirichlet distribution, was used then. If no measure of uncertainty was available for the beta or gamma distribution, the conservative assumption was made that the standard deviation defining the distribution equals the mean as recommended by Briggs et al. (2006) (4).

Base case analysis

In the base case analysis, probabilistic estimates were generated. The specifications of the model were reported in Table 5.1. The probabilistic analysis was based on 10000 samples. A random seed of 345 was used. Results were reported as incremental costs and incremental QALYs per patient. The results were scaled up to population level to demonstrate the economic burden of the HPE type for NHS England. The NHS dashboard reports the number of patients with concomitant OAC and NSAID treatment in England by quarters of the financial year (55). The estimate for the fourth quarter 2019/2020 (last available quarter at time of analysis) was used as a prevalence estimate of the HPE. The number of patients with the HPE was applied in the probabilistic model in *TreeAge Pro Healthcare 2021* to generate the incremental costs and QALYs the patients with the HPE generate compared with the hypothetical scenario that they were not exposed to the HPE. The results are presented in a scatter plot.

Deterministic analysis

A deterministic analysis was conducted using the expected mean values without the uncertainty distributions applied in the probabilistic analysis. One deterministic analysis was conducted using a lifetime horizon as in the base case analysis. To test if the effects of

changing time horizons are comparable in the deterministic model and the probabilistic model, an additional deterministic analysis with a ten-year time horizon was run.

5.2.6 Sensitivity analysis

Various assumptions were changed to test robustness of the results in the probabilistic model. A summary of the sensitivity analyses and the values that changed is reported at the end of this section.

Assumptions on model specifications

The time horizon of the model was varied to understand how this impacts the results. First, a shorter time horizon was tested. Patients were not only followed until they reached 100 years of age or died but for five, ten and 20 years. A five-year time horizon was used in the state-transition models applied in the PINCER economic evaluation (149). Results were thought to be more comparable if the same time horizon was chosen. A subgroup analysis investigated the impact of a change in the start age to 80 years to see if the intervention effect is age dependent. For the base case, costs and outcomes were discounted. In sensitivity analysis, a discount rate of 0% was tested, compared to 3.5% in the base case analysis.

Assumptions on resolving the HPE

In the base case, it was assumed for the HPE state-transition model that the NSAID was removed after a maximum of one year in no adverse event with the increased probabilities in the presence of the NSAID. This was used to account for the fact that practices might identify the HPE and remove the NSAID as part of routine monitoring. The cohort in the HPE model is therefore only for a maximum of four cycles at a higher risk of the ADEs. From the fifth cycle on, the same probabilities were applied as in the non-HPE model. In the sensitivity analysis, the cost-effectiveness analysis was run without this assumption, such that patients could experience the HPE over a longer time period. This is referred to as the 'No HPE correction scenario'.

Assumptions on input parameter choices

The risk ratios that describe the increase in the likelihood of serious GI events and stroke were used from the cohort study in Chapter Four. The results of the sensitivity analysis conducted in Chapter Four showed varying results when specific assumptions were changed. To identify the impact these changes have not only on the HR but also on the incremental costs and QALYs, these were investigated in this chapter. The first assumption tested was changing the grace period added to the calculated prescription length and exposure periods from 30 days to 60 days (HR Chapter Four – 60-day grace). The second assumptions tested was using IPTW instead of PSM (HR Chapter Four – IPTW). The third assumption tested was a change in the washout window applied for NSAID users from three months to six months (HR Chapter Four – NSAID washout 6 months). The fourth assumptions tested was and additional exclusion criterion restricting the cohort to patients that have not had the respective outcome before the index date (HR Chapter Four – Exclusion of patients with outcome before index). In addition to varying assumptions in the analysis conducted in Chapter Four, the impact of using the HRs identified in the subgroup analysis of the ARISTOTLE trial were also tested (HR – RCT subgroup analysis) (309). A detailed comparison of the subgroup analysis of the ARISTOTLE trial and the observational cohort study conducted for this dissertation is presented in in Chapter Four [4.5.2].

The risk ratio of the impact of NSAIDs on the incidence of GI discomfort was also varied in sensitivity analysis. An alternative source, a systematic review and meta-analysis of RCTs, reviewing studies published until 1997, that defined dyspepsia as diagnoses of epigastric pain, dyspepsia, discomfort, nausea, bloating and anorexia reported a risk ratio of 1.19 (95% CI 1.03 to 1.39) for NSAID users (548). The RCT data were not used for the base case because the estimate was older, and the meta-analysis did not represent the NSAID prescribing patterns in the UK. The utility associated with symptomatic ulcer was also varied. In the base case analysis, estimates in a population with bowel-disease were used as a proxy for symptomatic ulcer. In sensitivity analysis, the estimates reported for arthritis patients with and without symptomatic ulcer were used (538).

Table 5.8: Descriptions of sensitivity analyses conducted in the probabilistic model

Scenario analysis	Base case parameter value	Scenario analysis parameter value	Source
No HPE correction	HPE related changes in transition probabilities and costs applied for first 4 cycles in 'no adverse event' state. Then the non-HPE values are used.	HPE related changes are applied as long as patients stay in 'no adverse event' state in the HPE model.	N/A
Time horizon: 20 years	Lifetime (120 cycles)	20 years (80 cycles)	N/A
Time horizon: 10 years	Lifetime (120 cycles)	10 years (40 cycles)	N/A
Time horizon: 5 years	Lifetime (120 cycles)	5 years (20 cycles)	N/A
Start age: 80 years	70 years of age	80 years of age	N/A
Discount rate: 0%	3.5% discount rate	0% discount rate	N/A
HR Chapter Four - 60-day grace	HR serious GI event 2.96 (1.60; 5.46); HR stroke 2.48 (1.36; 4.53)	HR serious GI event 2.53 (1.66; 3.86); HR stroke 1.79 (1.13; 2.84)	Chapter Four
HR Chapter Four - IPTW	HR serious GI event 2.96 (1.60; 5.46); HR stroke 2.48 (1.36; 4.53)	HR serious GI event 2.47 (1.40; 4.34); HR stroke 1.71 (0.95; 3.08)	Chapter Four
HR Chapter Four - NSAID washout 6 months	HR serious GI event 2.96 (1.60; 5.46); HR stroke 2.48 (1.36; 4.53)	HR serious GI event 3.81 (2.04; 7.09); HR stroke 3.11 (1.61; 6.01)	Chapter Four
HR Chapter Four - Exclusion of patients with outcome before index	HR serious GI event 2.96 (1.60; 5.46); HR stroke 2.48 (1.36; 4.53)	HR serious GI event 3.37 (1.81; 6.25); HR stroke 2.33 (0.97; 5.59)	Chapter Four
HR RCT dataset	HR serious GI event 2.96 (1.60; 5.46); HR stroke 2.48 (1.36; 4.53)	HR serious GI event 1.26 (0.59; 2.73; HR stroke 1.44 (0.63; 3.28)	Dalgaard 2020 (309)
HR GI discomfort	HR GI discomfort 2.12 (1.73; 2.58)	HR GI discomfort 1.19 (1.03; 1.9)	Strauss 2000 (548)
Utility of symptomatic ulcer	Utility decrement of 0.13	Utility decrement of 0.37	Maetzel 2003 (538)

HR: hazard ratio with 95% confidence interval; GI: gastro-intestinal; IPTW: inverse probability of treatment weighting; RCT: randomised controlled trial

5.3 Results

5.3.1 Base case analysis

The results of the analysis are reported in Table 5.9. The base case analysis reported incremental costs of £244 (2.5% to 97.5% credible interval -£149 to £1073) and incremental QALYs of -0.04 (2.5% to 97.5% credible interval -0.17 to 0.05). Both the deterministic and probabilistic analyses estimate higher costs and less QALYs for people who are co-prescribed NSAIDs. The results of the deterministic and the probabilistic analysis produced similar results for a lifetime horizon and a ten-year time horizon.

Table 5.9: Costs and QALYs accrued in the HPE and the non-HPE cohort

Analysis	Costs generated per patient, £		QALYs generated per patient, QALY		Incremental	
	HPE	Non-HPE	HPE	Non-HPE	Costs, £	QALYs
Base case analysis[d]						
Probabilistic analysis	2090 (464; 5731)	1846 (383; 4922)	5.80 (5.28; 6.03)	5.84 (5.42; 6.09)	244 (-149; 1073)	-0.04 (-0.17; 0.05)
Deterministic analyses						
Time horizon: lifetime	2080	1856	5.80	5.84	224	-0.04
Time horizon: 10 years	1602	1386	4.80	4.83	216	-0.03
Sensitivity analyses in Probabilistic model [d]						
No Camacho correction	2774 (582; 8489)	1846 (383; 4922)	5.70 (4.94; 5.96)	5.84 (5.42; 6.09)	928 (-703; 4186)	-0.14 (-0.56; 0.11)
Time horizon: 20 years	2070 (460; 5683)	1826 (379; 4874)	5.77 (5.26; 6.00)	5.81 (5.39; 6.05)	244 (-149; 1073)	-0.04 (-0.17; 0.05)
Time horizon: 10 years	1614 (344; 4558)	1379 (264; 3735)	4.80 (4.46; 4.96)	4.83 (4.57; 5.00)	235 (-150; 1050)	-0.03 (-0.14; 0.04)
Time horizon: 5 years	1026 (196; 3045)	815 (132; 2306)	3.06 (2.90; 3.14)	3.08 (2.97; 3.16)	211 (-158; 1010)	-0.02 (-0.08; 0.03)
Start age 80	1369 (254; 4006)	1105 (195; 3093)	2.87 (2.65; 3.01)	2.91 (2.73; 3.07)	265 (-140; 1195)	-0.04 (-0.13; 0.03)
Discount rate: 0%	2628 (594; 7135)	2364 (499; 6288)	7.15 (6.46; 7.47)	7.20 (6.63; 7.54)	264 (-146; 1131)	-0.05 (-0.21; 0.06)

Analysis	Costs generated per patient, £		QALYs generated per patient, QALY		Incremental	
	HPE	Non-HPE	HPE	Non-HPE	Costs, £	QALYs
HR Chapter Four - 60-day grace	1956 (431; 5313)	1846 (382; 4922)	5.82 (5.36; 6.04)	5.84 (5.42; 6.09)	110 (-203; 645)	-0.02 (-0.11; 0.06)
HR Chapter Four - IPTW	1959 (429; 5379)	1846 (382; 4920)	5.82 (5.36; 6.05)	5.84 (5.42; 6.09)	114 (-210; 690)	-0.02 (-0.11; 0.06)
HR Chapter Four - NSAID washout 6 months	2301 (516; 6283)	1845 (385; 4924)	5.78 (5.16; 6.02)	5.84 (5.42; 6.09)	456 (-84; 1779)	-0.06 (-0.27; 0.04)
HR Chapter Four - Exclusion of patients with outcome before index	2098 (457; 5840)	1846 (382; 4922)	5.81 (5.30; 6.03)	5.84 (5.42; 6.09)	252 (-166; 1249)	-0.04 (-0.17; 0.05)
HR - RCT dataset[b]	1888 (417; 5058)	1845 (385; 4908)	5.83 (5.38; 6.05)	5.84 (5.42; 6.09)	42 (-260; 506)	-0.01 (-0.11; 0.07)
HR GI discomfort[a]	2095 (464; 5748)	1846 (385; 4915)	5.80 (5.28; 6.03)	5.84 (5.42; 6.09)	249 (-150; 1099)	-0.04 (-0.17; 0.05)
Utility symptomatic ulcer[c]	2090 (464; 5731)	1846 (383; 4922)	5.80 (5.28; 6.03)	5.84 (5.42; 6.09)	244 (-149; 1073)	-0.04 (-0.17; 0.05)

[a]Hazard ratio from Strauss et al. (2002) (548); [b]hazard ratio from analysis of the ARISTOTLE trial dataset (309); [c]utility from Maetzel et al. (2003) (538); [d]all expected values describe the mean with 2.5% and 97.5% percentile, the credible interval; HPE: hazardous prescribing event (Concomittant OAC and NSAID use); HR: hazard ratio; GI: gastro-intestinal; IPTW: inverse probability of treatment weighting; NSAID: non-steroidal anti-inflammatory drugs; RCT: randomised controlled trial; QALY: quality-adjusted life-years

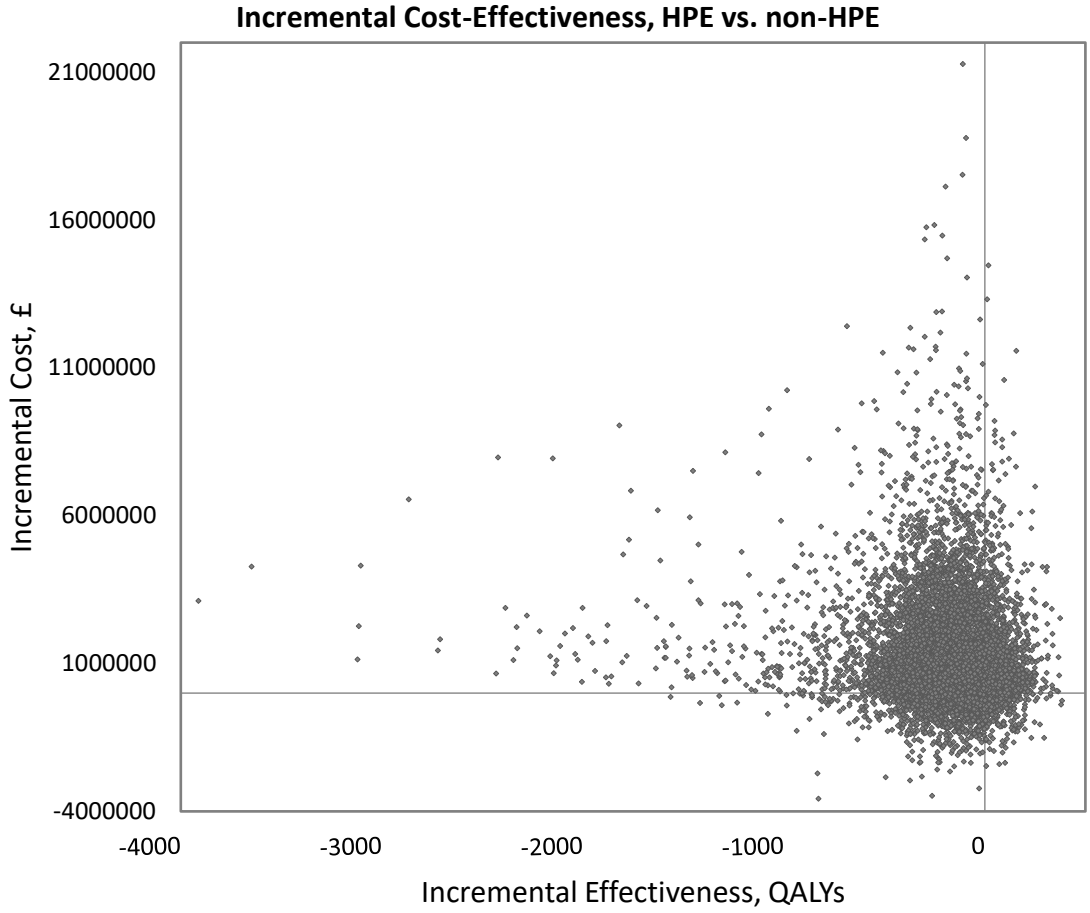
5.3.2 Sensitivity analysis

Results of the sensitivity analysis are reported in Table 5.9. Releasing the assumption that HPEs are detected and resolved after at least one year, generated 3.5 times more QALYs and 3.8 times higher costs than the base case. With increasing time horizon, the incremental costs and QALYs increased as well. After ten years (40 cycles), 50.7% of the patients died in the non-HPE cohort and 51.7% in the HPE cohort (results not reported). At 100 years of age (lifetime horizon), all patients were dead in both scenarios. Changing the source for the HR of GI discomfort, the utility of symptomatic ulcer or applying HRs for stroke and serious GI events from an analysis that excludes patients with the respective outcome prior to the index date did not impact the incremental costs and QALYs. The smaller HRs for serious GI events and stroke (IPTW and 60-day grace period), as well as the HRs from the subgroup analysis of the ARISTOTLE trial reduced the incremental costs and QALYs. Excluding patients with NSAID use up to six months before index date in Chapter Four increased the HR for stroke and serious GI events and resulted in higher incremental costs and QALYs.

Incremental costs and QALYs at population level

According to the NHS dashboard, 13399 patients are exposed to the HPE with a concomitant NSAID and OAC prescription (336). Over their lifetimes, these patients are expected to incur over £3 million more and generate more than 500 fewer QALYs than would be expected without exposure to the HPE. The state-transition model assumed that the HPE is corrected after a maximum of one year. The probabilistic results for 10000 iterations are reported in Figure 5.2. Over 70% of the data points are in the North-West quadrant, resulting in higher costs and lower QALYs for patients with the HPE.

Figure 5.2: Scatter plot for the probabilistic analysis of 13999 patients with and without the HPE with 10000 iterations



5.4 Discussion

5.4.1 Principal findings

In this chapter, a state-transition model was constructed modelling treatment pathways related to ADEs associated with NSAID use in anticoagulated patients [Dissertation Objective Four]. The key ADEs associated with NSAID use in anticoagulated patients were GI discomfort, symptomatic ulcer, serious GI events and stroke. Where possible, input parameters were generated from data of the cohort study conducted in Chapter Four, e.g., the increased likelihood of serious GI events and stroke associated with NSAID use. The probabilistic analysis suggests that exposure to the HPE generates £244 (2.5% to 97.5% credible interval -£149 to £1073) of incremental costs to NHS England and -0.04 (2.5% to 97.5% credible interval -0.17 to 0.05) QALYs over a lifetime, assuming the HPE is detected

and resolved after a maximum exposure of 12 months [Dissertation Objective Five]. Even under the assumption that the HPE is automatically identified and resolved after a maximum of 12 months, the incremental costs increased and the incremental QALYs decreased with longer time horizons. This suggests that even if the HPE is resolved after a maximum of one year, the long-term consequences of the HPE, in this example mainly the long-term care costs and increased mortality after a stroke event, increase with time. Changing the start age to the age of 80 increased the incremental costs and QALYS compared to the start age of 70 years. This could be a result of the decreasing utility with age. Utility decrements were applied as absolute decrements; hence, the relative decrease is higher when the base utility is smaller.

The correction of the HPE after a maximum of 12 months sojourn time in the no adverse event state with the HPE was found to generate less than a third of the incremental costs and QALYs compared with the sensitivity analysis without this assumption. The HPE correction was assumed to acknowledge the potential that HPEs were detected in annual patient reviews in the practices. However, the HPE was associated with higher costs and fewer QALYS even with the conservative adjustment.

While the incremental costs and QALYs were robust to most changes tested in sensitivity analysis, the results were sensitive to the source for the HRs for serious GI events and stroke. Serious GI events and strokes are the ADEs associated with the highest costs and loss of quality of life in this chapter. Changing the impact of the HPE on the risk of these events was expected to affect the results in this chapter. Depending on the method applied in Chapter Four to generate the HRs, the extend of the effect of the HPE on incremental costs and QALYs changed. Smaller HRs, such as in the analysis using IPTW or a 60-day grace period, resulted in smaller incremental effects. When non-significant HRs were applied from RCT subgroup analysis, the incremental economic impact was smaller. Reasons why the results from the RCT subgroup analysis might have been diluted were discussed in Chapter Four [4.5.2].

Exposure to the HPE was associated with a substantial financial burden to the healthcare provider. If prevalent cases of the HPE were avoided and these anticoagulated patients

would not have had a concomitant NSAID prescription, NHS England could save about £3 million in healthcare costs and health related quality of life could be improved by over 500 additional QALYs. While this estimate describes all prevalent cases, future research could investigate incident cases per year to estimate the yearly economic burden to NHS England.

5.4.2 Comparison to other studies

Only one study was found that used a cohort state-transition model to estimate the impact of NSAID use in anticoagulated patients (152). Foy et al. (2020) assessed the economic impact of various NSAID related HPEs including NSAID use in anticoagulated patients to inform a cost-effectiveness analysis of an intervention. The same state-transition model was used by Foy et al. (2020) for all NSAID related HPEs. Hence, the model structure was not specific to the population of anticoagulated patients. The results were not reported for the individual models, so no direct comparison of the results could be made. The model structure was based on a previously published model in a clinical guideline for osteoarthritis from 2008 (492). In addition to the ADEs modelled in this chapter, Foy et al. (2020) also modelled heart failure and myocardial infarction as ADEs affected by NSAID use. In the OAC specific model in this chapter, there was no indication that the risk of myocardial infarction or heart failure was increased with NSAID use in anticoagulated patients (308, 309) and no other OAC population specific estimates were available. The risk ratios used in Foy et al. (2020) to describe the increased risk with NSAIDs were derived from electronic health records from the CPRD. The dataset was not linked with hospital data as was done in Chapter Four. The number of GI bleeding events in primary care alone is known to underestimate the risk of GI bleedings (431), which might have overestimated the costs and QALY loss associated with the HPE type.

Similar to this study, Foy et al. (2020) assumed that patients are switched from the NSAID to paracetamol once an ADE occurs. Different was that patients in the ASPIRE model sojourn in a post-GI event state until the end of the time horizon or until they die. Hence in contrast to this model, no further events are possible once any of the post-GI event states is reached. This simplification of the model by Foy et al. (2020) might result in an underestimation of the overall costs and QALYs generated in the model.

Another difference between the models was how long-term effects of ADEs were modelled. The ASPIRE model assumes no long-term effect on utilities for all ADEs except for stroke, similar to the model in this study. However, Foy et al. (2020) assumed long-term costs associated with symptomatic ulcer and serious GI events of £20.00 per three-months cycle over the lifetime of the patients. The costs are based on prescriptions costs of continuous treatment with PPIs (492). The most common full-dose PPI, however, costs £3.79 for three months (528). It was unclear how the £20.00 were derived. In the *de novo* state transition model in this thesis, no long-term costs were assumed for these health states. According to the NICE guidelines on peptic ulcer, treatment with PPIs is only recommended for eight weeks and therefore covered in the cost of the acute health state (522).

In the published literature, two additional models were identified that assessed the economic impact of NSAID related HPEs (151, 227). Elliott et al. (2014) presented cost per QALY associated with NSAID use without a PPI in patients with a history of peptic ulcer compared with concomitant NSAID and PPI use (227). A history of peptic ulcer is associated with an increased risk of serious GI events, similar to OAC use but not with a risk of stroke. A prescription of the gastroprotective PPIs is assumed to reduce this impact. The model structure is very similar to the structure chosen for this chapter. However, the comparator is slightly different because the HPE is not resolved by removing the NSAID but by adding a PPI. The authors did not find a difference in QALYs but identified higher costs in the HPE cohort of £399. This suggests that this NSAID related HPE has a smaller impact on healthcare costs compared with NSAID use in anticoagulated patients. However, the comparison has to be interpreted with caution. Some key assumptions differed in the analyses, such as the time horizon that was only five years in the study by Elliott et al. (2014), and the costs associated with the ADEs that were much lower than today.

Moriarty et al. (2019) built a state-transition model to estimate incremental costs and QALYs associated with NSAID use in patient at least 65 years of age (151). NSAID users generated incremental costs of €806 and 0.07 less QALYS compared to paracetamol users. These estimates were higher compared with the incremental costs and QALYs in this chapter (incremental costs: £244; incremental QALYS: 0.04). This comparison, however,

has to be treated with caution because the model structure and design differed, such as a time horizon of 35 years and no correction of the HPE after one year's exposure in Moriarty et al. (2019). The estimates in this chapter without the HPE correction yielded estimates more similar to the result in Moriarty et al. (2019) with regards to costs (incremental costs: £928) but not with regards to QALYs (incremental QALYs: 0.14). The higher QALY gain estimated for NSAID users in this chapter could be a result of the baseline characteristics of the OAC cohort. The OAC population did not only include the risk factor for serious GI events that was OAC use but also included patients over 65 years, which is associated with an increased risk of bleeding on its own and was therefore investigated by Moriarty et al. (2019).

5.4.3 Strengths and limitations

This was the first study to estimate projected harm and healthcare costs of NSAIDs in a population with OACs using specific input parameters for the OAC population. The various sensitivity analyses suggested that the conclusions from the model were robust to changes in model specifications and most parameter changes [Table 5.9]. The model was conceptualised in cooperation with clinicians and patient representatives to face-validate required assumptions. The final model was also reviewed by another health economist to improve internal validity.

Another key strength of this model was the availability of real-world evidence from the cohort study conducted in the linked CPRD/HES/ONS dataset. Evidence on HPE related harm is sparse (31), and one of the main limitations in previous models was the lack of data to populate the model with appropriate estimates of HPE related harm (227). A challenge in cohort models is the heterogeneity of studies that contribute input parameters to the model (440). Even though the state-transition model in this thesis still incorporated data from other studies, key input parameters that define the difference between the HPE and non HPE model were derived from the same cohort, the same methods and were from UK specific data. This increased the internal validity of the estimates and estimates could be adjusted to the specific requirements of the model. For example, baseline transition probabilities were generated for patients with and without NSAIDs. This was not reported

in other studies. The example of the ASPIRE study demonstrates the challenges of using estimates from the literature that are often not fit for purpose. In ASPIRE, instead of assessing the increased likelihood of ADEs in anticoagulated patients, the authors used the likelihood of events from a normal population with NSAIDs and paracetamol. This approach has various limitations compared to assessing the risk of NSAIDs in anticoagulated patients directly as was done in Chapter Four. First of all, the model by Foy et al. (2020) applied a risk ratio that does not describe what it was used for in the model. The risk ratio described the increased risk of serios GI events of OAC and NSAIDs compared to no treatment at all. Secondly, the authors also assumed that OACs have the same effect on transition probabilities of symptomatic ulcer as on serious GI events. In the reviews conducted as part of this chapter, no evidence was found that OACs effect the risk of symptomatic ulcer (460, 496, 497). Finally, the impact of OAC treatment on stroke was not accounted for. Overall, the model by Foy et al. (2020) did adjust probabilities for some events to the increased risk of ADEs in patients with OACs but did not adjust the increased likelihood associated with the HPE to the specific model population. Compared to the model by Foy et al. (2020), the strength of this chapter was that it reports HPE and non-HPE probabilities for the specific OAC population. This was only possible because the specific harm estimates could be generated from linked health records to inform the state-transition model. Estimates from the literature are often not fit for purpose, as with the model by Foy et al. (2020), and methods are often not clearly described to appraise the quality of data difficult.

A general limitation of cohort models, such as the *de novo* state-transition model in this chapter, is that they only represent a simplification of reality to be workable (440). More complex patient level simulation models (micro-simulation models) can accommodate patient heterogeneity and it is easier to model event history compared with cohort models. In order to accommodate patient heterogeneity, micro-simulation models require patient level data to describe outcomes based on specific baseline characteristics. In the cohort of anticoagulated patients, data on each different type of OAC would be required because the risk of serious GI events depends on the type of OAC used. If these data are available, micro-simulation models simulate outcomes for each patient with specific characteristics and uses these to estimate a distribution of outcomes in a potentially heterogenous sample (549). Another advantage of micro-simulation models is the ability to ‘memorize’ prior

health states. In state-transition models, input parameters can only depend on the current health state not on those before. This is particularly helpful in state-transition models that become too complex when multiple tunnel states and post-GI event states are included. However, most ADEs associated with NSAID use were not associated with long term impact on input parameters and data availability around subsequent events was very low. Therefore, the ability of a micro-simulation model to account for patient history, would not have been required.

An example for simplification of reality in the *de novo* state-transition model, was the assumption that NSAID use lasted for one year until the HPE was detected by the practice or an ADE occurred. NSAIDs in the UK are usually prescribed on a short-term basis for example for episodes of pain or gout, especially in patients over 65 years of age. These episodes are often recurring and if the patient had no ADEs in the first NSAID treatment episode, the GP might prescribe the NSAID again. This recurring pattern was not represented in the model. It was assumed that a patient in the no adverse event state could have a maximum of 12 months of NSAID prescriptions after which the HPE would have been detected. For the model, it made no difference if this one-year NSAID use was incorporated in a recurring pattern or in the first four cycles. Experts contacted for face validation of the model assumptions agreed that this was a reasonable approach, but it does not represent reality.

5.4.4 Implications for thesis

This chapter projected harm and costs associated with NSAID use in anticoagulated patients to estimate incremental costs and QALYs. To assess the economic impact a reduction of HPE rates has on healthcare costs and patient outcomes, this research is pivotal. The state-transition model can be combined with the decision-analytic model [3.2.1.4] to answer the question of cost-effectiveness of SMASH including consequences of HPEs that were not reported or incorporated in the cost-effectiveness analysis in Chapter Three.

5.4.5 Implications for policy makers

The Department of Health and Social Care already acknowledged the high risk and potential burden of NSAID use in anticoagulated patients by choosing it as one of the key HPE types as part of a programme of work to reduce medication error and promote safer use of medicines (55). Part of this programme of work was the NHS Medication Safety Dashboard. The dashboard links the HPE with hospitalisation rates for GI bleeding. The study in this chapter takes this approach to quantify the burden of HPEs further. By incorporating multiple ADEs associated with the HPE, primary and secondary care managed ADEs, and modelling the consequences over a lifetime provides an estimate of the burden of this HPE type in England. The patients exposed to the HPE cost the NHS about £3 million, and more than 500 QALYs are lost over their lifetime. These estimates are likely to inform policy decisions on which HPEs should be targeted by interventions. Compared to the incremental costs and QALYs associated with the HPE of NSAID use in patients with peptic ulcer by Elliott et al. (2014) and NSAID use in patients over 65 years of age, this HPE seems to have a higher impact on costs and outcomes (149, 151). It is therefore important to develop interventions that effectively reduce HPE types with a high economic burden as demonstrated in this dissertation for NSAID use in anticoagulated patients. Policy makers should focus on supporting the implementation of interventions that target HPE types that are associated with a high economic burden and occur frequently in England.

5.4.6 Considerations for future work

One of the limitations of this state-transition model was the data availability already pointed out in earlier models (149). While some of the gaps could be filled by generating the estimates from electronic health records [Chapter Four], other gaps remained. Utility estimates for the GI adverse event states were under researched and up to date utilities in UK populations assessed using EQ-5D-3L questionnaires rarely available. Of those estimates reported, methods were often not described transparently and sample sizes were small (537, 538). Future research could focus on assessing utilities for the GI adverse event to reduce uncertainty in the model. However, sensitivity analysis showed only minor effects of changing the utility of symptomatic ulcer. A value of information analysis could be performed to assess which parameters are worth studying in more detail.

Future research could also look at the economic impact of other HPE types. As part of the PROTECT programme grant, state-transition models are already being developed for other HPE types targeted by SMASH and PINCER. Comparing the economic impact of the different HPE types will be informative to identify those where a reduction in HPE rates would be most favourable. Net monetary benefits at the UK WTP threshold could be used not only to identify the most severe HPE types but also to get an understanding of an upper bound of potential intervention costs. An intervention resolving all HPEs cannot cost more than the incremental net monetary benefit estimated for the HPE types it targets to be cost-effective at the WTP threshold.

5.4.7 Conclusion

NSAID use in anticoagulated patients is not recommended and patients with concomitant NSAID and OAC treatment generate higher costs and less QALYs compared with anticoagulated patients without NSAID use. Under the premises that HPEs and their economic impact are mostly avoidable, resolving the HPEs could reduce the overall healthcare costs and improve quality of life in anticoagulated patients.

This chapter reported a method to utilise state-transition models to project long term consequences of HPEs associated with HPE related ADEs managed in primary and secondary care. How this can be achieved was illustrated for the HPE of NSAID use in anticoagulated patients. The method can also be used to estimate the economic impact of other types of HPEs.

Chapter 6 - Cost-utility analysis of SMASH

Chapter Six reports an economic evaluation of SMASH in reducing NSAID use in anticoagulated patients. It is described how harm from NSAIDs in anticoagulated patients is combined with the cost and effectiveness of the intervention to estimate incremental costs and QALYs. The discussion elaborates on the interpretation and the wider implications of this study.

6.1 Introduction

The economic analysis presented in Chapter Three generated the cost per HPE avoided by SMASH. This estimate is informative in identifying and valuing the resource use associated with SMASH but does not allow an interpretation of the overall effect of the intervention on patient related health outcomes and healthcare costs.

A key limitation of existing economic evaluations of patient safety programmes was either not including overall effect of the intervention on patient related health outcomes and healthcare costs, or having to rely on the scarce literature available on linking HPE with harm, patient outcomes and healthcare costs if they did (31, 146, 241) [Chapter Two]. Cost-effectiveness analyses of patient safety programmes rarely used patient outcomes as a measure of effectiveness and quality of life estimates were often derived from studies with weak study designs as pointed out in a review of economic evidence under the mandate of the European Commission (241). In the NICE reference case, utility weights for health related quality of life are recommended as the outcome measure of choice. The authors highlight the need to use quality of life (measured in QALYs) as the effectiveness outcome measure to enable comparability of interventions and to incorporate the potential benefit of interventions on health related quality of life. Additionally, cost estimates of ADEs associated with HPEs were diverse and variability of the reported results was high (146, 241). In the literature review of decision-analytic models assessing the economic impact of HPEs [2.8], the link between harm and the associated costs and outcomes was found to be one of the main sources of uncertainty. The majority of studies relied on estimates from

the literature on the cost of ADEs by ADE severity (156, 220, 222, 224, 225) and relied on expert elicitation for QALYs for each severity level (156, 219, 222, 226). The severity levels were estimated by experts or by using severity distributions from the literature. There are several limitations to this approach. First, the data availability was a major limitation (30, 441). Estimates by severity level were derived from studies that were run in different countries, all of which were outdated and based on small sample sizes (241). These were all factors considered to have a major impact on the cost of ADEs according to a systematic review of cost of ADEs (30). Consequently, it is important to use country specific and up to date estimates to avoid biased cost estimates for ADEs. Another limitation of existing cost estimates was the focus on costs of secondary care events alone. This underestimates the true economic burden by not including ADEs managed in primary care or the impact on death rates (441). A third limitation of this approach was that the cost of ADEs in the literature are too variable to obtain precise cost estimates (30). Often studies did not investigate individual HPE types, and the use of cost estimates assessed for a different type of HPE or a group of HPE types was not considered appropriate because of the huge variation between cost of ADEs associated with a specific HPE type.

The literature review in Chapter Two also identified key features of the study design that enabled studies to conduct cost-utility analyses. The first one was to focus on specific HPEs that can be linked to an actual measure of harm, for instance, a specific ADE associated with the HPE type. Studies modelling the economic impact of each HPE type individually, instead of different types of HPEs grouped by severity, were able to use more robust estimates on costs and QALYs. The second feature was to attach costs and utilities not to the HPE itself but to attach them to specific health states related to harm from the HPE. Four studies were found that used this approach, and this chapter utilises these features (149-152). Not any type of ADE is simulated, but the occurrences of specific ADEs associated with each HPE targeted by the intervention are simulated.

The cohort study conducted for this dissertation [Chapter Four], aimed to fill the gap in the literature on harm from HPEs. In Chapter Five, the economic impact that results from harm associated with this HPE type was assessed. In this chapter, the cost per HPE avoided analysis conducted in Chapter Three was extended to enable further interpretation of the

consequences of reductions of NSAID use in anticoagulated patients by SMASH. The overall aim was to determine the cost-effectiveness of SMASH in reducing NSAID use in anticoagulated patients compared with standard practice beyond the data from the primary effectiveness study. This aim was achieved by combining the HPE specific state-transition model [Chapter Five] with the cost per HPE avoided analysis [Chapter Three] to identify probabilistic cost per QALY generated by SMASH. SMASH targets a defined set of HPE types. For the analysis reported in this thesis, NSAID use in anticoagulated patients was used as an example case study to demonstrate how these different analyses can be combined.

6.2 Methods

The methods describe the design of the economic evaluation, the input parameters for the effectiveness of SMASH, the intervention cost of SMASH and the economic impact of the HPE type and how these are combined and analysed to assess the cost-effectiveness of SMASH in reducing NSAID use in anticoagulated patients.

6.2.1 Design of the economic evaluation

This chapter reports on a cost-utility analysis of SMASH compared with standard care in reducing NSAID use in anticoagulated patients in primary care. Cost-utility analysis were described in Chapter Two [2.8]. The economic evaluation followed guidelines from CHEERS (278). The CHEERS checklist can be found in Appendix C, where sections addressing key CHEERS criteria in this document are reported. The state-transition model conceptualised in Chapter Five and used in this analysis followed AdViSHE reporting and validating standards (442). The AdViSHE checklists and how its requirements were met is reported in Appendix L.

6.2.1.1 Target population

The target population in this cost-utility analysis were patients at risk of the HPE type in an average practice in Salford. Patients at risk had a mean age of 70 years, 55% were men and 45% were women based on population characteristics in Chapter Four [4.4.1]. The mean

age and gender distribution identified in the cohort of anticoagulated patients in CPRD practices is reported to be representative of the UK population (342, 343). The same cohort characteristics at baseline were used in Chapter Five in the state-transition model estimating the economic impact of the HPE type.

6.2.1.2 Strategies compared

The intervention investigated in this chapter was SMASH, a pharmacist-led electronic audit and feedback intervention. The alternative strategy that SMASH was compared to in this economic evaluation was standard care. In standard care, no interventions or measures to reduce HPEs, other than standard annual reviews by the GP, were in place. The intervention was described in Chapter Two [2.9], and the intervention and comparing strategies were discussed in detail Chapter Three [3.2.1.2].

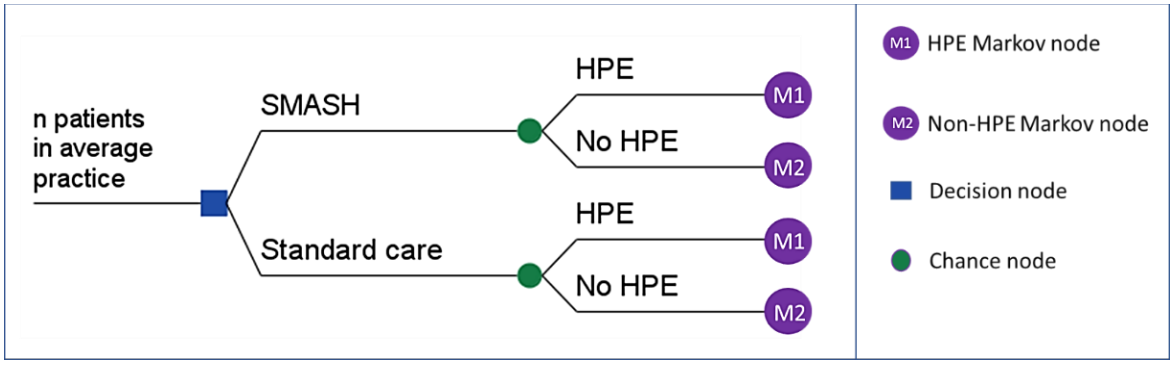
6.2.1.3 Primary outcomes

The primary health outcome was health related quality of life measured in QALYs [2.8] as recommended in the NICE reference case (3). Costs were estimated from the NHS/PSS perspective for the 2019 cost year. Costs and QALYs were used to measure the incremental cost-effectiveness of SMASH versus standard care. Costs and benefits were discounted using a 3.5% discount rate from the treasury (550) as recommended by NICE (3).

6.2.1.4 Structure of decision-analytic model

The underlying decision-analytic model of the cost-utility analysis is reported in Figure 6.1. It consisted of two stages: (i) the cost per HPE avoided analysis based on a decision tree, and (ii) the HPE specific state-transition model. In the decision tree, the short-term impact of the decision to implement SMASH, or not, on HPE rates (12-months time horizon) and in the state-transition model the longer-term sequelae of HPE exposure were modelled (lifetime horizon). The HPE specific state-transition model is populated with input parameters for the HPE cohort (M1) or input parameters for the non-HPE cohort (M2).

Figure 6.1: Decision-analytic model of the two-stage economic analysis estimating cost per QALY generated by SMASH compared with standard care



Details of the input parameters that differ between HPE and non-HPE cohort were reported in Chapter Five [5.2.4]. In the decision tree, the absolute number of patients exposed to an HPE and not exposed but at risk of the HPE in an average practice was assessed. Combined with the patient-level expected costs and QALYs in the state-transition model, the incremental costs and QALYs associated with SMASH in reducing NSAID use in anticoagulated patients was generated for an average practice.

6.2.2 Effectiveness of SMASH

The effectiveness of SMASH was described by the reduction in HPE rates as described in detail in Chapter Three [3.2.2]. In Chapter Three, effectiveness was reported for a composite of all HPEs. In this chapter, only the effectiveness of SMASH in reducing NSAID prescribing in anticoagulated patients is relevant. The probability of NSAID use in anticoagulated patients ($HPEprobability_{SMASH}$) was based on the regressed ITSA results for SMASH and the denominator ($No\ of\ patients\ at\ risk$) based on the observed number of patients with an OAC prescription at 12 months.

6.2.3 Cost of SMASH

The cost of SMASH was estimated in Chapter Three. The majority of cost components were fixed and independent of the number of HPEs reviewed: (i) server costs, (ii) pharmacist training, (iii) initial practice training, and (iv) IT support. The costs for managing HPEs, on the other hand, was variable dependent on the number of HPEs reviewed. Including the overall costs of managing all HPEs in this HPE specific cost-utility analysis would

overestimate the costs generated for only one HPE type. The variable costs were therefore allocated to the specific HPE (NSAID/OAC HPE). The resource use items required to estimate the cost component of managing HPEs were reported in detail in Chapter Three [3.2.3.3]. In summary, resource use for the management of HPEs entailed (i) the number of HPEs reviewed, (ii) the percentage of HPEs requiring patient contact, and (iii) the time pharmacists and GPs spent with managing HPEs that require patient contact and those that do not require patient contact. To allocate the cost, of managing HPEs to the specific NSAID/OAC HPE type, the number of HPEs reviewed was adjusted to represent only NSAID/OAC HPEs compared with the analysis in Chapter Three. The number of NSAID/OAC HPEs reviewed were assessed based on the number of new patients flagged by SMASH that had concomitant treatment with OACs and NSAIDs within 12 months. The time spent managing HPEs and the percentage of HPEs requiring patient contact were not adjusted. It was assumed that these are the same for all type of HPEs because no HPE specific data was available.

6.2.4 Economic impact of NSAID use in anticoagulated patients

The state-transition model conceptualised to assess the economic impact of NSAID use in anticoagulated patients was reported in Chapter Five. The impact the HPE had on the incidence of HPE specific ADEs was modelled. HPE specific ADEs were GI discomfort, symptomatic ulcer, serious GI events and stroke. The economic impact of NSAID use in anticoagulated patients was reported as incremental costs and QALYs.

6.2.5 Incremental economic analysis

The unit of analysis in this economic evaluation was an average practice in Salford. The observed practice-level HPE rates and costs at 12 months (base case) for the impact of SMASH versus standard care were incorporated into the HPE-specific model to generate an estimate of incremental QALYs and NHS costs associated with a reduction of NSAID use in anticoagulated patients, per practice, for SMASH versus standard care.

Base case analysis

The incremental cost per additional QALY generated by SMASH was denoted as

$$\text{Equation 6: } \frac{(Total\ cost_{SMASH} - Total\ cost_{standard\ care})}{(Total\ QALYs_{SMASH} - Total\ QALYs_{\ standard\ care})}$$

Monte Carlo simulation with 10000 iterations was conducted, utilising Microsoft Excel to obtain the 2.5% and 97.5% percentiles of the cost per QALY generated distribution. The incremental costs and QALYs from the probabilistic analysis were plotted on a cost-effectiveness plane. The probability of SMASH being cost-effective at different WTP thresholds was illustrated in a cost-effectiveness acceptability curve (CEAC). Results were also presented as incremental NMB for the WTP threshold of £20000 (551).

Equation 7: $Incremental\ NMB = Incremental\ effect * £20000 - Incremental\ cost$

A threshold analysis of the cost of SMASH, the absolute difference in HPE rates between SMASH and standard care, and the number of patients at risk of a HPE were conducted in *TreeAge Pro Healthcare 2021* to identify the scenarios under which SMASH is cost-effective in reducing NSAID use in anticoagulated patients at the WTP threshold of £20000.

6.2.6 Sensitivity analysis

Scenario analyses in probabilistic model

The PSA was run to identify how different scenarios of model specifications or key structural assumptions would change the incremental costs and QALYs. Parameter changes or structural assumptions that impacted the results in the cost per HPE avoided analysis [Chapter Three] or the economic impact analysis of the HPE [Chapter Five] were tested in this cost-effectiveness analysis, as well as different allocation methods of the intervention costs. The former sensitivity analyses were described in the respective chapters and the relevant assumptions are reported in Table 6.1.

Table 6.1: Description of the scenarios tested in the probabilistic model

Scenario	Assumption
Decision tree	
Delivery of SMASH for 6 months	The cost of delivering SMASH, the HPE rate with SMASH and the HPE rate with standard care were changed to the estimates assessed at 6 months after intervention start.
State-transition model	
Time horizon: 10 years	Instead of 120 cycles, the state-transition model was run for 40 cycles[a].
Time horizon: 5 years	Instead of 120 cycles, the model was run for 20 cycles[a].
No HPE correction	The NSAID was only removed after ADEs occurred. The HPE was not detected and resolved after a maximum of 12 months as assumed in the base case.
Discount rate: 0%	The base case analysis applies a 3.5% discount rate to costs and outcomes. To test what impact his makes, a scenario was run with no discounting.
State-transition model – HR stroke/serious GI source[a]	
HR Chapter Four - 60-day grace	Extending the grace period after calculated prescription stop dates to 60 days in the cohort study in Chapter Four; HR serious GI event 2.53 (1.66; 3.86); HR stroke 1.79 (1.13; 2.84).
HR Chapter Four - IPTW	Using IPTW instead of propensity score matching in the cohort study in Chapter Four; HR serious GI event 2.47 (1.40; 4.34); HR stroke 1.71 (0.95; 3.08).
HR Chapter Four - NSAID washout 6 months	Extending the washout period for NSAID use before index date to 6 months in the cohort study in Chapter Four; HR serious GI event 3.81 (2.04; 7.09); HR stroke 3.11 (1.61; 6.01).
HR - RCT dataset (Dalgaard (2020))	Using increased risk ratios from a subgroup analysis of the ARISTOTLE arm instead of results from the cohort study in Chapter Four; HR serious GI event 3.37 (1.81; 6.25); HR stroke 2.33 (0.97; 5.59).
Allocation of costs	
Total cost by number of HPE types	Fixed and variable costs depend on the number of HPE types (n=10) targeted by the intervention. Each HPE type contributes to the total cost in the same way.
Variable cost by number of HPE types	Only variable costs depend on the number of HPE types (n=10) targeted by the intervention. Each HPE type contributes to the variable costs in the same way.
Total cost by proportion of HPE type	Fixed and variable costs depend on the number of HPEs (n=1283) identified at baseline. How much each HPE type contributes to the total cost is dependent on its proportion of all HPEs. Of all HPEs, 3.43% were NSAID/OAC HPEs.
Variable cost by proportion of HPE type	Only variable costs depend on the number of HPEs (n=1283) identified at baseline. How much each HPE type contributes to the total cost is dependent on its proportion of all HPEs. Of all HPEs 3.43% were NSAID/OAC HPEs.

[a]Details were reported in Chapter Five [5.3.2]; ADEs: adverse drug events

Three different methods were tested to allocate the costs compared with the analysis in Chapter Three. In the base case analysis, costs and HPE rates were adjusted to describe only the cost and HPE rate of the specific NSAID/OAC HPE. In one scenario analysis, the total intervention cost of £2135 as estimated in Chapter Three [3.3.3], was divided by the number of HPE types targeted by SMASH, assuming each of these ten HPE types contributed the same to the total intervention cost. In a second sensitivity analysis, the intervention costs were allocated to the HPE type by using a weight based on the prevalence of the HPE. This prevalence weight was generated from the proportion of NSAID/OAC HPEs among all HPEs at baseline. The absolute number of HPEs in the SMASH study in all practices was 1283 at baseline and 44 (3.43%) of these were anticoagulated patients with concomitant NSAID treatment (48). The third sensitivity analysis of an allocation method of costs adjusted only the variable cost of £1840. The first and second scenario adjusted the total cost, fixed and variable, of the intervention. The variable costs were the cost for HPE management that depended on the number of HPEs reviewed. These were allocated to the specific HPE type by dividing them by the total number of HPE types that was ten in one scenario and by the proportion of the HPE among all HPEs in another scenario. The intervention costs derived from each allocation method are reported in the results.

One-way sensitivity analysis in the deterministic model

Parameter uncertainty was already accounted for in the PSA. Additionally, one-way sensitivity analysis was conducted in the deterministic model to identify how individual parameters impact the incremental NMB. Parameters were tested for the input for the decision tree (effectiveness estimate, number of patients at risk and intervention costs) and the state-transition model (health state costs, health state utilities and decrements, mortality estimates, increased risk ratios for ADEs associated with the HPE and the discount rate). The parameter values applied in the sensitivity analysis are reported in Table 6.2. The confidence intervals or standard deviations reported in the respective sources were used to identify ranges for the one-way sensitivity analysis. Where no measure of uncertainty was reported in the source, a range of ±25% was used.

Table 6.2: Parameter ranges applied in one-way sensitivity analysis

Variable in decision-analytic model	Expected value	Low range	High range	Source of uncertainty
Input decision tree				
Absolute difference, HPE rate with SMASH vs. historical comparator	0.08%	-0.12%	0.26%	95% CI
HPE rate SMASH	1.12%	0.82%	1.50%	95% CI
Number of patients at risk per practice	91.16	73.00	109.00	SD
Cost of SMASH, £	380	73	478	Allocation method of costs
Input state-transition model				
Health state cost, £				
No adverse event, with HPE	7.68	5.76	9.6	±25%
No adverse event, no HPE	19.06	14.30	23.83	±25%
Serious GI event	3137.45	2652.00	3623.00	95% CI
Symptomatic ulcer	571.36	428.52	714.20	±25%
GI discomfort	34.18	25.64	42.73	±25%
Stroke	10645.86	8937.12	12354.60	95% CI
Post-stroke	235.20	212.80	257.60	95% CI
Health state utilities and decrements				
Utility no adverse event, age 70-74	0.78	0.77	0.82	95% CI
Utility no adverse event, age 75-79	0.75	0.74	0.76	95% CI
Utility decrement, serious GI events	0.05	0.03	0.06	±25%
Utility decrement, GI discomfort	0.02	0.02	0.03	±25%
Utility decrement, symptomatic ulcer	0.03	0.05	0.07	95% CI
Utility decrement, stroke	0.06	0.05	0.07	±25%
Utility decrement, post-stroke	0.05	0.03	0.07	95% CI
Mortality				
Mortality, serious GI event, age 70-74	6.37%	4.77%	0.08%	±25%
Mortality, serious GI event, age 75-79	11.23%	8.42%	0.14%	±25%
Mortality, symptomatic ulcer, age 70-74	2.47%	1.85%	0.03%	±25%
Mortality, symptomatic ulcer, age 75-79	4.33%	3.25%	0.05%	±25%
Mortality, stroke, age 70-74	5.64%	4.23%	0.07%	±25%
Mortality, stroke, age 75-79	9.96%	7.47%	0.12%	±25%
Mortality, post-stroke, age 70-74	1.70%	1.49%	0.02%	95% CI
Mortality post-stroke, age 75-79	3.01%	2.80%	0.03%	95% CI
Mortality no adverse event, age 70-74	1.15%	0.99%	1.32%	95% CI
Increased risk ratios for adverse drug events				
HR serious GI event	2.60	1.39	4.87	95% CI
OR GI discomfort	2.12	1.73	2.58	95% CI
OR symptomatic ulcer	1.70	1.49	1.94	95% CI
HR stroke	2.98	1.62	5.50	95% CI
Discounting of costs and outcomes				
Discount rate	1.04	1.00	1.07	No or twice the discount rate

CI: confidence interval; GI: gastro-intestinal; HR: hazard ratio; OR: odds ratio; SD: standard deviation

6.2.7 Validation

Model validation is required to ensure that the model represents the treatment pathway and the system accurately and that results are plausible. In the NICE guidance on the methods of technology appraisal, results and methods of model validation are required (3). Nevertheless, NICE does not provide further detail on what the validation should incorporate. The AdViSHE tool (442) was developed to set a framework of model validation that is necessary to ensure model validity. Various types of validity assessment are possible: (i) face validity; (ii) cross validity; (iii) internal validity; and (iv) external validity.

(i) Face validity: To determine face validity, the model structure was presented to clinical experts (GPs, pharmacists), health economists, and lay members of the PROTECT study team (other researchers, patients) in oral presentations. The participants were asked whether the structure mirrors the treatment pathway of the population under investigation and whether assumptions on long-term effects, input parameters and distributions made within the model and model outcomes are reasonable. Additionally, the model was validated throughout the development process from health economist and pharmacists in the research team.

(ii) Cross validity: The construction process of the model included a literature review of models published on the cost-effectiveness of interventions aiming to reduce HPEs [2.5] and on HPE specific models [Appendix K]. The identified models were used to compare assumptions they make and transition states and data input they use.

(iii) Internal validity: To verify internal validity, the model went through a rigorous debugging process. The correctness of chosen equations, statistical distributions, coding language and consistency with the model specification was checked by different parties from the research team at various time points. The state-transition model was examined by a health economist, who has experiences with NSAID related HPE models from various previous projects. Results of this validation were reported in Appendix N. The results were checked for unrealistic results, such as negative costs, more events than possible or more QALYs than life years that undermine the credibility of the model. This was done by using algorithms to verify that transition probabilities per state equal one or that the number of

patients stayed constant in each cycle. Sensitivity analyses were carried out to identify whether the model behaves as expected or whether any anomalies were found (552). One-way sensitivity analysis was used changing input values beyond a plausible range in order to identify any bugs in the model. Transition probabilities and utilities were tested for every value between zero and one and values for costs from zero to extreme high values.

(iv) External validity: An independent external validation was not possible as part of this programme of work.

6.3 Results

6.3.1 Effectiveness of SMASH

SMASH reduced the overall probability of NSAID exposure by 0.08% (95% CI -0.12% to 0.26%) to 1.12% (95% CI 0.82% to 1.50%) after 12 months [Table 6.3]. The relative HPE reduction was 5.37% (absolute reduction of 0.08%), resulting in 1.1 HPE avoided by SMASH per practice with 91 patients with an OAC prescription per practice.

Table 6.3: HPE rate at 6 and 12 months for SMASH and standard care calculated from Peek et al. (2020) with the resulting number of HPEs avoided per practice (48)

Time point	Mean with 95% CI		Distribution
Observed HPE rate with SMASH (in % of all at risk)			
At 12 months (base case)	1.12 (0.82; 1.50)		Beta
At 6 months	1.11 (0.80; 1.51)		Beta
HPE reduction with SMASH			
	Absolute difference[a]	Number of HPEs avoided[b]	
At 12 months (base case)	0.08 (-0.12; 0.26)	1.1 (1.6; 0.4)	Normal[c]
At 6 months	0.13 (-0.01; 0.24)	0.7 (1.1; 0.2)	Normal[c]

[a]Interrupted time series analysis output from Peek et al. (2020); [b]number of HPE avoided generated by multiplying the number of patients at risk with the absolute difference in HPE rates between SMASH and standard care (not used in the decision-analytic model); [c]distribution applied to absolute difference; HPE: hazardous prescribing event; SMASH: Safety Medication Dashboard

6.3.2 Cost of SMASH

Resource use and unit costs were reported in detail in Chapter Three [1023.3.2]. In the base case analysis in this chapter, the costs were allocated to be representative for SMASH in reducing NSAID use in anticoagulated patients. The only parameter that changed compared with Chapter Three was the number of HPEs reviewed. Pharmacists are assumed to review three (SD ±3) and five (±4) HPEs at six and 12 months, respectively. The total cost of SMASH adjusted to the specific HPE was £380 (2.5% to 97.5% credible interval £347 to £417) per practice after 12 months [Table 6.4]. Cost of SMASH applied in scenarios analyses are also reported in Table 6.4. Allocating the variable costs by number of NSAID/OAC HPEs reviewed over 12 months (base case), by HPE type or by the proportion of the HPE among all HPEs at baseline resulted in higher costs, compared to allocation methods that allocated the total costs.

Table 6.4: Total cost of SMASH for reviewing anticoagulated patients with NSAID use

Cost item	Total cost	Low range[a]	High range[a]	SD	Distribution[b]
Base case					
Total cost at 12 months[c]	380	347	417	19	Gamma
Scenario: Delivery of SMASH for 6 months					
Total cost at 6 months[c]	321	301	343	11	Gamma
Scenario: Different allocation methods					
Total cost by number of HPE types	215	49	579	143	Gamma
Variable cost by number of HPE types	478	312	842	143	Gamma
Total cost by proportion of HPE type	73	17	197	49	Gamma
Variable cost by proportion of HPE type	356	299	479	49	Gamma

[a]2.5% (low range) to 97.5% (high range) credible interval; [b]distribution applied in probabilistic analysis; [c]allocated by adjusting the number of HPEs reviewed to only NSAID/OAC HPEs; HPE: hazardous prescribing event; SD: standard deviation, used to describe gamma uncertainty distributions in *TreeAge Pro Healthcare 2021*; SMASH: Safety Medication Dashboard

6.3.3 Economic impact of NSAID use in anticoagulated patients

The state-transition model with applied transition probabilities, health state utilities and costs with distributions is reported in Chapter Five [5.2.2, 5.2.4].

6.3.4 Incremental economic analysis

The results generated by the decision-analytic model comparing SMASH and standard care in reducing NSAID use in anticoagulated patients are reported in Table 6.5. SMASH generated higher costs (£363, 2.5% to 97.5% credible interval £248 to £424) and more QALYs (0.003 QALYs, 2.5% to 97.5% credible interval -0.008 to 0.019) compared with standard care resulting in an ICER of £141411.

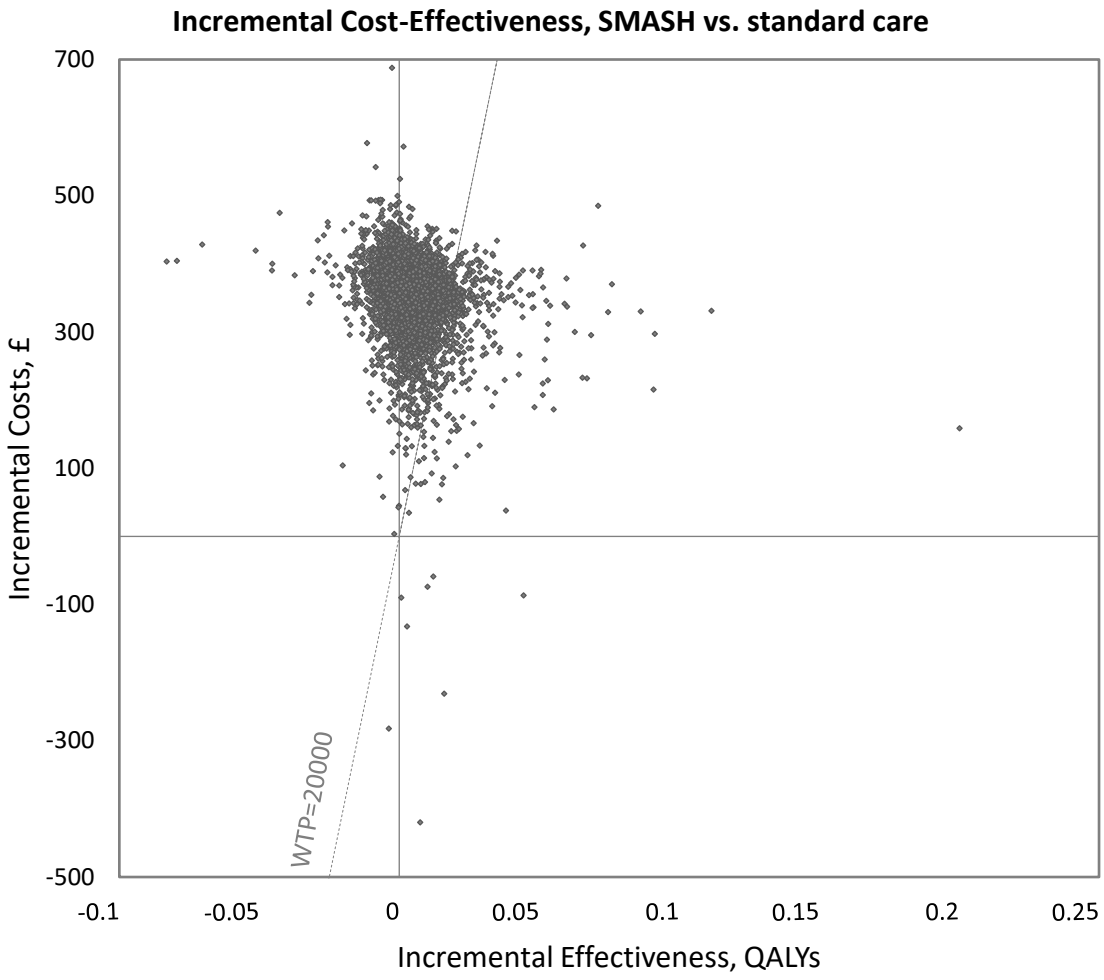
Table 6.5: Results of cost-effectiveness analysis of SMASH compared with standard care per practice in deterministic and probabilistic analysis (n=91 patients at risk)[a]

	Deterministic	Probabilistic [b]
Costs generated, £		
SMASH	169771	169333 (34632; 449511)
Standard care	169408	168970 (34251; 449148)
Incremental costs	363	363 (248; 424)
Effectiveness, QALYs		
SMASH	532.185	532.414 (493.573; 554.539)
Standard care	532.182	532.412 (493.572; 554.531)
Incremental effectiveness	0.003	0.003 (-0.008; 0.019)
Cost-effectiveness		
ICER, £/QALY	142012	141411[c]
Incremental NMB, £	-312	-311 (-542; 73)

[a]Lifetime horizon of state-transition model, costs and outcomes discounted at 3.5%; cost and HPE rates for SMASH after 12 months; [b]mean estimates with bootstrapped 2.5% to 97.5% percentile credible interval; [c]no confidence interval reported because they are not interpretable if some of the ICERs are <0; ICER: incremental cost-effectiveness ratio; NMB: net monetary benefit; QALYs: quality-adjusted life-years; SMASH: Safety Medication Dashboard

The incremental costs and QALYs of the 10000 simulations are presented in Figure 6.2 in a scatter plot. Of all the estimates 0.07% were in the south-east quadrant (SMASH more effective, less costly: dominant) compared with 65.25% in the north-east quadrant (SMASH more effective, more costly), 0.01% in the south-west quadrant (SMASH less effective, less costly) and 34.67% in the north-west quadrant (SMASH less effective, more costly: dominated).

Figure 6.2: Scatter plot for the probabilistic analysis with 10000 iterations

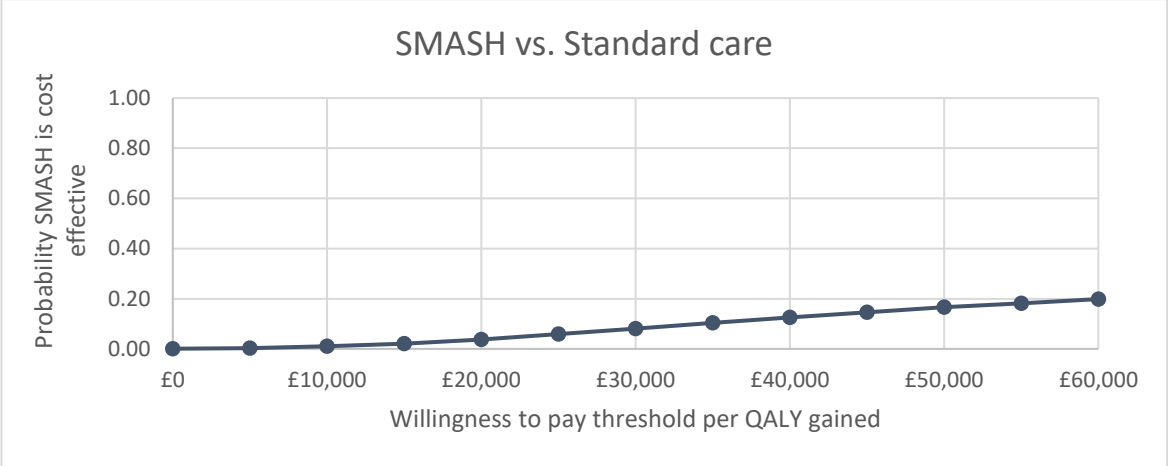


Quadrant	North-east	South-east	South-west	North-west
Percentage in quadrant	65.25%	0.07%	0.01%	34.67%

Estimates that fall in the north-east (higher costs, higher effectiveness) quadrant can be cost-effective if the ICER is below a specified WTP threshold. Those estimates that fall in the south-west quadrant are considered cost-effective if the ICER is greater the threshold.

The cost-effectiveness acceptability curve of SMASH compared with standard care presents the probability of SMASH being cost-effective at different WTP thresholds [Figure 6.3]. At the UK WTP threshold of £20000 per QALY gained, the probability of SMASH being cost-effective in reducing NSAID use in anticoagulated patients was 4%. The incremental NMB was -£311 (2.5% to 97.5% credible interval -£542 to £73) at a WTP of £20000 [Table 6.5].

Figure 6.3: Cost-effectiveness acceptability curve (CEAC) – base case analysis



The threshold analysis identified that the cost of SMASH would have to be £68 or lower to be cost-effective. The absolute difference in HPE rates would need to be at least 0.45% for SMASH to be cost-effective in reducing NSAID use in anticoagulated patients. With a number of patients at risk of at least 513 patients with OAC treatment, SMASH would be cost-effective at the WTP threshold.

6.3.5 Sensitivity analysis

Sensitivity analysis in the probabilistic model

Results of the sensitivity analysis in the probabilistic model on the impact of different assumptions are reported in Table 6.6. The analyses changed the results as expected. None of the changes in assumptions changed the fact that SMASH generated higher costs and more QALYs compared with standard care. The incremental NMB showed that SMASH was not cost-effective in reducing NSAID use in anticoagulated patients in all scenarios. The scenario under which SMASH generated the largest incremental NMB in reducing NSAID use in anticoagulated patients was when the total cost of SMASH was allocated by the proportion of the NSAID/OAC HPE among all HPEs (incremental NMB, -£5). The second highest incremental NMB was generated, when the assumption that the HPE is corrected without the occurrence of an ADE after a maximum of one year was released. The scenario resulted in an incremental NMB of -£105. In this scenario, patients exposed to the HPE are at an increased risk of ADEs until they die or experience an ADE and not only for a maximum of 12 months.

Table 6.6: Results of probabilistic sensitivity analysis of SMASH compared with standard care in reducing NSAID use in anticoagulated patients

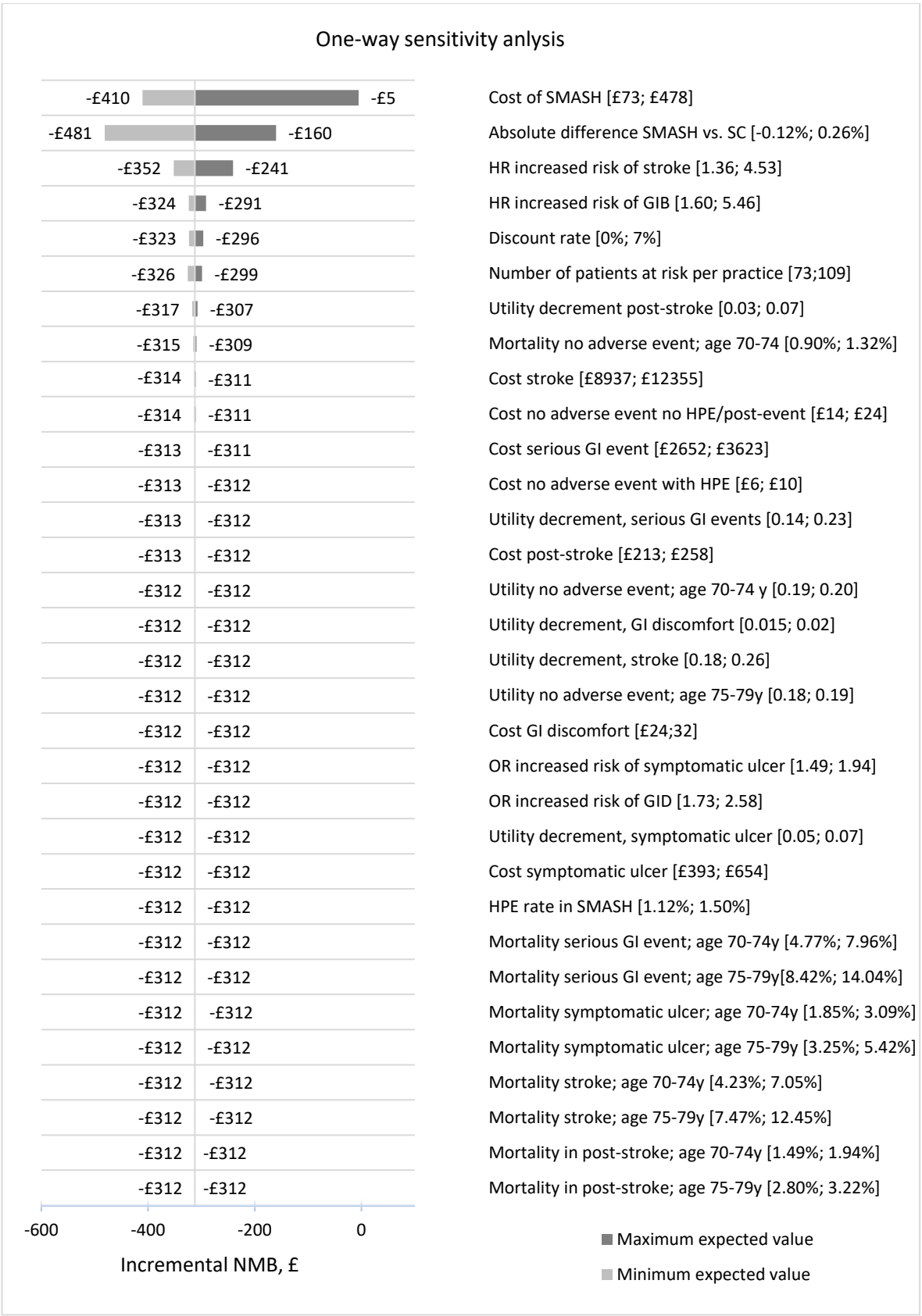
Analysis	**Costs generated per patient, £**		**QALYs generated per patient, QALY**		**Incremental**	
	SMASH	Standard care	SMASH	Standard care	Costs, £	QALYs
Base case analysis						
Probabilistic	169333 (34633; 449511)	168970 (34251; 449148)	532.41 (493.57; 554.54)	532.41 (493.57; 554.53)	363 (248; 424)	0.003 (-0.008; 0.019)
				ICER, NMB:	*£141411 per QALY, -£311*	
Deterministic	169772	169408	532.18	532.18	363	0.003
				ICER, NMB:	*£142012 per QALY, -£312*	
Sensitivity analysis in probabilistic model						
No HPE correction	170189 (34753; 451812)	169888 (34407; 451923)	532.31 (493.44; 554.35)	532.30 (493.43; 554.35)	301 (-132; 489)	0.010 (-0.021; 0.061)
				ICER, NMB:	*£30728 per QALY, -£105*	
Delivery of SMASH for 6 months	169271 (34566; 449423)	168977 (34274; 449166)	532.42 (493.57; 554.54)	532.41 (493.57; 554.54)	294 (178; 346)	0.004 (-0.007; 0.023)
				ICER, NMB:	*£66016 per QALY, -£205*	
Time horizon: 10 years	125783 (24522; 341824)	125420 (24154; 341829)	440.39 (416.12; 455.22)	440.38 (416.12; 455.23)	362 (247; 424)	0.002 (-0.006; 0.015)
				ICER, NMB:	*£188501 per QALY, -£324*	
Time horizon: 5 years	74539 (12717; 209419)	74175 (12343; 208929)	280.42 (270.58; 288.05)	280.42 (270.58; 288.05)	364 (254; 424)	0.001 (-0.004; 0.009)
				ICER, NMB:	*£330672 per QALY, -£311*	
Start age 80	103683 (17591; 288044)	103321 (17236; 287637)	265.45 (248.30; 279.83)	265.45 (248.29; 279.82)	362 (242; 426)	0.003 (-0.006; 0.018)
				ICER, NMB:	*£131648 per QALY, -£307*	
Discount rate: 0%	217584 (44998; 574574)	217222 (44644; 574274)	656.33 (603.40; 686.62)	656.33 (603.38; 686.62)	362 (245; 425)	0.003 (-0.010; 0.024)
				ICER, NMB:	*£110222 per QALY, -£296*	

Analysis	Costs generated per patient, £		QALYs generated per patient, QALY		Incremental	
	SMASH	Standard care	SMASH	Standard care	Costs, £	QALYs
Increased risk ratios for stroke and serious GI events						
HR Chapter Four - 60-day grace	169200 (34597; 449260)	168828 (34210; 448983)	532.43 (493.65; 554.55)	532.43 (493.66; 554.55)	373 (302; 424)	0.001 (-0.008; 0.014)
				ICER, NMB:	*£269264 per QALY, -£345*	
HR Chapter Four - IPTW	169206 (34583; 448694)	168834 (34194; 448315)	532.43 (493.59; 554.55)	532.43 (493.59; 554.54)	372 (301; 425)	0.001 (-0.008; 0.014)
				ICER, NMB:	*£269885 per QALY, -£345*	
HR Chapter Four - NSAID washout 6 m	169544 (34711; 449615)	169196 (34317; 449399)	532.39 (493.39; 554.48)	532.38 (493.39; 554.47)	348 (181; 435)	0.005 (-0.008; 0.029)
				ICER, NMB:	*£76723 per QALY, -£257*	
HR - RCT dataset (Dalgaard (2020))	169140 (34567; 449531)	168763 (34187; 449156)	532.44 (493.86; 554.54)	532.44 (493.86; 554.54)	377 (320; 428)	0.001 (-0.009; 0.013)
				ICER, NMB:	*£475760 per QALY, -£361*	
Allocation method of costs for SMASH						
Total cost by number of HPE types	169168 (34575; 449319)	168970 (34251; 449148)	532.41 (493.57; 554.54)	532.41 (493.57; 554.53)	198 (-8; 570)	0.003 (-0.008; 0.019)
				ICER, NMB:	*£77028 per QALY, -£146*	
Variable cost by number of HPE types	169433 (34651; 449629)	168970 (34251; 449148)	532.41 (493.57; 554.54)	532.41 (493.57; 554.53)	463 (208; 788)	0.003 (-0.008; 0.019)
				ICER, NMB:	*£180395 per QALY, -£411*	
Total cost by proportion of HPE type	169026 (34269; 449172)	168970 (34251; 449148)	532.41 (493.57; 554.54)	532.41 (493.57; 554.53)	56 (-69; 191)	0.003 (-0.008; 0.019)
				ICER, NMB:	*£21829 per QALY, -£5*	
Variable cost by proportion of HPE type	169309 (34607; 449458)	168970 (34251; 449148)	532.41 (493.57; 554.54)	532.41 (493.57; 554.53)	339 (208; 455)	0.003 (-0.008; 0.019)
				ICER, NMB:	*£132062 per QALY, -£287*	

[a]Source for HRs describing the increased risk of stroke and serious GI events associated with NSAID use in patients with oral anticoagulants; GI: gastro-intestinal; HPE: hazardous prescribing event; ICER: incremental cost-effectiveness ratio; NMB: incremental net monetary benefit; SMASH: Safety Medication Dashboard; QALY: quality-adjusted life-year

One-way sensitivity analyses in deterministic model

Figure 6.4: Tornado diagram on impact of individual input parameters on the incremental net monetary benefit (£312)



The results of the one-way sensitivity analysis are reported in Figure 6.4. The results were most sensitive to changes in the cost of SMASH, the absolute difference of the HPE rates in SMASH and standard care, and the risk increase of stroke with NSAIDs. The uncertainty around the absolute difference of SMASH vs. standard care ranged from -£481 to -£160. The minimum and maximum expected values for the HR of stroke changed the incremental NMB from -£352 to -£241, respectively. All other parameters did not vary the incremental NMB by more than £33.

6.4 Discussion

6.4.1 Principal findings

The reduction of NSAID use in anticoagulated patients by SMASH generated higher costs and more QALYs compared with standard care. The cost-effectiveness plane showed that incremental cost and incremental QALY estimates were distributed across all four quadrants. At a WTP threshold of £20000, SMASH had a 4% probability of being cost-effective in reducing NSAID use in anticoagulated patients with an incremental NMB of -£311 (2.5% to 97.5% credible interval -£542 to £72). The absolute effect of SMASH on reducing NSAID use in anticoagulated patients was small and not significant with an absolute difference in HPE rates of 0.08% between SMASH and standard care. Threshold analysis showed that an absolute difference of at least 0.45% would be required for SMASH to be cost-effective in reducing NSAID use in anticoagulated patients at the WTP threshold. The average absolute reduction of HPE rates with SMASH among all ten HPEs was 0.96% [3.3.1]. SMASH was less effective in reducing NSAID use in anticoagulated patients compared with other HPE types targeted by SMASH. One possible explanation for this result is that the ITSA was underpowered to detect a change in the relatively low prevalence of the HPE type. The rollout of SMASH to practices in Greater Manchester will show if an increased sample size can detect a change if a change is associated with SMASH.

The deterministic sensitivity analysis showed that uncertainty around most of the chosen estimates did not change the overall results. Only three input parameter changes varied the incremental NMB by more than £30 [Figure 6.4]. The incremental NMB of SMASH was most sensitive to the allocation method of the costs of SMASH, the absolute difference

between the HPE rates in the two strategies and the increased risk of stroke and serious GI events associated with NSAID use. The uncertainty range of the increased risk ratios derived from the cohort study in Chapter Four was substantial. In the state-transition model, they describe the increased risks of the ADEs associated with the highest economic and health related burden. This explains the impact changes in these parameters have on the incremental NMB. The results were also sensitive to the number of patients at risk of the HPE per practice. This was expected because a larger sample size increases the number of HPEs avoided by SMASH. In a practice with a minimum of 513 patients at risk of the HPE type, SMASH was found to be cost-effective at the WTP threshold of £20000.

Changing key model assumptions in the probabilistic sensitivity analysis affected the results as anticipated. A key assumption that impacted the incremental costs and QALYs was the correction of the HPE after a maximum of one year in the no adverse event state exposed to the HPE. This conservative assumption was made to acknowledge the potential that HPEs were detected in annual patient reviews in the practices. The scenario analysis showed that the incremental NMB was -£5, hence almost cost-effective for this scenario. However, the less conservative assumption is unlikely to be reasonable because NSAIDs are rarely prescribed indefinitely.

6.4.2 Comparisons with prior work

The results do not represent the overall cost-effectiveness of SMASH. The results of this cost-effectiveness analysis were, therefore, not compared to results of other economic evaluations aiming to reduce HPEs. The aim of this chapter was to demonstrate how the decision tree and the state-transition model can be combined. This method of linking the process indicator of HPE rates with patient outcomes and costs is, therefore, compared to methods in the literature to quantify the impact of HPEs. This process is divided into (i) a comparison with methods to quantify harm from HPEs and (ii) a comparison with CUAs modelling harm of specific HPE types.

Comparison with methods to quantify harm from HPEs

In previous studies, HPEs were grouped not by type of HPE but by severity or their potential to cause an ADE. In the literature, harm was either estimated by (i) expert elicitation (225), or (ii) based on severity distributions from the literature unrelated to the actual HPEs detected (156, 222-224). The limitations of these approaches were discussed in Chapter Two [2.4; 2.8] and lead to the decision to estimate harm for each individual HPE type targeted by SMASH separately as was done for NSAID use in anticoagulated patients in this chapter.

The first method of expert elicitation can estimate the potential to cause harm for individual cases, but some of these HPEs will be intercepted and will not reach the patient. Assessment by experts was also found not to produce reproducible estimates of harm [2.4]. Besides these general limitations of the methods, they were also not actionable in the context of the SMASH effectiveness study. In SMASH, the severity of occurring HPEs was not assessed as part of the intervention (48). Asking experts to estimate the potential harm of each HPE identified by SMASH would be an expensive and time-consuming process. Additional staff resources would be required that would further increase the cost of the intervention. Consequently, assessing the potential severity of HPEs by experts to estimate harm from HPEs does not only have methodological limitations but was also not practical for SMASH.

The other method economic evaluations in the published literature applied to estimate harm from HPEs, was the use of severity distributions from other sources. The use of severity distributions from the literature relies on the assumptions that the type of HPEs and their severity are the same in the population in the economic evaluation and the population in the study that estimated the severity distribution of HPEs. However, the populations are often not comparable because the underlying types of HPEs differ. HPE types and their distributions were found to differ between countries (553), settings and available care or HPE management (554) [Chapter Two]. For SMASH, the type of HPEs reduced was known and there were no UK specific estimates linking exactly this set of HPE types with harm.

Because neither expert elicitation or severity distributions from the literature were considered appropriate and robust methods to estimate harm from HPEs avoided by SMASH, a third option identified in the literature review was used. The alternative approach in the literature was to estimate harm for each HPE type separately. This study demonstrated this for NSAID use in anticoagulated patients. Harm from the HPE types was measured as the increased risk of specific ADEs associated with the HPE type [Chapter Four]. Using linked primary and secondary care data, the cohort study in Chapter Three was able to link HPEs in primary care with specific ADEs in secondary care (e.g., serious GI events), instead of generic harm outcomes (e.g., hospitalisations). This allowed a more robust link between harm and HPE, than expert elicitation or severity distributions from the literature.

Comparison with CUAs modelling harm of specific HPE types

Three cost-effectiveness studies were found in the literature that modelled costs and outcomes of specific types of HPEs to estimate cost-effectiveness of interventions (150, 152, 227). Elliott et al. (2014) and Foy et al. (2020) assessed the cost-effectiveness of an intervention aiming to reduce multiple HPE types (152, 227). To estimate harm from each HPE type, separate state-transition models were built. Both studies relied on estimates from the literature to inform the increased risk of harm associated with each HPE type and highlighted the scarce data availability as a limitation. Foy et al. (2020) built a model on the same HPE as presented in this chapter (152). Due to lack of available data the link of the HPE with patient harm was solely based on the risk increases of ADEs associated with NSAIDs in general, not with the specific population of anticoagulated patients. As discussed in Chapter Five, the availability of the results from the cohort study in this economic evaluation enabled a more precise estimation of harm that acknowledges the impact OAC use has on the ADEs related to NSAIDs. Elliott et al. (2014) applied results from a meta-analysis of RCTs of the relative impact of adding a gastroprotective agent on endoscopic ulcer occurrence (227). The relative risk was applied to serious GI events in the state-transition model. It was assumed that the relative effect of gastroprotective agents on serious GI events was the same as for endoscopic ulcers because no other data were available.

The third study by Forster et al. (2018) investigated an intervention aiming to reduce administration errors of asthma inhalers. The CRITIKAL study measured harm associated with the exact administration error targeted by the intervention investigated by Forster et al. (2018) (555). Similar to the cohort study in Chapter Four, the CRITIKAL study used observational data. The relationship between administration errors and harm outcomes was estimated using logistic regression. The types of outcomes considered in the economic evaluation were restricted to those measured in the CRITIKAL study. Other potential ADEs or consequences of the outcome, such as mortality, were not included in the state-transition model. In the *de novo* state-transition model developed as part of this thesis, primary care managed ADEs (symptomatic ulcer and GI discomfort) were also included, even though these were not measured in the cohort study. The projected costs and quality of life estimates in Forster et al. (2018) were also limited to one year, the time period of the CRITIKAL study. The advantage of the modelling approach chosen in this chapter was that it allowed extrapolation of harm beyond the primary data. This way, a long-term estimate of the consequences of HPEs was possible beyond the 12 months observation period in the SMASH effectiveness study (48).

These three studies were selected for discussion because they estimated the effectiveness of strategies to reduce HPEs by modelling harm from HPEs using state-transition models. Essential to this economic evaluation was the availability of linked primary and secondary care data. This allowed to estimate the increased risk of ADEs for the specific population at risk of the HPEs and filled the gap in the literature where estimates on the increased risk of harm from HPEs in general is scarce. The economic evaluation in this thesis did not rely on the best available data from the literature on the increased risk of the major ADEs associated with the HPE. This was one of the major limitations of the economic evaluations by Elliott et al. (2014), Forster et al. (2018) and Foy et al. (2020). Especially in areas with scarce literature, such as harm from HPEs, evidence from the literature might not ideally describe the population or outcomes modelled. Assessing the required input parameters as part of the economic evaluation allowed complete control over the methods and outcome definitions. Consequently, less assumptions had to be made to find suitable data or to adjust available data to the requirements of the state-transition model. This improves robustness of the economic evaluation.

6.4.3 Strengths and limitations

One of the key strengths of the economic evaluation in this chapter was that it projected costs and patient outcomes for an effectiveness study where only the process indicator of HPE rates was available. The decision tree reported in Chapter Three [3.2.1.4] presents the results of the quasi-experimental study on the effectiveness of SMASH. The availability of electronic health records allowed to inform model parameters previously under-reported in the literature [Chapter Four]. The combination with the state-transition model built in Chapter Five [5.2.2] enables the analysis reported in this chapter to extrapolate harm and costs associated with the process indicator of HPE rates. The challenge with combining these to analyses were the different time horizons. The cost per HPE avoided represent the delivery of SMASH over 12 months and the incremental cost per QALY estimated in the presence of the HPE were estimated over a life time horizon. In this thesis, it was assumed that the cost and changes in HPE rates related to SMASH appear in the beginning of the first cycle of the state transition mode instead of gradually over the first 12 months. As a result, some of the patients the state transition model considers to be unexposed to the HPE from the beginning might actually be still exposed. Only later by the end of the 12 months these HPEs are definitely resolved. This double counting could result in an overestimation of the effect of the intervention. However, the effectiveness data of the intervention reported by Peek et al. (2020) indicates a rapid reduction of the HPE rates in the first months and suggests a flooring effect after five months (48). The double counting is, therefore, only present in these first five months before the HPE rates stay almost constant. There was almost no difference in the HPE rate at six and 12 months for NSAID use in anticoagulated patients [Table 6.3]. In theory, this double counting could be problematic if the effect size is very large and huge numbers of patients are affected. However, this is not the case for SMASH. Per practice only one OAC/NSAID HPE is avoided by SMASH over 12 months [Table 6.3] in 91 patients at risk of the HPE type.

A strengths of the economic evaluation, reported in this dissertation, was the validation on multiple levels and stages of the project to improve robustness of the results. In addition to continuous discussions with the supervisory team and the health economists' team from the PROTECT programme, experts were consulted throughout the economic evaluation. Assumptions made when retrospectively estimating the cost of SMASH were based on

interviews with two pharmacists involved in the interventions and the final decisions were run by the interviewees for validation. The use of results from ITSA and the costing of SMASH were reviewed by a health economist and discussed at a conference from the Health Economists' Study Group (HESG). GPs and pharmacist were also included in decisions made to identify confounders in the cohort study in Chapter Four, and results were presented to clinicians and health economists on conferences. The conceptualisation of the decision-analytic model also underwent continuous validation processes throughout the development stages. Face validation was checked by clinical experts (GPs and pharmacists), patient representatives, health economists, as well as qualitative researchers. Assumptions and choices of input parameters were found to be reasonable [Appendix M]. The final work product was reviewed by an experienced health economist not involved in this programme of work, who checked the model structure, the chosen input parameters, the computerised model in *TreeAge Pro Healthcare 2021* and the results. The feedback and how identified problems were dealt with is reported in Appendix N.

As described in the section on comparison with prior work, methodological strengths were the focus on generating HPE-specific harm estimates, instead of generic harm measures. Modelling each HPE individually is considered to generate more precise estimates of harm but has the limitation that the process is time consuming. In an intervention targeting a limited number of HPE types, such as the intervention by Forster et al. (2018) that only looked at the consequences of two administration errors, modelling them separately is manageable. For the original CUA of the PINCER trial, six state-transition models were developed using the best available data from the literature (227). This thesis developed the state-transition model and generated the required increased risk estimates from electronic health records instead of relying on estimates from the literature. In a complex intervention targeting ten HPE types, such as SMASH, considerable time resources are required to conduct this for each HPE type.

Allocation of costs of SMASH

In the economic evaluation, the impact on only a subset of the complex SMASH interventions was analysed. Costs, therefore, had to be allocated to represent the costs relevant for the subset of the intervention’s functions relevant to the study in this chapter.

The probabilistic sensitivity analysis demonstrated how sensitive the incremental NMB of SMASH in reducing NSAID use in anticoagulated patients was to assumptions about the allocation of intervention costs [6.3.4]. This study extrapolated the harm and costs of one HPE type and investigated at the ability of SMASH to reduce NSAID use in anticoagulated patients, allocating a hypothetical portion of the overall cost of SMASH. Where the variable costs that are dependent on the number of HPEs were allocated, the scenarios resulted in higher interventions costs compared to those allocation methods that allocated the total costs of SMASH [Table 6.4]. Allocating the total costs by the proportion of the NSAID/OAC HPE among all HPEs at baseline resulted in the largest incremental NMB. Allocating the total costs assumes that for each HPE type only a proportion of the fixed costs is required. These scenarios are thought to underestimate the intervention cost because the cost for the server, the initial meeting, as well as the training needs would be required independent of the number of HPE types the interventions targets or of the number of HPEs the pharmacists review. Allocating only the variable costs assumes that the fixed costs stay the same because these are not dependent on the number of HPEs or HPE types. However, this represents a hypothetical cost estimate. The intervention was not implemented to only target one HPE type. In reality, SMASH has been designed to reduce the incidence of a suite of HPEs.

These assumptions on how to allocate costs to represent only a subset of the intervention's functions is complex and often subjective. For the base case analysis, the more conservative assumption of allocating the variable costs was chosen. Because this might overestimate the cost of the intervention, the results of other allocation methods were reported alongside. The allocation of costs is not only a problem in this study. In complex interventions, often only subsets of the available functions are analysed or different programme streams are evaluated separately (556). Evaluating subsets allows for comparison between different interventions that target similar subsets, even if the overall interventions include different functions. SMASH and PINCER, for example, aim to reduce hazardous prescribing by targeting specific HPE types. As part of the PROTECT programme grant, these interventions are compared to the clinical decision support system, Optimse-Rx (53). Since this intervention not only includes a function to identify hazardous prescribing as PINCER and SMASH do but it also includes functions to prescribe in a more

cost-effective way, these different parts of the intervention need to be assessed separately. To compare the cost-effectiveness of Optimise-Rx in reducing hazardous prescribing alone to the cost-effectiveness of PINCER and SMASH, the fixed costs need to be allocated to the specific subset of functions. In this chapter, the challenges of allocating the joint costs to a subset were illustrated showing the need to report different allocation methods. This allows an understanding of the impact of the different allocation method on the cost-effectiveness results.

6.4.4 Implications for policy

Subsequent to the third WHO safety challenge aiming to reduce HPE related patient harm by 50% in five years, the UK started various initiatives to reach this goal (33). A new policy objective by the Department of Health and Social Care encourages new interventions aiming to reduce HPEs, such as the PINCER intervention (35). While the new policy calls for the wider rollout of proven interventions in primary care that reduce HPEs, there is no mentioning in the policy's key priorities that this should be achieved in a cost-effective way. Nevertheless, when not all interventions that are effective in reducing HPE rates can be funded, CEAs will gain importance. The metric of cost per QALY is a useful tool to identify which intervention is more cost-effective in reducing HPEs (557). On a UK national level, the measure of cost per QALY is used to prioritise new technologies. The measure of cost per HPE avoided as estimated in Chapter Three is less relevant to UK decision makers at the national level because no willingness to pay threshold is available for cost per HPE avoided.

The measure of cost per QALY, however, might not be useful for decision making processes on all levels (557). Cost per QALY was found to be difficult to interpret (558, 559) and often not relevant on a local level. On a local level, such as for CCGs, measures that directly affect their budgets, and provide short term outcomes, are more important in decision making processes (560). Budget impact analyses focussing on the economic consequences of new interventions in the specific setting and over a short time period are more useful on a local level (561).

The cost per QALY for individual HPE types, as reported in this chapter, could also aid decision making on whether to include a HPE type in the set of HPE types targeted by SMASH. The type of HPEs targeted by SMASH can be adapted to the specific needs of the practices that implement SMASH. The number of HPE types should not overwhelm healthcare professionals working with SMASH to avoid alert fatigue, but the types of HPEs targeted can be changed. This was already done in practice by the Salford CCG. For the implementation of SMASH in Salford, the original set of HPE types was adjusted to the specific requirements of the Salford CCG [2.9.1]. So far, HPE types were included if they were considered to be likely to cause harm, be computable and relevant for GPs (42, 43) or were requested by the practices. The economic evidence provided in this chapter could take this further by choosing HPEs under the premises of maximising cost-effectiveness. An optimal set of HPEs would depend on the prevalence, the harm associated with the HPE type, the costs associated with correcting the HPE and the effectiveness of the intervention at reducing the HPE. Once the cost-effectiveness of multiple HPE types is assessed, decision makers could use this information to inform decisions on which HPE types to include when SMASH is implemented.

The HPE investigated in this dissertation was associated with a high burden for patients and healthcare providers. NSAID use in anticoagulated patients was found to have a substantial impact on patient health and healthcare costs [5.3.1]. This suggests that preventing this HPE is important for patients and decision makers. However, SMASH was not found to significantly reduce the occurrence of NSAID use in anticoagulated patients [6.3.1] and was not cost-effective [6.3.4]. Consequently, the intervention would need to be adapted to provide different information on this HPE to increase effectiveness in reducing its occurrence. Further research could aim to understand the extent to which success or failure of the intervention are a result of features of the intervention or rather the type of HPE. The development and delivery of the intervention was consistent with best practice guidance (245) but there could be features with potential for improvement to make SMASH more effective for this HPE type. Alternatively, other interventions aiming to reduce the occurrence of this severe HPE type would need to be developed.

6.4.5 Implications for patients

After an HPE is identified by SMASH, a pharmacist takes actions to resolve the HPE. For NSAID use in anticoagulated patients, this action involves stopping the NSAID. On a case by case basis, it was checked if the NSAID was necessary and if it was what alternative treatment strategies were available. After consultation with GPs, the patients are most likely switched to paracetamol. These changes can affect patients in several ways. As part of this study, a patient familiar with NSAID treatment was consulted to understand the patient view towards removing the NSAID. The patient raised the sensitive issue of pain management. Individual patients might be reluctant to change a beneficial treatment that allows them to reduce the pain to a manageable level. Paracetamol is known to have a slightly smaller effect on pain reduction than NSAIDs (562). The second issue raised by the patient was the inconvenience of taking paracetamol. While the most common NSAID in the UK, naproxen, is taken once a day, paracetamol is taken three times a day. There might be considerable effects on the patient's adherence to treatment if the dosing frequency is increased (563). Because of these inconveniences of removing the NSAID, patients might be reluctant to resolve the HPE. This could have contributed to the small and non-significant effect SMASH had on the occurrence of this HPE type. The patient highlighted the need to improve communication of the risks associated with NSAIDs. Future research on how to communicate medication changes to resolve HPEs could support effective actions to resolve HPEs.

6.4.6 Considerations for future work

This dissertation assessed the economic impact of the presence and the absence of NSAID use in anticoagulated patients and risk estimates from routinely collected health data were generated for this HPE type. The overall cost-effectiveness of SMASH can only be estimated once the economic impact of each of these HPE types is assessed. Because this comprehensive approach is time consuming, this dissertation focused on one specific HPE type. A similar approach is used for the other HPE types targeted by SMASH and PINCER as part of the five-year PROTECT research programme grant (53). Post-doctoral work will focus on working on estimating the economic impact of the other HPE types as part of PROTECT and to generate an overall estimate of cost-effectiveness of SMASH. The use of

this approach for the other HPE types raises the question of generalisability to other HPE types targeted by SMASH. The following describes the specific steps required to apply the approach reported in this dissertation to the remaining HPEs and hazardous monitoring events identified by SMASH. This section is divided by the different steps of the CUA conducted for this dissertation: (i) generalisability of the cohort study design, (ii) generalisability of the cohort study analysis, and (iii) generalisability of the methods to assess the economic impact of HPEs. Subsequently, it is briefly described how the overall cost-effectiveness of SMASH can be assessed. The final section touches on the impact of SMASH not only on health related outcomes but also on non-health related outcomes.

(i) Generalisability of cohort study design

The first step to assess the economic impact of an HPE type was the quantification of ADE risks in the presence and absence of the HPE. Where high quality data from randomised trials or observational studies from the UK is available on the increased risk of ADEs associated with the HPE type, these can be applied in the economic analysis. For example, for the HPE of prescribing a non-selective beta-blocker to a patient with asthma, high quality estimates were available from the CPRD on the incidence in the at-risk populations and the increased risk of asthma exacerbations in the presence of the HPE (564-566). Where no appropriate data are available, a cohort study as conducted in Chapter Four can be a useful tool to estimate the increased risk of the ADE in the at-risk population and the patients exposed to the HPE. This is possible for all HPE types that include a history of specific diagnoses, information about age or current drug use. All of these are recorded in the CPRD. Clinical code lists were already generated to inform the queries for the dashboard and will not need to be developed from scratch. The only HPE types that potentially require additional code sets and data beyond those recorded in the CPRD were those that included test results. Test results were required for the HPE of the prescription of an oral NSAID to a patient with chronic renal failure (eGFR <45 ml/min). There are diagnoses codes for chronic renal failure, but the analysis could benefit from adding records from CPRD test files that include measurements of eGFR to verify the diagnoses was really associated with a reduced eGFR of less than 45 ml/min.

(ii) Generalisability of the cohort study analysis

After the dataset is structured as time series data with exposed and unexposed time periods for each patient, the data can be analysed. For each HPE type and ADE outcome, the potential confounders could be identified using a DAG and could be validated by clinicians to identify the appropriate set of variables to include in the generation of the propensity score. Five other HPEs are associated with an increased bleeding risk, and risk factors for the GI bleeding outcome were already assessed in this thesis. For the three HPE types, where exposure is defined as a prescription of an NSAID and the associated health outcomes are serious GI events, the same potential confounders can be used as for the propensity score assessment in Chapter Four. For each HPE type, however, other health outcomes might be associated with the HPE that should be considered for analysis. For each health outcome, new ICD-10 code lists would need to be generated and checked if HES data is the appropriate dataset to identify the ADE cases.

Consequently, and depending on the HPE, different adaptions of the identification of exposure times, covariates or outcomes is required, but the overall approach with using CPRD/HES/ONS to identify the patients and use time series analysis to estimate HRs is possible. The ISAC protocol, written as part of this dissertation that was approved in 2018, included data usage for four of the HPE types. For these, no additional costs would occur for the usage of the data. The analysis of these other HPE types will be conducted in post-doctoral work.

(iii) Generalisability of the methods to assess the economic impact of HPEs

The state-transition model structure used to estimate the economic impact of NSAID use in anticoagulated patients was based on ADEs that were associated with the HPE type. The ADEs from the cohort study in Chapter Four and two ADEs predominantly managed in primary care (symptomatic ulcer and GI discomfort) were modelled to project patient related harm and costs. For the other five HPE types associated with an increased risk of serious GI events, the economic impact analysis is based on a similar model. This is work in progress conducted within the PROTECT economics research team. In all six HPE types, the hazardous prescription was an antiplatelet or an NSAID, which were associated with similar

implications on GI related ADEs (serious GI events, symptomatic ulcer and GI discomfort) (567, 568). However, as part of the face validation of the action taken to resolve the HPE conducted for this thesis, it was also assessed how other HPE types would have been resolved that were implicated in an increased risk of serious GI events. Contrary to the removal of the NSAID as the dominant action to resolve the HPE once detected, the addition of a gastroprotective agent was often found to be the action of choice by the pharmacists and GPs questioned. In this case, assumptions on the choice of health states might need to be adapted to account for this change of medication. In the state-transition model on NSAID use in anticoagulated patients in this study, for example, the addition of a gastroprotective agent would reduce the risk of the GI related ADEs, but there is no clinical evidence that this would affect the risk of stroke (569-573). Hence, even though the hazardous prescription of the NSAID might be associated with an increased risk of stroke, this would not be included in the model because the action to resolve the HPE does not affect cardiovascular outcomes. Independent of the possibility to use the state-transition model structure developed for NSAID use in anticoagulated patients, the input parameters would need to be adapted to fit the specific HPE cohort. For the other HPE types, new state-transition models are being constructed as part of the PROTECT programme grant.

Estimating the overall cost-effectiveness

The final step will entail the construction of a composite model. This has been done in the earlier economic evaluation of the PINCER trial (45). By combining the state-transition models, the cost-effectiveness of SMASH can be estimated. The incremental costs and QALYs generated in the presence and absence of the HPE will be weighted by the number of patients with the HPE in an average practice.

Feasibility of assessing the cost-effectiveness of SMASH

Overall, this programme of work showed the complex process of how to estimate the economic impact associated with one HPE type. The methods are assumed to be applicable to the other HPE types as well. In the PROTECT programme, substantial workforce is committed to analyse the health records and to conceptualise the state-transition models for each of the HPE types where no such data or model structures exist.

Non-health benefits of SMASH

The cost-utility analysis is used as a metric to assess the value for money of an intervention. Under the assumption that the change in HPE rates translated into a change in health related quality of life, measured in QALYs, the cost per QALY was estimated. This restricts the value of the intervention to outcomes that affect health status. Non-health benefits, potentially associated with the intervention, are not captured by health related quality of life, which could underestimate the value of the intervention (574, 575). In the context of patient safety interventions, such as SMASH, other factors have been found to impact the value of the interventions and the WTP for the improvements (157). In addition to health related consequences, preventability of the incidences and trust in safety devices or systems were key aspects for patients and decision makers to estimate the value of an intervention. As described in Chapter Two [2.6], evidence on non-health benefits of safety improvements is rare. Future research could identify and quantify the relative importance that patients attach to the outcomes of safer prescribing. This could support the interpretation of the relevance of the cost per QALY generated. Future research could use discrete choice experiments that are increasingly used in health economics and a useful tool to estimate stated preferences of patients and the general population (172, 173). Alternatively, contingent valuation methods have been used valuating safety in health. Compared with DCEs, contingent valuation methods directly assess stated preferences compared with the revealed preference methods DCEs are based on (170). Both methods could be a useful tool to generate an understanding of how patients value non-health benefits with regards to patient safety and reductions of HPEs.

6.4.7 Conclusion

The overall aim of this chapter was to determine the cost-effectiveness of SMASH in reducing NSAID use in anticoagulated patients compared with standard practice beyond the data from the primary effectiveness study. This aim was achieved by combining the HPE specific state-transition model [Chapter Five] with the cost per HPE avoided analysis [Chapter Three] to generate probabilistic cost per additional QALY generated by SMASH. SMASH targets a defined set of HPE types. For the analysis reported in this thesis, NSAID

use in anticoagulated patients was used as an example case study to demonstrate how these different analyses can be combined.

SMASH was found to improve health outcomes at a greater direct costs to the NHS compared to standard care in the hypothetical scenario where SMASH only targets NSAID users with concomitant OAC treatment. For policy making, however, other outputs might need to be accounted for, such as policy incentives, e.g., the WHO third global patient safety challenge and potential non-health benefits of improving safety.

Chapter 7 - Discussion

This chapter summarizes the principal findings reported in previous chapters and describes how the objectives outlined in Chapter One were met. Specific strengths and weaknesses of the different studies were detailed in the individual chapters and are not discussed here. The subsequent section reflects on the methods used in this dissertation to use observational data in economic evaluations. This is divided into a section reflecting on the use of the ITSA design to estimate effectiveness and the feasibility of using routinely collected health data to project patient outcomes from the process indicator of HPE rates. The final section of this chapter discusses the implications of this work for policy. A summary of the key contributions and an overall conclusion of this dissertation is provided at the end of this section.

7.1 Summary of key findings

The aim of this dissertation was to conduct an economic evaluation of a system level DHI as part of a quasi-experimental study that relied on routinely collected health data to measure exposure and outcomes. The intervention in focus was an e-A&F intervention with an integrated pharmacist service aiming to reduce hazardous prescribing in primary care, called SMASH. The cost-effectiveness of SMASH in reducing the number of patients with a specific type of HPE was used as an example to illustrate the different steps required to estimate the cost-effectiveness using routinely collected health data. The specific type of HPE under investigation included patients receiving oral anticoagulation with a concomitant hazardous prescription of an NSAID. To achieve the aim, several objectives were stated in Chapter One. This section contains a summary of the key findings of this dissertation ordered by the objectives proposed in Chapter One.

Objective One was to identify costs associated with the provision and implementation of the SMASH intervention. No guidelines exist on how to conduct economic evaluations of complex system level DHIs where multiple healthcare professionals are involved, such as SMASH (272), and various reviews found existing evidence on cost-effectiveness to be

scarce and of low quality (274-277). One key recommendation from systematic reviews of cost-effectiveness studies of DHIs was the need for more transparent reporting of costing studies (274, 276). Chapter Three reports the micro-costing approach of the cost components identified, as well as the assumptions required to estimate resource use in detail. Key cost components were server costs, the training of pharmacists, an initial meeting with the practice to introduce SMASH, the management of the HPEs identified by SMASH, and IT services. Healthcare professionals involved in the delivery of the interventions validated key assumptions made to increase face validity of the costing approach and were consulted throughout the study. The cost of SMASH per practice was estimated to be £241 during set-up and £1891 for maintaining the intervention for 12 months [Chapter Three].

Objective Two was to assess the cost per HPE avoided by combining the cost of SMASH with the effectiveness data from the quasi-experimental study in a decision-analytic model [Chapter Three]. The method used to derive effect size in the quasi-experimental study was based on a historical comparator, and no information on resource use in the comparator was available. In the cost-effectiveness analysis, SMASH was compared with standard care, assuming that standard care involved no measures that aim at reducing HPEs. The incremental costs of SMASH were £2149 (2.5% to 97.5% credible interval £487 to £5790) at practice-level, and the number of HPEs was reduced by eleven compared with standard care, on average, resulting in £205 (2.5% to 97.5% credible interval £46 to £559) per HPE avoided by SMASH. Expressing outcomes in natural units (cost per HPE avoided) does not allow decision-makers to compare cost-effectiveness estimates across different interventions for patient safety targets other than HPEs avoided. To generate an analysis consistent with the NICE reference case and to inform allocative efficiency of population resources for healthcare, the impact of SMASH on patient outcomes was estimated.

This thesis reports how consequences of HPEs related to patient harm and healthcare costs were projected for one of the HPE types. To quantify patient harm associated with NSAID use in anticoagulated patients, Objective Three was to measure the risk difference in harm outcomes of patients exposed to the HPE and those at risk of the HPE but not exposed using routinely collected health data [Chapter Four]. NSAID use in anticoagulated patients

was associated with an increased risk of serious GI events (HR 2.96, 95% CI 1.60 to 5.46) and stroke (HR 2.48, 95% CI 1.36 to 4.53) based on results from the cohort study using routinely collected health data from linked CPRD/HES and ONS records [Chapter Four]. This is the first UK study that assessed the ADE risk increase associated with NSAID use in anticoagulated patients. This estimate contributes to the evidence on harm from HPEs, which was identified as a gap in the literature in Chapter Two.

In the subsequent step (Objective Four), a state-transition model was conceptualised to model potential treatment pathways related to the consequences of NSAID use in anticoagulated patients informed by findings on ADEs related to the HPE [Chapter Five]. Objective Five was to generate input parameters for estimated harm to populate the state transition model on HPE consequences to estimate the economic impact [Chapter Five]. ADEs managed in secondary care, i.e., stroke and serious GI events, were generated from estimates in the cohort study in Chapter Four. The state-transition model also included ADEs associated with NSAID use managed in primary care: symptomatic ulcer and GI discomfort. This is the first state-transition model that estimates the economic impact of ADEs associated with NSAID use in anticoagulated patients. In previous economic evaluations, the occurrence of ADEs associated with HPEs was often not available and harm from HPEs had to be modelled through proxies, such as assessment of preventability of ADEs or expert elicitation of estimates of potential future harm [Chapter Two]. The overall economic impact of the presence of the HPE was estimated as £244 (2.5% to 97.5% credible interval -£149 to £1073) incremental costs and 0.04 (2.5% to 97.5% credible interval -0.17 to 0.05) reduction in QALYs per patient, from an NHS/PSS perspective, over a life time horizon, with both costs and outcomes discounted at 3.5% [Chapter Five]. Based on the prevalence of the HPE in the English population (estimate from 2019), consequences of the HPE generate additional costs to NHS England of almost £3 million over the life time of these patients and patients lose more than 500 QALYs. This estimate accounts for the probability that the HPE is identified and resolved after a maximum of a year.

To meet Objective Six, the cost-effectiveness of the SMASH intervention in reducing this specific HPE was assessed by combining results on cost per HPE avoided by SMASH with the modelled patient harm and costs associated with the HPE occurrence [Chapter Six].

In the cost-utility analysis, the cost per additional QALY and costs associated with the HPE were incorporated. The cost-utility analysis of SMASH combined the decision-analytic model developed in the cost per HPE avoided analysis [Chapter Three] with the state-transition model on the economic impact of HPEs [Chapter Five]. The cost of SMASH per practice in reducing NSAID use in anticoagulated patients were allocated to the specific HPE type (£380, 2.5% to 97.5% credible interval £347 to £417). SMASH was associated with higher costs (£363, 2.5% to 97.5% credible interval £248 to £424) and 0.003 (2.5% to 97.5% credible interval -0.008 to 0.019) more QALYs compared with standard care under the (fictional) assumption that SMASH only targets NSAID use in anticoagulated patients. At the UK WTP threshold of £20000, the incremental NMB of SMASH in reducing the HPE was estimated to be -£311 (2.5% to 97.5% credible interval -£542 to £73). SMASH would therefore not be cost-effective at the UK WTP and should not be implemented for this one type of HPE alone. However, as discussed in Chapter Six, SMASH was not designed for this one HPE type alone, thus, decisions on whether SMASH overall is cost-effective are only possible once projected harm and costs are available for all HPE types targeted by SMASH. Post-doctoral work will focus on combining the state-transition models for the other HPE types that are developed as part of the PROTECT programme grant with the cost per HPE avoided analysis of SMASH in Chapter Three. Cost-effectiveness of SMASH in reducing NSAID use in anticoagulated patients was mainly driven by the effectiveness of SMASH in reducing HPE rates [Chapter Six]. SMASH reduced overall HPE rates significantly (p-value < 0.005) and did not reduce HPE rates significantly for NSAID use in anticoagulated patients. It is therefore assumed that SMASH can be cost-effective even though it was not cost-effective at the UK WTP for this specific HPE type.

Objective Seven aimed to reflect on conducting an economic evaluation as part of a quasi-experimental study that relied on routinely collected health data to measure exposure and outcomes. The next two sections reflect on the use of evidence from observational data [1.1] from a quasi-experimental effectiveness study and from routinely collected health data [7.4].

7.2 Reflections on using a quasi-experimental effectiveness study

This thesis relied on the use of effectiveness estimates from a quasi-experimental ITSA in the economic evaluation of SMASH. Various quasi-experimental methods have been reported, and these might differ in how they are used in economic evaluations and with regards to strengths and limitations of importance (10). This section reflects on the quasi-experimental method ITSA with historical controls as performed for SMASH. The segmented regression using a historical comparator was the strongest available option given practical and ethical considerations, the lack of a concurrent comparator, and the availability of sufficient pre- and post-intervention measurement points (13). The strengths and limitations of quasi-experimental study designs compared with RCTs have been discussed in the literature in multiple publications (7, 267, 270, 576), and are summarised in Chapter Two [2.9.2]. Strengths and limitations of the quasi-experimental method of ITSA used to evaluate SMASH were described in detail in Chapter Three [3.4.4]. This section focuses on the specific implications of the use of a historical comparator for the economic evaluation.

For SMASH, no concurrent control was available. The intervention was implemented across all practices in the Salford CCG and because of the unique healthcare infrastructure in Salford no suitable comparator existed. The analysis therefore relied on the use of a historical comparator. A recent systematic review by Hategeka et al. (2020) of health system quality improvement interventions using ITSA methods found only 18.3% (22 out of 120) of interventions to use controlled designs (264). Without a concurrent comparator, as a result of potential unmeasured time-varying confounders, it cannot be ruled out that the measured effect would have occurred without the intervention and can be explained away due to confounding factors (259, 577). Unmeasured time-varying confounding can be caused by other unexpected events or intercurrent events and undermine assumptions on causal inference. The use of concurrent controls can minimise the risk of time-varying confounding. As described in Chapter Two [2.9.2] and Chapter Three [3.4.4], controlled ITSA methods were found to produce more robust estimates because they allow within-group and between-group comparisons. The use of a historical comparator has not only implications on the robustness of the effectiveness estimate of the intervention

evaluated [1.1], but it also had implications on the interpretation for the economic evaluation and the cost assessment.

In longitudinal observational studies, a change of the at-risk population over time can be a problem because this violates the key assumption that the pre-intervention trend of the historical comparator is predictive of the current outcome measure (13). In the SMASH effectiveness study, this was not a problem because the at-risk population did not change over the follow-up time of 12 months. This has been criticised in other ITSAs and is difficult to control for (264, 578). The time horizon can be adjusted to only measure the effect in a time interval where the at risk population is constant or some form of concurrent control is necessary in addition to the historic control (259).

Another challenge of the historical control design in this specific example was that resource use of the comparator was not available (48). This had consequences on the cost assessment of the comparator in the economic evaluation. With no information on what measures were in place to reduce HPE rates before the intervention, it was assumed that there were none, so no additional costs were incurred by the comparator. If some of the services of SMASH substituted services already part of standard care before SMASH, the incremental costs of SMASH would have been smaller as mentioned by pharmacists during interviews conducted as part of the cost assessment in Chapter Three.

On reflection, some challenges were encountered in the economic evaluation as a result of the ITSA design. These were mainly a consequence of the use of a historical comparator. With robust study methods to minimise bias and sensitivity analysis to test the impact of assumptions where possible, this dissertation addressed the challenges of using effectiveness estimates from a quasi-experimental study design. However, the use of ITSA data in economic evaluation relies on the assumption that the effectiveness estimate is robust and not biased because of time-varying confounding. Compared with controlled quasi-experimental designs, the ITSA design using historical controls is considered to be less robust (579). With no concurrent control available, the results from the ITSA provide the best available estimate. One way to check validity of the effect size is to compare results with similar studies. PINCER reduced the HPE rate (composite of all HPEs) from 3% to 2%

(OR 0.78, 95% CI 0.64 to 0.94) after 12 months compared with simple feedback (45). SMASH reduced the HPE rate from 2% to 1% compared with standard care before intervention start (48). This suggests that the measured effects in SMASH are of the same order as another similar intervention, which increases confidence in the results to some extent.

7.3 Feasibility of the use of routinely collected health data to link process indicators with patient outcomes

Only process indicators, in the form of HPE rates, were available from the effectiveness study by Peek et al. (2020) (48). While SMASH directly impacts HPE rates, its indirect impact on patient outcomes is not known. This section examines the feasibility of the use of routinely collected health data to link those process indicators to patient outcomes to determine the probability of the patient outcome, given the frequency of the HPE. Firstly, it is reflected on the use of routinely collected health data in the specific case study of NSAID use in anticoagulated patients. Then, the ability to generalise this approach beyond the immediate context of the SMASH intervention is explored. The feasibility of the assessment of health outcomes associated with the HPE presence depends on the definitions of the cohort, the exposure and outcome, the routinely collected health data available and the study design.

Definition of cohort and exposure

The intervention specifies how the cohort and exposure are defined. The cohort at risk and the HPE types were designed as described in Chapter Two [2.9.1] to be identifiable from electronic health records. Consequently, definitions for patients at risk of the HPE (denominator) and exposed to the HPE (numerator) are clearly defined and can, thus, be included in search algorithms. The electronic health records from primary (CPRD Gold) and secondary care (HES) utilised in this thesis provide the data not only to identify patients with the HPE but also variables required to identify periods of exposure to the HPE. This allowed an analysis of exposure as a time-varying variable. Assumptions made to identify exposure start and stop dates were tested in sensitivity analysis and the results were robust to changes in these assumptions [4.4.3].

Dataset

The dataset is crucial to ensure a robust analysis. The results can only be as good as the data available. This has been referred to as a 'fit for purpose' dataset (580). The CPRD dataset has been found to broadly represent the UK population in terms of age, sex and ethnicity (342, 343). With about 600 practices in the UK covering 11.3 million patients, CPRD Gold is one of the largest available longitudinal datasets in primary care (341, 342). This enables analysis of rare outcomes that require large sample sizes. The linkages enable researchers to follow patients through electronic health records in primary and secondary care until death. Validity of diagnostic coding is high (344, 345), as well as the quality of the detailed information on prescribing data (581). For the outcome of stroke, for example, the positive predictive value of a recorded stroke diagnosis in linked CPRD/HES data was high with 79% and a negative predictive value of 100% (433). The linked dataset contains data on patient demographics, prescriptions, diagnoses, referrals, tests, symptoms and life style factors (342, 581). For the analysis, it is important that outcomes, exposure variables and potential confounders are recorded in the dataset.

The assessment of ADEs associated with HPEs avoided by SMASH in this dataset was reasonably robust because the dataset is representative of the study population, sufficiently large, contains high quality data, and provides a comprehensive set of variables, i.e., prescribing records, diagnoses, patient demographic and life-style factors. However, an observational dataset was used that always imposes the risk of unmeasured confounding, missing data and measurement error as discussed in Chapter Four [4.5.4]. While positive predictive values of diagnoses used in the analysis were high, patients without a recorded diagnosis might still have the condition but where never diagnosed. A recent observational study conducted in England demonstrated the impact the COVID-19 pandemic had on recorded diagnosis for different conditions in the Salford Integrated Record (SIR) system, the electronic health record available in the area where SMASH was rolled out (582). The electronic health records recorded much less diagnoses compared with previous years, not because less patients had these conditions, but because they did not report them to their GP. While the COVID-19 pandemic is an extreme example and no such events were present at the time of analysis in this thesis, the example illustrates the potential impact of measurement error in electronic health records.

Study design

A robust study design and analysis plan are key to robust results. Especially in observational studies, these are crucial to minimise selection and confounding bias. In Chapter Four, it was described how different types of confounding were addressed in the cohort study [4.3.7]. This section focuses on providing evidence on whether specific methods used to minimise bias were effective. One study by Kreif et al. (2013) was found that provided a critical appraisal tool to evaluate the study design of observational studies that are used in economic evaluations (11). Kreif et al. (2013) defined five key items relevant to critically appraise statistical methods in observational studies that are used in economic evaluations. Because this was the only guidance found to evaluate if observational studies appropriately address confounding to be used in economic evaluations, this section uses the checklist by Kreif et al. (2013) to critically appraise the study design in Chapter Four.

For the case study in Chapter Four, a robust study design consisting of propensity score matched cohorts was used to balance out observable patient characteristics at baseline. The knowledge derived from clinicians and the literature on potential confounders was used to identify variables to predict the propensity score. PSM is a widely accepted method to reduce treatment selection bias (583). The overlap of baseline covariates between comparators was assessed with histograms and standardised differences before and after the matching as recommended by Kreif et al. (2013). After matching on the propensity score, balance tests showed no substantial difference between the two cohorts at baseline demonstrating the success of the matching process [Table 4.4; Appendix H]. If this can be achieved through matching, the average treatment effect, similar to an RCT, can be estimated (260). If this balance cannot be achieved, other measures such as further regression adjustment would have been required.

However, balanced baseline characteristics only demonstrate that observable variables were evenly distributed between the matched cohorts. Unobserved confounding could still be present. Unobserved confounding is a major challenge of observational studies (428, 584). The study would not be feasible if the risk of unobserved confounding was high. This study addressed unmeasured confounding in different ways. GPs and pharmacists of the PROTECT team were involved in the identification of potential confounders. Potential

confounder and collider relationships were investigated in detail by drawing a DAG based on external literature and discussions with clinicians [Appendix G]. Clinicians involved in the validation were asked if they can think of confounders or colliders not observable in the dataset. No relevant confounder was reported by the clinicians. In addition, sensitivity analysis tested how strong the relation of outcome and exposure with the unmeasured confounder had to be to negate the measured effect using e-values (424). Assessment of the e-value showed that the relation of the unmeasured confounder with the outcome and the exposure must be strong to explain away the identified effect compared with other risk factors observable in the dataset or in the literature (423). The statistics as well as the clinicians suggested that the risk of unmeasured confounding was small [4.5.4]. According to Kreif et al. (2013), the combination of evidence from the literature, involvement of clinical experts and the use of DAGs is the preferred approach to address unmeasured confounding (11).

According to Kreif et al. (2013), it is essential to assess the specification of the regression model. In this thesis, statistical tests and plots were used to assess if the proportional hazard function hold [Appendix H]. The tests were used to identify possible violations of underlying assumptions and indicated no such violations in Chapter Four.

Another pivotal checklist item involved structural uncertainty of the statistical method. To address structural uncertainty, Chapter Four provided estimates for different matching methods, different regression models and changes in exclusion criteria [4.4.3]. The different estimates were use in sensitivity analysis in the economic evaluation and how they impacted the cost-effectiveness analysis is reported in Chapter Six [6.3.5].

In accordance with the checklist by Kreif et al. (2013), the study design in Chapter Four fulfilled all requirements to sufficiently address confounding in observational studies. Kreif et al. (2013) also reviewed the literature on economic evaluations using observational data and critically appraised the statistical methods of these studies according to their checklist (11). The authors found study designs to rarely address confounding appropriately. The study design in Chapter Four, however, addressed confounding appropriately and can be used in economic evaluations.

Summary of feasibility of the use of routinely collected health data

Overall, the assessment of health outcomes associated with the HPE rates measured in the SMASH effectiveness study was found to produce robust estimates. The definitions specified for denominator and numerator by the intervention, the fit for purpose dataset with regards to relevance, availability and reliability of records, and the chosen study design allowed a robust analysis of the increased likelihood of ADEs associated with the HPE. Where no RCT data is available or no RCT is possible, such as in the analysis of ADEs associated with hazardous prescribing, and the analysis relies on the use of real-world evidence, the use of routinely collected health data to link process indicators with patient outcomes was feasible. It is essential, however, to decide on feasibility of the use of evidence from routinely collected health data on a case to case basis. The study designs and, thus, quality of observational studies are highly variable (557), and only robust methods as used in this thesis can provide appropriate evidence to inform economic evaluations.

Generalisability of the applied methods to other settings

The use of routinely collected health data to estimate health outcomes as proposed in this dissertation is not restricted to HPE types targeted by SMASH and PINCER. There are, however, certain criteria that are crucial for this approach to work.

(i) Denominator and numerator need to be clearly defined and be observable in the dataset. The denominator describes the population at risk and the numerator the presence of the process indicator if this is a binary outcome. If the presence of the process indicator is variable over time, for example, the onset and end of an infection or drug treatment, start and stop dates of the exposure periods need to be defined and be available in the dataset. In the linked CPRD/HES dataset, this is manageable for process indicators that are defined by the presence of a prescription or a disease and even some test results. The linked CPRD/HES dataset does not provide information on administration of medications or adherence to treatment. It is therefore not suitable for process indicators, such as administration errors.

(ii) The patient outcome needs to be clearly defined and be observable in the dataset. If only primary care data is available, and the patient outcome is usually managed in hospital, the completeness of recording of these secondary care events in the primary care record needs to be known. Examples for patient outcomes recorded in the CPRD/HES/ONS dataset are diagnosis codes, causes of hospital admissions and mortality. Not recorded in the dataset are life style changes (e.g., diet, quality of life measures or non-health related outcomes). Other datasets might be restricted to only diagnoses recorded in primary care, so that outcomes mainly managed in secondary care might not be appropriate to be investigated in this dataset.

(iii) The patient outcome must be linkable to the process indicator. A causal link between the exposure to the process indicator and the patient outcome needs to be likely, so that the association measured between patient outcome and exposure can be attributable to the exposure. A relevant factor in causality assessment is physiological plausibility (101).

(iv) Availability of data on potential confounders in the dataset is required. One of the key limitations of the use of observational data in general are unobservable confounders. While randomisation balances baseline characteristics between two groups on observable and unobservable confounders, this is not possible using observational data. The methods proposed mimic a randomised process by using PSM. This matching process, however, can only balance out characteristics that are observable in the dataset. An example is confounding by indication, where often not all confounding factors are observed in the dataset and can therefore not all be adjusted for (585). In contrast to RCTs, where exposure is randomly allocated, in routinely collected datasets the exposure assignment by a prescriber is driven by the patient's health state (586, 587). Prescribers are, for example, more likely to prescribe gastroprotective agents to patients at higher bleeding risk, and studies investigating the association of gastroprotective agent use in patients with NSAIDs on GI bleeding risk are at risk of confounding by indication (588). In the example study, this resulted in an increased risk of GI bleedings in patients exposed to the gastroprotective agent. It is not possible in all cases to fully resolve confounding by indication with adjustment methods, such as regression adjustment or matching (585). Observational

datasets are only useful to assess an association between exposure to a medication if confounding by indication is small or can be minimised.

(v) Balance between patient groups must be achievable. Creating balanced cohorts is not always possible. If the proposed propensity score method does not provide balanced comparator groups, other balancing methods can be tested. If patient characteristics differ considerably and sample sizes are low, appropriate matches might be limited.

In under-researched areas, primary data collection, such as in RCTs, with complete control over study design and analysis is not always practical given resource and time constraints and ethical considerations. The use of retrospectively collected data from electronic health records, such as linked CPRD/HES data, enable to analyse a large sample size, requiring considerably less resource use than primary data collection at the cost of no control over the quality of the data recorded. Compared with other alternatives, such as expert elicitation, routinely collected health data provides a less subjective measure of association that can be used in economic evaluation. In conclusion, acknowledging the limitations of routinely collected health data with regards to confounding, the proposed analysis was a valid alternative. The analysis is only feasible for other interventions if the available data source contains comprehensive information on the relevant cohorts, required patient outcomes and potential confounders. Collaboration with clinicians is highly recommended to get an understanding of confounder relations and if the dataset contains the required data.

7.4 Implications of this work for policy

The main objective of this dissertation was to use observational data for the economic assessment of a system level DHI intervention where it was not possible to conduct an RCT. This section outlines implications the use of quasi-experimental methods has for policy decision making.

How evidence from quasi-experimental studies inform policy decisions

Policy decisions should be made on the basis of the best available evidence. There are two main scenarios relevant for policy makers. The first one is to evaluate a new technology or intervention and the second is to evaluate one that is already introduced. The policy decision to be made is slightly different if the intervention is already in place compared to evaluating a new technology. Effectiveness and cost-effectiveness estimates of new technologies assessed alongside an RCT aim to aid decision making on market authorisation (e.g., MHRA in the UK) or reimbursement (e.g., NICE in the UK) of the new treatment. Decision makers then decide whether this new treatment should be used or not. The gold standard for this type of decision making are RCTs (3, 557, 589). In the example of a pragmatic 'real world' study, where the intervention is already in place, the nature of the decision to be made is different (557). The decision to use or implement the intervention has already been made. Effectiveness and cost-effectiveness evidence then aid decision making on whether to roll out the intervention further or to revoke the intervention.

The decision maker on a CCG level in Salford, Greater Manchester, decided to implement SMASH. After implementation, the effectiveness of SMASH was evaluated based on the longitudinal data of this rollout. The evidence on effectiveness of SMASH in reducing HPE rates suggests that SMASH is effective (48) but comes at a higher cost compared with standard care [Chapter Three]. The evidence on cost-effectiveness of SMASH overall, is not assessed yet. Once the economic impact of all HPE types is know, this can provide information of the impact of SMASH not only on HPE rates but also on patient outcomes. Once the cost-utility analysis of SMASH is complete, the results can inform future policy decisions by the Salford CCG. If SMASH overall is cost-effective, the evidence supports decisions to continue with the programme and potentially roll SMASH out further. However, if SMASH is not found to be cost-effective, the evidence from the cost-utility analysis would suggest revoking the intervention. For CCGs, the decision to revoke an existing intervention that was found to be effective and that aims to improve patient safety could be objected by general practices that use and value SMASH. Improving patient safety by reducing HPE rates is not only associated with health related improvements, as they are measured in the cost-utility analysis, but could also have non-health related benefits, for instance, improving trust in the healthcare system [2.6]. The decisions by policy makers

can, thus, often not be made solely on the basis of the cost-effectiveness evidence but need to consider other factors.

Other factors with implications for policy making

The context of when the intervention is being developed and implemented plays a major role in policy decisions. The PINCER intervention is used as an illustrative example where a quasi-experimental study informed policy decisions, where the context played a major role in policy decisions. The effectiveness and cost-effectiveness of PINCER were first assessed in an RCT (45). The positive findings informed policy decisions on a regional rollout in the East Midlands (253). The regional rollout was evaluated using quasi-experimental methods because it was not justifiable to withhold the proven intervention from a subset of practices. The evidence on the effectiveness of PINCER in reducing the process indicator of HPE rates from the quasi-experimental study informed policy decisions on a national rollout of PINCER without any evidence on cost-effectiveness. An important factor that contributed to the national rollout of PINCER was the topicality of the aim of PINCER. The WHO's call to action by policy makers to reduce harm from medication errors in its third global patient safety challenge (33) and the subsequent UK policy objectives to support implementation of interventions aiming to reduce medication related harm, such as PINCER (35), potentially contributed to the policy decision to nationally roll out PINCER.

Additionally, policy developments pushing the introduction of more clinical pharmacists into general practices might have contributed to the decision of rolling out PINCER nationally. The NHS has encouraged the employment of clinical pharmacists in general practices. Since 2015, the Clinical Pharmacists in General Practice Scheme and since 2019, the GP Contract Framework reform the Network Contract Directed Enhanced Services were introduced (244). PINCER and SMASH both support the work of clinical pharmacists in general practices, with SMASH taking full advantage of continuous surveillance by a pharmacist compared with the occasional pharmacist review in PINCER.

The context could have also contributed to the rollout of SMASH to Greater Manchester before evidence on its cost-effectiveness was available. The topicality of the reduction of

HPE rates and the support for practices to employ clinical pharmacists seem to have driven the decision to roll out SMASH on the basis of the effectiveness evidence alone.

In the long term, PINCER and SMASH, however, will need to demonstrate their value for money in order to justify continuous funding needs. The Department of Health and Social Care might need to amend their policy (35) to prioritize interventions that reduce harm from HPEs in a cost-effective way. So far, the UK policy only calls for the wider rollout of interventions such as PINCER and SMASH. Estimates on cost per QALY generated by interventions can enable a comparison of the cost-effectiveness of interventions even if they, for example, reduce different set of HPEs and, thus, avoid other ADEs. The ability to compare the value for money of interventions is essential under the prospect that more and more interventions aiming to reduce hazardous prescribing will be developed in the future.

Guidelines in the use of evidence from quasi-experiments

The NICE evidence standard framework for effectiveness and cost-effectiveness data acknowledges that quasi-experimental evidence can be appropriate to inform policy decisions in interventions with a low risk of harm for the patient (12). The possibility of using quasi-experimental evidence is acknowledged, but no exact guidance on how to use this evidence is provided. A targeted grey literature review of key stakeholder and decision maker websites by Jaksa et al. (2021) identified a lack of comprehensive guidance on how 'high-quality' real world evidence can and should be generated (580). Existing recommendations were found to be fragmented and often did not provide the required level of detail to conduct such analysis. As a result, Deidda et al. (2019) developed the first checklist with best practice guidance for economic evaluations alongside quasi-experimental studies (10). The checklist was designed for prospective economic evaluations that estimate costs and effectiveness outcomes within the same study. Deidda et al. (2019) highlight the need for more comprehensive frameworks that describe a best standard or common approach for the use of quasi-experimental cost-effectiveness evidence (10).

For the specific example in this thesis where costs and patient outcomes were projected based on changes in the process indicator of HPE rates, and costs were estimated retrospectively, the guidance from Deidda et al. (2019) cannot be applied. Future frameworks should address not only challenges with economic evaluations where patient outcomes can be measured alongside the quasi-experimental study but also challenges of economic evaluations that rely on the use of process indicators.

It is possible, and it would be useful, to provide more guidance on the choice of quasi-experimental designs and analysis methods by summarising available designs and methods, comparing their strengths and limitations, and describing implications for the interpretation of economic evidence. For interventions, such as SMASH, the evidence standard framework mentions the use of a high-quality quasi-experimental study as an alternative to RCTs without defining what this term includes (12). In Chapter Two [2.9.2], quasi-experimental methods were described, and their different biases and how they can be mitigated are explained. Not for all studies the same study design is appropriate or can be considered to be of high quality. Controlled designs, e.g., DID and controlled ITSA, were identified as the strongest designs (13, 257, 259, 264). With a suitable control, these controlled designs are able to mitigate time-variant confounding and can be considered high-quality quasi-experimental studies. With no appropriate control, selection bias is introduced that cannot be mitigated and the study design cannot be considered high quality. It was assumed that a high-quality quasi-experimental study as recommended by NICE is defined as a study design that appropriately addresses confounding. In the example of SMASH, where no appropriate control was available, the evaluation relied on the use of historical data from the same practices. For the economic evaluation, this was the only data available. The lack of data from a concurrent control potentially introduces time-variant confounding from intercurrent events that cannot be mitigated. After careful considerations, the study design assessing the effectiveness of SMASH by Peek et al. (2020), was considered to appropriately address potential confounding and to be at low risk of bias [3.4.4; 1.1]. The statistical analysis minimised biases due to regression to the mean, seasonality, maturation bias and time-invariant confounding [2.9.2]. The risk of time-variant confounding was considered low because practices introduced the interventions at different points in time, and the follow-up period was short with 12 months. There was no

indication that the population at risk changed over the study period and no intercurrent events were identified. Under the assumption that studies that appropriately address confounding are considered to be of high quality as recommended by NICE, the ITSA in Peek et al. (2020) is a valid alternative to experimental designs for the specific case of evaluating SMASH.

Guidance, however, is not only required for the choice of study design and method of analysis but it is also pivotal to identify if the data set is fit for purpose. Recommendations exists that describe the need for elements, i.e., relevance, reliability, accuracy and validity of records but these lack detail of how this translates into practice and do not provide thresholds or minimum criteria how to meet these elements (580). A comprehensive guidance could define thresholds of what level of missingness of variables, for example, is still appropriate to generate 'high-quality' results. The thresholds should be specific to the type of variables, such as those used to identify the study population, exposure or covariates used for balancing scores or regression adjustment. In Chapter Four, more than 50% of BMI measurements were missing. High BMI was not considered a confounder but clinicians indicated that it is an indirect risk factor of the serious GI event outcome. It was assumed that the high level of missingness was acceptable and patients with missing BMI were assigned to a 'missing BMI' category for the subsequent PSM. In sensitivity analysis it was explored how excluding patients with missing values would affect the results. The small difference in the estimated HRs was interpreted as an indication that missingness only affects the results to a small extent. For future studies, it would be useful if guidance existed that defines clear criteria on how to meet the proposed standards for data sets to be fit for purpose (580).

A comprehensive guideline from institutions, such as NICE, on how to conduct 'high-quality' real-world evidence studies, should specify criteria for (i) fit for purpose study designs, such as which biases are problematic with different study designs, (ii) appropriate methods of analysis, such as how biases need to be mitigated, and (iii) fit for purpose datasets, such as minimal levels of completeness for specific variables. Future research should focus on comparing the consequences of different quasi-experimental designs for economic evaluations. Consensus methods with researchers and decision makers could be used to

identify best-practice guidelines (580). Including researchers in the development is important to understand methodological strengths and limitations. The inclusion of decision makers would ensure that the recommendations are helpful and acceptable to aid decision making.

7.5 Summary of key contributions

The key contributions of this thesis are as follows:

- Prescribing NSAIDs to anticoagulated patients is a serious hazardous prescribing event that is not only associated with an increase in ADE rates, such as serious GI events and major bleeding, but also with a decreased effectiveness of the anticoagulant therapy in reducing the risk of stroke [Chapter Four].
- The consequences of prescribing NSAIDs to anticoagulated patients impose a health economic burden to the NHS England with higher healthcare costs and fewer QALYs in the presence of the HPE. This highlights the need to develop and implement interventions that effectively and cost-effectively reduce this HPE type [Chapter Five].
- SMASH was not found to be cost-effective in reducing NSAID use in anticoagulated patients. This does not automatically imply that SMASH is not cost-effective overall but indicates that the intervention as it is now might not be suitable to reduce this specific HPE type [Chapter Three, Chapter Six].
- For interventions in cross-therapeutic areas outcomes are variable, as it is often the case for safety interventions. For SMASH, where the HPE types cover different therapeutic areas, a single outcome, such as hospitalisations or mortality, is not useful to fully explain the effect of the intervention [Chapter Two]. This thesis showed the importance to explore consequences of each HPE type individually and illustrated how this can be done using routinely collected data [Chapter Two to Chapter Six].
- To date there are no comprehensive guidance or standards for reliable real world evidence studies (580). This thesis illustrated a robust and transparent approach to conduct an economic evaluation that relied on real world evidence from routinely collected data to estimate exposure and outcomes [Chapter Seven].

7.6 Conclusion

Overall, the methods proposed in this thesis of conducting an economic analysis as part of a quasi-experimental 'real world' study that relies on routinely collected data to measure exposure and outcomes was feasible for the type of HPE investigated in this particular study. The objectives set out at the beginning of this dissertation were met, and this is the first study providing an estimate of cost per HPE avoided for SMASH and illustrating the process of how patient outcomes (QALYs) can be projected from this estimate.

As detailed in Chapter Six [6.4.6], the method of using routinely collected data to inform state-transition models on the economic impact of HPE types is applicable to other HPE types targeted by SMASH. Interventions that aim to reduce the same or similar type of HPEs and where effectiveness is assessed using quasi-experimental study designs, can use the same approach. As part of the PROTECT programme grant, state-transition models for other HPE types are being developed and where no appropriate data on the increased likelihood of ADEs associated with these types of HPEs exists, routinely collected health data is used to estimate this. The state-transition models developed as part of this dissertation and as part of the PROTECT programme grant will be used to estimate the overall cost-effectiveness of SMASH and PINCER in post-doctoral work.

The proposed methods do not depend on the type of HPE or the specific interventions per se. The methods are not restricted to patient safety interventions and patient outcomes associated with HPE rates as a process indicator. The methods are feasible in various potential interventions, especially for those where the process indicator is related to prescribing. This is not limited to process indicators measured in quasi-experimental studies and can be applied to RCTs reporting process indicators as well. Nevertheless, the methods are only feasible if certain criteria are fulfilled relating to data quality and availability of necessary variables in the data source.

References

1. Department of Health. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. Australia; 2016.
2. Ministry of Health and Long-Term Care. Ontario Guidelines for drug submission and evaluation. Ontario; 2016.
3. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London; 2013.
4. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford, UK: Oxford University Press; 2006.
5. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford, UK: Oxford University Press; 2014.
6. Raftery J, Young A, Stanton L, et al. Clinical trial metadata: defining and extracting metadata on the design, conduct, results and costs of 125 randomised clinical trials funded by the National Institute for Health Research Health Technology Assessment programme. Southampton (UK); 2015.
7. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ. 2015;350:h2750.
8. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, et al. Alternatives to the randomized controlled trial. Am J Public Health. 2008;98(8):1359-66.
9. Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. Psychiatry Res. 2020;283:112452.
10. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. Soc Sci Med. 2019;220:353-61.
11. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. Health Econ. 2013;22(4):486-500.
12. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies. 2019.
13. Hinde S, Bojke L, Richardson G. Understanding and addressing the challenges of conducting quantitative evaluation at a local level: a worked example of the available approaches. BMJ Open. 2019;9(11):e029830.
14. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc. 2006;13(1):16-23.
15. Council of the European Union. Council Recommendations of 9 June 2009 on patient safety, including the prevention and control of healthcare associated infections (2009/C 151/01). Brussels; 2009.
16. European Commission. Report form the Commission to the Council on the basis of Member States' reports on the implementation of the Council Recommendation (2009/C 151/01) on patient safety, including the revention and control of healthcare associated infections. Brussels; 2012.
17. NHS Business Services Authority. Prescribing costs 2019 [Available from: https://www.nhsbsa.nhs.uk/prescribing-costs-2019-published].
18. Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events: Implications for prevention. JAMA. 1995;274(1):29-34.
19. Hug BL, Witkowski DJ, Sox CM, Keohane CA, Seger DL, Yoon C, et al. Adverse drug event rates in six community hospitals and the potential impact of computerized physician order entry for prevention. J Gen Intern Med. 2010;25(1):31-8.
20. Jolivot PA, Pichereau C, Hindlet P, Hejblum G, Bige N, Maury E, et al. An observational study of adult admissions to a medical ICU due to adverse drug events. Annals of Intensive Care. 2016;6(1):1-12.

21. Kongkaew C, Noyce PR, Ashcroft DM. Hospital admissions associated with adverse drug reactions: a systematic review of prospective observational studies. Ann Pharmacother. 2008;42(7):1017-25.
22. Hakkarainen KM, Hedna K, Petzold M, Hagg S. Percentage of patients with preventable adverse drug reactions and preventability of adverse drug reactions - a meta-analysis. PLoS One. 2012;7(3):e33236.
23. World Health Organization (WHO). Medication error: Technical Series on Safer Primary Care. Geneva 2016.
24. National Coordinating Council for Medication Error Reporting and Prevention. About medication errors: What is a medication error [Available from: http://www.nccmerp.org/about-medication-errors].
25. Winterstein AG, Sauer BC, Hepler CD, Poole C. Preventable drug-related hospital admissions. Ann Pharmacother. 2002;36(7-8):1238-48.
26. Howard RL, Avery AJ, Slavenburg S, Royal S, Pipe G, Lucassen P, et al. Which drugs cause preventable admissions to hospital? A systematic review. Br J Clin Pharmacol. 2007;63(2):136-47.
27. Kohn L, Corrigan J, Donaldson M. To err is human-building a safer health system. Washington DC: Institute of Medicine; 1999.
28. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. Qual Saf Health Care. 2006;15(3):174-8.
29. Donaldson LJ. An organisation with a memory. Clin Med. 2002;2(5):452-7.
30. Agbabiaka TB, Lietz M, Mira JJ, Warner B. A literature-based economic evaluation of healthcare preventable adverse events in Europe. Int J Qual Health Care. 2017;29(1):9-18.
31. Elliott RA, Camacho E, Campbell F, Jankovic D, Martyn St James M, Kaltenthaler E, et al. Prevalence and economic burden of medication errors in the NHS in England. Rapid evidence synthesis and economic analysis of the prevalence and burden of medication error in the UK. Policy Research Unit in Economic Evaluation of Health and Care Interventions; 2018.
32. National Institute for Health and Care Excellence. Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes. 2015.
33. Donaldson LJ, Kelley ET, Dhingra-Kumar N, Kieny M-P, Sheikh A. Medication Without Harm: WHO's Third Global Patient Safety Challenge. The Lancet. 2017;389(10080):1680-1.
34. World Health Organization (WHO). Metrics: Disability-Adjusted Life Year (DALY) 2020 [Available from: https://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/].
35. Short Life Working Group. The report of the Short Life Working Group on reducing medication-related harm. London; 2018.
36. European Directorate for the Quality of Medicines & HealthCare (EDQM). Pharmaceutical care, policies and practices for a safer, more responsible and cost-effective health system Strasbourg; 2012.
37. Howard RL, Avery AJ, Howard PD, Partridge M. Investigation into the reasons for preventable drug related admissions to a medical admissions unit: observational study. Quality and Safety in Health Care. 2003;12:280-5.
38. Avery AJ, Ghaleb M, Barber N, Dean Franklin B, Armstrong SJ, Serumaga B, et al. The prevalence and nature of prescribing and monitoring errors in English general practice: a retrospective case note review. Br J Gen Pract. 2013;63(613):e543-53.
39. Akbarov A, Kontopantelis E, Sperrin M, Stocks SJ, Williams R, Rodgers S, et al. Primary care medication safety surveillance with integrated primary and secondary care electronic health records: a cross-sectional study. Drug Saf. 2015;38(7):671-82.
40. Williams R, Keers R, Gude WT, Jeffries M, Davies C, Brown B, et al. SMASH! The Salford medication safety dashboard. J Innov Health Inform. 2018;25(3):183-93.
41. National Institute for Health and Care Excellence. Atrial fibrillation: management. 2014. Report No.: ISBN: 978-1-4731-0603-1.
42. Avery AJ, Dex GM, Mulvaney C, Serumaga B, Spencer R, Lester HE, et al. Development of prescribing-safety indicators for GPs using the RAND Appropriateness Method. Br J Gen Pract. 2011;61(589):e526-36.
43. Spencer R, Bell B, Avery AJ, Gookey G, Campbell SM. Identification of an updated set of prescribing--safety indicators for GPs. Br J Gen Pract. 2014;64(621):e181-90.

44. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Boyd M, Cresswell K, et al. PINCER trial: a cluster randomised trial comparing the effectiveness and cost-effectiveness of a pharmacist-led IT-based intervention with simple feedback in reducing rates of clinically important errors in medicines management in general practices. Birmingham; 2010.
45. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. The Lancet. 2012;379(9823):1310-9.
46. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Ann Intern Med. 2016;164(6).
47. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):Cd000259.
48. Peek N, Gude WT, Keers RN, Williams R, Kontopantelis E, Jeffries M, et al. Evaluation of a pharmacist-led actionable audit and feedback intervention for improving medication safety in UK primary care: An interrupted time series analysis. PLoS Med. 2020;17(10):e1003286.
49. Guthrie B, McCowan C, Davey P, Simpson CR, Dreischulte T, Barnett K. High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross sectional population database analysis in Scottish general practice. BMJ. 2011;342:d3514.
50. Dreischulte T, Grant AM, McCowan C, McAnaw JJ, Guthrie B. Quality and safety of medication use in primary care: consensus validation of a new set of explicit medication assessment criteria and prioritisation of topics for improvement. BMC Clin Pharmacol. 2012;12:5-.
51. Dreischulte T, Grant A, Donnan P, McCowan C, Davey P, Petrie D, et al. A cluster randomised stepped wedge trial to evaluate the effectiveness of a multifaceted information technology-based intervention in reducing high-risk prescribing of non-steroidal anti-inflammatory drugs and antiplatelets in primary medical care: The DQIP study protocol. Implementation Science. 2012;7(1):24.
52. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer Prescribing — A Trial of Education, Informatics, and Financial Incentives. N Engl J Med. 2016;374(11):1053-64.
53. Avery A, Elliott R. Avoiding patient harm through the application of prescribing safety indicators in English general practices University of Nottingham: NIHR Programme Grants for Applied Research; 2016.
54. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18820 patients. BMJ. 2004;329(7456):15-9.
55. NHS Business Services Authority. Medication Safety Indicators Specification. 2018.
56. Stocks SJ, Kontopantelis E, Akbarov A, Rodgers S, Avery AJ, Ashcroft DM. Examining variations in prescribing safety in UK general practice: cross sectional study using the Clinical Practice Research Datalink. BMJ. 2015;351:h5501
57. Bradley MC, Fahey T, Cahir C, Bennett K, O'Reilly D, Parsons C, et al. Potentially inappropriate prescribing and cost outcomes for older people: a cross-sectional study using the Northern Ireland Enhanced Prescribing Database. Eur J Clin Pharmacol. 2012;68(10):1425-33.
58. Cahir C, Fahey T, Teeling M, Teljeur C, Feely J, Bennett K. Potentially inappropriate prescribing and cost outcomes for older people: a national population study. Br J Clin Pharmacol. 2010;69(5):543-52.
59. Dean B, Barber N, Schachter M. What is a prescribing error? Qual Health Care. 2000;9(4):232-7.
60. Beers MH. Explicit criteria for determining potentially inappropriate medication use by the elderly: An update. Arch Intern Med. 1997;157(14):1531-6.
61. van der Hooft CS, Jong GW, Dieleman JP, Verhamme KM, van der Cammen TJ, Stricker BH, et al. Inappropriate drug prescribing in older adults: the updated 2002 Beers criteria--a population-based cohort study. Br J Clin Pharmacol. 2005;60(2):137-44.
62. Gallagher P, Ryan C, Byrne S, Kennedy J, O'Mahony D. STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert doctors to Right Treatment). Consensus validation. Int J Clin Pharmacol Ther. 2008;46(2):72-83.

63. American Geriatrics Society Beers Criteria Update Expert Panel. Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2015;63(11):2227-46.
64. Hamilton H, Gallagher P, Ryan C, Byrne S, O'Mahony D. Potentially inappropriate medications defined by STOPP criteria and the risk of adverse drug events in older hospitalized patients. Arch Intern Med. 2011;171(11):1013-9.
65. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. Age Ageing. 2015;44(2):213-8.
66. Ble A, Masoli JA, Barry HE, Winder RE, Tavakoly B, Henley WE, et al. Any versus long-term prescribing of high risk medications in older people using 2012 Beers Criteria: results from three cross-sectional samples of primary care records for 2003/4, 2007/8 and 2011/12. BMC Geriatr. 2015;15:146.
67. Gandhi TK, Seger DL, Bates DW. Identifying drug safety issues: from research to practice. Int J Qual Health Care. 2000;12(1):69-76.
68. Barber N, Franklin BD, Cornford T, Klecun E, Savage I. Safer, faster, better? Evaluating electronic prescribing. London; 2006.
69. Flynn E, Barker K, Pepper G, Bates D, Mikeal R. Comparison of methods for detecting medication errors in 36 hospitals and skilled-nursing facilities. Am J Health Syst Pharm. 2002;59(5):436-46.
70. James KL, Barlow D, McArtney R, Hiom S, Roberts D, Whittlesea C. Incidence, type and causes of dispensing errors: a review of the literature. Int J Pharm Pract. 2009;17(1):9-30.
71. Thomas EJ, Petersen LA. Measuring Errors and Adverse Events in Health Care. J Gen Intern Med. 2003;18(1):61-7.
72. Lewis PJ, Dornan T, Taylor D, Tully MP, Wass V, Ashcroft DM. Prevalence, incidence and nature of prescribing errors in hospital inpatients: a systematic review. Drug Saf. 2009;32(5):379-89.
73. Franklin BD, O'Grady K, Paschalides C, Utley M, Gallivan S, Jacklin A, et al. Providing feedback to hospital doctors about prescribing errors; a pilot study. Pharm World Sci. 2007;29(3):213-20.
74. Cheung K-C, Bouvy ML, De Smet PAGM. Medication errors: the importance of safe dispensing. Br J Clin Pharmacol. 2009;67(6):676-80.
75. Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. Quality and Safety in Health Care. 2008;17(3):170-7.
76. Berdot S, Gillaizeau F, Caruba T, Prognon P, Durieux P, Sabatier B. Drug Administration Errors in Hospital Inpatients: A Systematic Review. PLoS One. 2013;8(6):e68856.
77. Olaniyan JO, Ghaleb M, Dhillon S, Robinson P. Safety of medication use in primary care. Int J Pharm Pract. 2015;23(1):3-20.
78. National Reporting and Learning System. General Practice; Patient Safety Incident Report Form: NHS Improvement; [Available from: https://report.nrls.nhs.uk/GP_eForm].
79. Franklin BD, Vincent C, Schachter M, Barber N. The Incidence of Prescribing Errors in Hospital Inpatients. Drug Saf. 2005;28(10):891-900.
80. Tully MP, Buchan IE. Prescribing errors during hospital inpatient care: factors influencing identification by pharmacists. Pharm World Sci. 2009;31(6):682-8.
81. Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman EJ, et al. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. Qual Saf Health Care. 2007;16(1):40-4.
82. Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, van der Wal G, de Vet HC. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. J Clin Epidemiol. 2010;63(1):94-102.
83. Hanskamp-Sebregts M, Zegers M, Vincent C, van Gurp PJ, de Vet HC, Wollersheim H. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. BMJ Open. 2016;6(8):e011078.

84. Ashcroft DM, Lewis PJ, Tully MP, Farragher TM, Taylor D, Wass V, et al. Prevalence, Nature, Severity and Risk Factors for Prescribing Errors in Hospital Inpatients: Prospective Study in 20 UK Hospitals. Drug Saf. 2015;38(9):833-43.
85. Dornan T, Ashcroft DM, Heathfield H, Lewis P, Miles J, Taylor D, et al. An in depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education. EQUIP study. Manchester: Hope Hospital (University of Manchester Medical School Teaching Hospital); 2009.
86. Keers RN, Williams SD, Vattakatuchery JJ, Brown P, Miller J, Prescott L, et al. Prevalence, nature and predictors of prescribing errors in mental health hospitals: a prospective multicentre study. BMJ Open. 2014;4(9):e006084.
87. Avery AJ, Ghaleb M, Barber N, Franklin BD, Armstrong S, Slight SP, et al. Investigating the prevalence and causes of prescribing errors in general practice: The practice study. Pharmacoepidemiol Drug Saf. 2012;21:4.
88. Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, et al. Inter-rater reliability of case-note audit: a systematic review. J Health Serv Res Policy. 2007;12(3):173-80.
89. Lisby M, Nielsen LP, Brock B, Mainz J. How are medication errors defined? A systematic literature review of definitions and characteristics. Int J Qual Health Care. 2010;22(6):507-18.
90. Miller MR, Robinson KA, Lubomski LH, Rinke ML, Pronovost PJ. Medication errors in paediatric care: a systematic review of epidemiology and an evaluation of evidence supporting reduction strategy recommendations. Qual Saf Health Care. 2007;16(2):116-26.
91. Ferner RE. The epidemiology of medication errors: the methodological difficulties. Br J Clin Pharmacol. 2009;67(6):614-20.
92. Lynskey D, Haigh SJ, Patel N, Macadam AB. Medication errors in community pharmacy: an investigation into the types and potential causes. Int J Pharm Pract. 2007;15(2):105-12.
93. Assiri GA, Shebl NA, Mahmoud MA, Aloudah N, Grant E, Aljadhey H, et al. What is the epidemiology of medication errors, error-related adverse events and risk factors for errors in adults managed in community care contexts? A systematic review of the international literature. BMJ Open. 2018;8(5).
94. Shah NS, Aslam M, Avery A. A survey of prescription errors in general practice. Pharm J. 2001;276:860-2.
95. Quinlan P, Ashcroft DM, Blenkinsopp A. Medication errors: a baseline survey of interventions recorded during the dispensing process in community pharmacies. Int J Pharm Pract. 2002;10(S1):R67-R.
96. NHS Business Services Authority. English Prescribing Dataset - Jun 2020. 2020.
97. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care. 2003;12(3):194-200.
98. Classen DC, Metzger J. Improving medication safety: the measurement conundrum and where to start. Int J Qual Health Care. 2003;15(suppl_1):i41-i7.
99. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv. 1995;21(10):541-8.
100. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. Qual Saf Health Care. 2003;12 Suppl 2(Suppl 2):ii39-45.
101. Pande S. Causality or Relatedness Assessment in Adverse Drug Reaction and Its Relevance in Dermatology. Indian J Dermatol. 2018;63(1):18-21.
102. Jones JK. Adverse drug reactions in the community health setting: approaches to recognizing, counseling, and reporting. 1982;5(2):58-67.
103. Kane-Gill SL, Kirisci L, Pathak DS. Are the Naranjo criteria reliable and valid for determination of adverse drug reactions in the intensive care unit? Ann Pharmacother. 2005;39(11):1823-7.
104. Kramer MS. A Bayesian Approach to Assessment of Adverse Drug Reactions: Evaluation of a Case of Fatal Anaphylaxis. Drug Inf J. 1986;20(4):505-17.
105. Kramer MS, Leventhal JM, Hutchinson TA, Feinstein AR. An Algorithm for the Operational Assessment of Adverse Drug Reactions: I. Background, Description, and Instructions for Use. JAMA. 1979;242(7):623-32.

106. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. 1981;30(2):239-45.
107. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions. Drug Saf. 2008;31(1):21-37.
108. Uppsala Monitoring Centre. The use of the WHO-UMC system for standardized case causality assessment. 2018 [Available from: https://www.who-umc.org/search/?s=criteria&searchButton=].
109. Naranjo CA, Lanctôt KL. A Consultant's View on the Role of Bayesian Differential Diagnosis in the Safety Assessment of Pharmaceuticals. Drug Inf J. 1992;26(4):593-601.
110. Davies EC, Rowe PH, James S, Nickless G, Ganguli A, Danjuma M, et al. An Investigation of Disagreement in Causality Assessment of Adverse Drug Reactions. Pharmaceut Med. 2011;25(1):17-24.
111. Hakkarainen KM, Andersson Sundell K, Petzold M, Hägg S. Methods for assessing the preventability of adverse drug events: a systematic review. Drug Saf. 2012;35(2):105-26.
112. Davies EC, Green CF, Mottram DR, Rowe PH, Pirmohamed M. Emergency re-admissions to hospital due to adverse drug reactions within 1 year of the index admission. Br J Clin Pharmacol. 2010;70(5):749-55.
113. Gallagher RM, Mason JR, Bird KA, Kirkham JJ, Peak M, Williamson PR, et al. Adverse Drug Reactions Causing Admission to a Paediatric Hospital. PLoS One. 2012;7(12):e50127.
114. Hallas J, Harvald B, Gram LF, Grodum E, BrØSen K, Haghfelt T, et al. Drug related hospital admissions: the role of definitions and intensity of data collection, and the possibility of prevention. J Intern Med. 1990;228(2):83-90.
115. Schumock GT, Thornton JP. Focusing on the preventability of adverse drug reactions. Hosp Pharm. 1992;27(6):538.
116. Hodkinson A, Tyler N, Ashcroft DM, Keers RN, Khan K, Phipps D, et al. Preventable medication harm across health care settings: a systematic review and meta-analysis. BMC Med. 2020;18(1):313.
117. Panagioti M, Khan K, Keers R, Abuzour A, Phipps D, Bower P, et al. Preventable patient harm across health care services: a systematic review and meta-analysis. 2017.
118. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. BMJ. 2019;366:l4185.
119. Alqenae FA, Steinke D, Keers RN. Prevalence and Nature of Medication Errors and Medication-Related Harm Following Discharge from Hospital to Community Settings: A Systematic Review. Drug Saf. 2020;43(6):517-37.
120. Morris CJ, Rodgers S, Hammersley VS, Avery AJ, Cantrill JA. Indicators for preventable drug related morbidity: application in primary care. Quality and Safety in Health Care. 2004;13(3):181-5.
121. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse Drug Reactions in Hospital In-Patients: A Prospective Analysis of 3695 Patient-Episodes. PLoS One. 2009;4(2):e4439.
122. Tangiisuran B, Davies JG, Wright JE, Rajkumar C. Adverse drug reactions in a population of hospitalized very elderly patients. Drugs Aging. 2012;29(8):669-79.
123. Gallagher AM, de Vries F, Plumb JM, Hass B, Clemens A, van Staa TP. Quality of INR control and outcomes following venous thromboembolism. Clin Appl Thromb Hemost. 2012;18(4):370-8.
124. Parekh N, Ali K, Stevenson JM, Davies JG, Schiff R, Van der Cammen T, et al. Incidence and cost of medication harm in older adults following hospital discharge: a multicentre prospective study in the UK. Br J Clin Pharmacol. 2018;84(8):1789-97.
125. National Institute for Health and Care Excellence. NICE glossary [Available from: https://www.nice.org.uk/glossary?letter=q].
126. World Health Organization (WHO). International Drug Monitoring: The Role of National Centers. Geneva; 1972.
127. Garfield S, Reynolds M, Dermont L, Franklin BD. Measuring the severity of prescribing errors: a systematic review. Drug Saf. 2013;36(12):1151-7.

128. Dale MA, Copeland R, Barton R. Prescribing errors on medical wards and the impact of clinical pharmacists. Int J Pharm Pract. 2003;11(1):19-24.
129. Shulman R, Singer M, Goldstone J, Bellingan G. Medication errors: a prospective cohort study of hand-written and computerised physician order entry in the intensive care unit. Crit Care. 2005;9(5):R516-21.
130. Dobrzanski S, Hammond I, Khan G, Holdsworth H. The nature of hospital prescribing errors. British Journal of Clinical Governance. 2002;7(3):187-93.
131. Abdel-Qader DH, Harper L, Cantrill JA, Tully MP. Pharmacists' interventions in prescribing errors at hospital discharge: an observational study in the context of an electronic prescribing system in a UK teaching hospital. Drug Saf. 2010;33(11):1027-44.
132. Dean B, Schachter M, Vincent C, Barber N. Prescribing errors in hospital inpatients: their incidence and clinical significance. Qual Saf Health Care. 2002;11(4):340-4.
133. Tully MP, Parker D, Buchan I, al. e. Patient Safety Research Programme: Medication errors 2: Pilot study. 2006.
134. Webbe DD, S.; Roberts, C.M. Improving junior doctor prescribing: the positive impact of a pharmacist intervention. Pharm J. 2007;278:136-9.
135. Haw C, Stubbs J. Prescribing errors at a psychiatric hospital. Pharm Pract. 2003;13:64–6.
136. Conroy S. Association between licence status and medication errors. Arch Dis Child. 2011;96(3):305-6.
137. Ridley SA, Booth SA, Thompson CM, Intensive Care Society's Working Group on Adverse Incidents. Prescription errors in UK critical care units. Anaesthesia. 2004;59(12):1193-200.
138. Stubbs J, Haw C, Taylor D. Prescription errors in psychiatry - a multi-centre study. J Psychopharm. 2006;20(4):553-61.
139. Sagripanti M, Dean B. An evaluation of the process-related medication risks for elective surgery patients from pre-operative assessment to discharge. Int J Pharm Pract. 2002;10(3):161-70.
140. Wang JK, Herzog NS, Kaushal R, Park C, Mochizuki C, Weingarten SR. Prevention of pediatric medication errors by hospital pharmacists and the potential benefit of computerized physician order entry. Pediatrics. 2007;119(1):e77-85.
141. Rees S, Thomas P, Shetty A, al. e. Drug history errors in the acute medical assessment unit quantified by use of the NPSA classification. Pharm J. 2007;279:469–71.
142. National Patient Safety Agency. Safety in doses: medication safety incidents in the NHS. London; 2007.
143. Franklin BD, Reynolds M, Shebl NA, Burnett S, Jacklin A. Prescribing errors in hospital inpatients: a three-centre study of their prevalence, types and causes. Postgrad Med J. 2011;87(1033):739-45.
144. Walsh KE, Harik P, Mazor KM, Perfetto D, Anatchkova M, Biggins C, et al. Measuring Harm in Health Care: Optimizing Adverse Event Review. Med Care. 2017;55(4):436-41.
145. Alldred D, Barber N, Buckle P, Carpenter J, Dickinson R, Franklin B. Care home use of medicines study (CHUMS) Medication errors in nursing and residential care homes–prevalence, consequences, causes and solutions. Report to the Patient Safety Research Portfolio. 2009.
146. Walsh EK, Hansen CR, Sahm LJ, Kearney PM, Doherty E, Bradley CP. Economic impact of medication error: a systematic review. Pharmacoepidemiol Drug Saf. 2017;26(5):481-97.
147. Kind P, Lafata JE, Matuszewski K, Raisch D. The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects. Value Health. 2009;12(s1):S27-S30.
148. Wallace E, McDowell R, Bennett K, Fahey T, Smith SM. Impact of Potentially Inappropriate Prescribing on Adverse Drug Events, Health Related Quality of Life and Emergency Hospital Attendance in Older People Attending General Practice: A Prospective Cohort Study. J Gerontol A Biol Sci Med Sci. 2017;72(2):271-7.
149. Elliott RA, Putman K, Franklin M, Verhaeghe N, Annemans L, Eden M, et al. Economic evaluation of a pharmacist-led IT-based intervention with simple feedback in reducing rates of clinically important errors in medicines management in general practices (PINCER). Birmingham, UK; 2013.

150. Forster R, Ratcliffe A, Lewis M, Crossley A, Bastida JL, Dunlop WCN. Cost-utility analysis of an intervention designed to reduce the critical handling error of insufficient inspiratory effort. Eur J Health Econ. 2018;19(9):1303-18.
151. Moriarty F, Cahir C, Bennett K, Fahey T. Economic impact of potentially inappropriate prescribing and related adverse events in older people: a cost-utility analysis using Markov models. BMJ Open. 2019;9(1):e021832.
152. Foy R, Willis T, Glidewell L, McEachan R, Lawton R, Meads D, et al. Developing and evaluating packages to support implementation of quality indicators in general practice: the ASPIRE research programme, including two cluster RCTs. Southampton (UK); 2020.
153. Patel K, Jay R, Shahzad MW, Green W, Patel R. A systematic review of approaches for calculating the cost of medication errors. European Journal of Hospital Pharmacy. 2016;23(5):294-301.
154. Cranshaw J, Gupta KJ, Cook TM. Litigation related to drug errors in anaesthesia: an analysis of claims against the NHS in England 1995-2007. Anaesthesia. 2009;64(12):1317-23.
155. Zaidi S, Mordaunt C, Durnin N, Cooke S, Blakey JD. Quantifying and reducing inhaler prescription errors in secondary care. Int J Clin Pharm. 2015;37(6):1028-32.
156. Karnon J, Campbell F, Czoski-Murray C. Model-based cost-effectiveness analysis of interventions aimed at preventing medication error at hospital admission (medicines reconciliation). J Eval Clin Pract. 2009;15(2):299-306.
157. Steuten L, Buxton M. Economic evaluation of healthcare safety: which attributes of safety do healthcare professionals consider most important in resource allocation decisions? Qual Saf Health Care. 2010;19(5):e6.
158. Shah K, Praet C, Devlin N, Sussex J, Appleby J, Parkin D. Is the aim of the English health care system to maximize QALYs? J Health Serv Res Policy. 2012;17(3):157-63.
159. Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. Mayo Clin Proc. 2014;89(8):1116-25.
160. Whittaker CF, Miklich MA, Patel RS, Fink JC. Medication Safety Principles and Practice in CKD. Clin J Am Soc Nephrol. 2018;13(11):1738-46.
161. Offe C. How can we trust our fellow citizens? (1999). In: Institutionen, Normen, Bürgertugenden. vol. 3. Wiesbaden: Springer Fachmedien Wiesbaden; 1999. p. 147-90.
162. Quick O. Outing Medical Errors: Questions of Trust and Responsibility. Med Law Rev. 2005;14(1):22-43.
163. Altice FL, Mostashari F, Friedland GH. Trust and the acceptance of and adherence to antiretroviral therapy. J Acquir Immune Defic Syndr. 2001;28(1):47-58.
164. Krause TM, Yay Donderici E, Ganduglia Cazaban C, Franzini L. Future expenditure risk of silent members: a statistical analysis. BMC Health Serv Res. 2016;16:319.
165. LaVeist TA, Isaac LA, Williams KP. Mistrust of health care organizations is associated with underutilization of health services. Health Serv Res. 2009;44(6):2093-105.
166. Birkhauer J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: A meta-analysis. PLoS One. 2017;12(2):e0170988.
167. Smith CP. First, do no harm: institutional betrayal and trust in health care organizations. Journal of multidisciplinary healthcare. 2017;10:133-44.
168. Thompson TL. Encyclopedia of Health Communication: SAGE Publications; 2014.
169. Slawomirski L, Auraaen A, Klazinga N. The Economics of Patient Safety: Strengthening a value-based approach to reducing patient harm at national level. 2017.
170. Perry-Duxbury M, van Exel J, Brouwer W. How to value safety in economic evaluations in health care? A review of applications in different sectors. Eur J Health Econ. 2019;20(7):1041-61.
171. Entwistle V, Quick O. Trust in the context of patient safety problems. Journal of health organization and management. 2006;20:397-416.
172. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.
173. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145-72.

174. Klose T. The contingent valuation method in health care. Health Policy. 1999;47(2):97-123.
175. O'Brien B, Gafni A. When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. Med Decis Making. 1996;16(3):288-99.
176. Halvorsen B. Ordering effects in contingent valuation surveys. Environ Resource Econ. 1996;8(4):485-99.
177. Shafie AA, Hassali MA. Willingness to pay for a pharmacist's dispensing service: a cross-sectional pilot study in the state of Penang, Malaysia. Pharm Pract (Granada). 2010;8(2):116-21.
178. Beattie J, Covey J, Dolan P, Hopkins L, Jones-Lee M, Loomes G, et al. On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1-Caveat Investigator. J Risk Uncertainty. 1998;17(1):5-26.
179. Kaur S, Mitchell G, Vitetta L, Roberts MS. Interventions that can Reduce Inappropriate Prescribing in the Elderly. Drugs Aging. 2009;26(12):1013-28.
180. Batuwitage BT, Kingham JG, Morgan NE, Bartlett RL. Inappropriate prescribing of proton pump inhibitors in primary care. Postgrad Med J. 2007;83(975):66-8.
181. Pimlott NJ, Hux JE, Wilson LM, Kahan M, Li C, Rosser WW. Educating physicians to reduce benzodiazepine use by elderly patients: a randomized controlled trial. CMAJ. 2003;168(7):835-9.
182. Brown B, Peek N, Buchan I. The Case for Conceptual and Computable Cross-Fertilization Between Audit and Feedback and Clinical Decision Support. Stud Health Technol Inform. 2015;216:419-23.
183. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17.
184. Gurwitz JH, Field TS, Ogarek J, Tjia J, Cutrona SL, Harrold LR, et al. An electronic health record-based intervention to increase follow-up office visits and decrease rehospitalization in older adults. J Am Geriatr Soc. 2014;62(5):865-71.
185. Murray MD, Harris LE, Overhage JM, Zhou XH, Eckert GJ, Smith FE, et al. Failure of computerized treatment suggestions to improve health outcomes of outpatients with uncomplicated hypertension: results of a randomized controlled trial. Pharmacotherapy. 2004;24(3):324-37.
186. Riordan DO, Walsh KA, Galvin R, Sinnott C, Kearney PM, Byrne S. The effect of pharmacist-led interventions in optimising prescribing in older adults in primary care: A systematic review. SAGE Open Med. 2016;4:2050312116652568.
187. Khalil H, Bell B, Chambers H, Sheikh A, Avery AJ. Professional, structural and organisational interventions in primary care for reducing medication errors. Cochrane Database Syst Rev. 2017;10:CD003942.
188. Garcia RM. Five ways you can reduce inappropriate prescribing in the elderly: a systematic review. J Fam Pract. 2006;55(4):305-12.
189. Cooper JA, Cadogan CA, Patterson SM, Kerse N, Bradley MC, Ryan C, et al. Interventions to improve the appropriate use of polypharmacy in older people: a Cochrane systematic review. BMJ Open. 2015;5(12):e009235.
190. Zermansky AG, Petty DR, Raynor DK, Freemantle N, Vail A, Lowe CJ. Randomised controlled trial of clinical medication review by a pharmacist of elderly patients receiving repeat prescriptions in general practice. BMJ. 2001;323(7325):1340-3.
191. Lowrie R, Mair FS, Greenlaw N, Forsyth P, Jhund PS, McConnachie A, et al. Pharmacist intervention in primary care to improve outcomes in patients with left ventricular systolic dysfunction. Eur Heart J. 2012;33(3):314-24.
192. Bryant LJ, Coster G, Gamble GD, McCormick RN. The General Practitioner-Pharmacist Collaboration (GPPC) study: a randomised controlled trial of clinical medication reviews in community pharmacy. Int J Pharm Pract. 2011;19(2):94-105.
193. Hanlon JT, Weinberger M, Samsa GP, Schmader KE, Uttech KM, Lewis IK, et al. A randomized, controlled trial of a clinical pharmacist intervention to improve inappropriate prescribing in elderly outpatients with polypharmacy. Am J Med. 1996;100(4):428-37.

194. Richmond S, Morton V, Cross B, Wong ICK, Russell I, Philips Z, et al. Effectiveness of shared pharmaceutical care for older patients: RESPECT trial findings. Br J Gen Pract. 2010;60(570):e10.
195. Taylor CT, Byrd DC, Krueger K. Improving primary care in rural Alabama with a pharmacy initiative. Am J Health Syst Pharm. 2003;60(11):1123-9.
196. Krska J, Cromarty JA, Arris F, Jamieson D, Hansford D, Duffus PR, et al. Pharmacist-led medication review in patients over 65: a randomized, controlled trial in primary care. Age Ageing. 2001;30(3):205-11.
197. Crotty M, Rowett D, Spurling L, Giles LC, Phillips PA. Does the addition of a pharmacist transition coordinator improve evidence-based medication management and health outcomes in older adults moving from the hospital to a long-term care facility? Results of a randomized, controlled trial. Am J Geriatr Pharmacother. 2004;2(4):257-64.
198. Brown BK, Earnhart J. Pharmacists and their effectiveness in ensuring the appropriateness of the chronic medication regimens of geriatric inpatients. Consult Pharm. 2004;19(5):432-6.
199. Rhoads M, Thai A. Physician acceptance rate of pharmacist recommendations to reduce use of potentially inappropriate medications in the assisted living setting. Consult Pharm. 2003;18(3):241-7.
200. de Araujo BC, de Melo RC, de Bortoli MC, Bonfim JRA, Toma TS. How to Prevent or Reduce Prescribing Errors: An Evidence Brief for Policy. Front Pharmacol. 2019;10:439.
201. Reis WC, Bonetti AF, Bottacin WE, Reis AS, Jr., Souza TT, Pontarolo R, et al. Impact on process results of clinical decision support systems (CDSSs) applied to medication use: overview of systematic reviews. Pharm Pract (Granada). 2017;15(4):1036.
202. Lainer M, Mann E, Sönnichsen A. Information technology interventions to improve medication safety in primary care: a systematic review. Int J Qual Health Care. 2013;25(5):590-8.
203. Manias E, Kusljic S, Wu A. Interventions to reduce medication errors in adult medical and surgical settings: a systematic review. Ther Adv Drug Saf. 2020;11:2042098620968309.
204. Royal S, Smeaton L, Avery AJ, Hurwitz B, Sheikh A. Interventions in primary care to reduce medication related adverse events and hospital admissions: systematic review and meta-analysis. Quality and Safety in Health Care. 2006;15(1):23-31.
205. Omer U, Danopoulos E, Veysey M, Crampton P, Finn G. A Rapid Review of Prescribing Education Interventions. Medical Science Educator. 2020.
206. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, et al. Educational outreach visits: effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2007;2007(4):Cd000409.
207. Greenhalgh T, Swinglehurst D, Stones R. Health Services and Delivery Research. Rethinking resistance to 'big IT': a sociological study of why and when healthcare staff do not use nationally mandated information and communication technologies. Southampton (UK): NIHR Journals Library; 2014.
208. Jeffries M, Phipps DL, Howard RL, Avery AJ, Rodgers S, Ashcroft DM. Understanding the implementation and adoption of a technological intervention to improve medication safety in primary care: a realist evaluation. BMC Health Serv Res. 2017;17(1):196.
209. Burgin A, O'Rourke R, Tully MP. Learning to work with electronic patient records and prescription charts: experiences and perceptions of hospital pharmacists. Res Social Adm Pharm. 2014;10(5):741-55.
210. Luetsch K, Rowett D, Twigg MJ. A realist synthesis of pharmacist-conducted medication reviews in primary care after leaving hospital: what works for whom and why? BMJ Quality & Safety. 2020:bmjqs-2020-011418.
211. Thielen FW, Van Mastrigt G, Burgers LT, Bramer WM, Majoie H, Evers S, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). Expert Rev Pharmacoecon Outcomes Res. 2016;16(6):705-21.
212. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak. 2007;7:16.

213. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. 2014.
214. National Institute for Health and Care Excellence. NICE guideline 50-Cirrhosis in over 16s-Appendices A–H. 2016.
215. National Institute for Health and Care Excellence. NICE guideline 81-Glaucoma-Appendices A-T. 2017.
216. Centre for Reviews and Dissemination. Search strategies 2019 [Available from: http://www.crd.york.ac.uk/crdweb/searchstrategies.asp#nhseedmedline].
217. National Institute for Health and Clinical Excellence. NICE guideline 59-Low back pain and sciatica in over 16s: assessment and management-Appendix. 2016.
218. Forrester SH, Hepp Z, Roth JA, Wirtz HS, Devine EB. Cost-effectiveness of a computerized provider order entry system in improving medication safety ambulatory care. Value Health. 2014;17(4):340-9.
219. Ghatnekar O, Bondesson A, Persson U, Eriksson T. Health economic evaluation of the Lund Integrated Medicines Management Model (LIMM) in elderly patients admitted to hospital. BMJ Open. 2013;3(1).
220. Maviglia SM, Yoo JY, Franz C, Featherstone E, Churchill W, Bates DW, et al. Cost-benefit analysis of a hospital pharmacy bar code solution. Arch Intern Med. 2007;167(8):788-94.
221. Nerich V, Borg C, Villanueva C, Thiery-Vuillemin A, Helias P, Rohrlich PS, et al. Economic impact of prescribing error prevention with computerized physician order entry of injectable antineoplastic drugs. J Oncol Pharm Pract. 2013;19(1):8-17.
222. Nuckols TK, Asch SM, Patel V, Keeler E, Anderson L, Buntin MB, et al. Implementing Computerized Provider Order Entry in Acute Care Hospitals in the United States Could Generate Substantial Savings to Society. Jt Comm J Qual Patient Saf. 2015;41(8):341-50.
223. Rosselli D, Rueda JD, Silva MD, Salcedo J. Economic Evaluation of Four Drug Administration Systems in Intensive Care Units in Colombia. Value Health Reg Issues. 2014;5:20-4.
224. Samp JC, Touchette DR, Marinac JS, Kuo GM. Economic Evaluation of the Impact of Medication Errors Reported by U.S. Clinical Pharmacists. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2014;34(4):350-7.
225. Westbrook JI, Gospodarevskaya E, Li L, Richardson KL, Roffe D, Heywood M, et al. Cost-effectiveness analysis of a hospital electronic medication management system. J Am Med Inform Assoc. 2015;22(4):784-93.
226. Yao GL, Novielli N, Manaseki-Holland S, Chen Y-F, van der Klink M, Barach P, et al. Evaluation of a predevelopment service delivery intervention: an application to improve clinical handovers. BMJ Quality & Safety. 2012;21(Suppl 1):i29-i38.
227. Elliott RA, Putman KD, Franklin M, Annemans L, Verhaeghe N, Eden M, et al. Cost Effectiveness of a Pharmacist-Led Information Technology Intervention for Reducing Rates of Clinically Important Errors in Medicines Management in General Practices (PINCER). Pharmacoeconomics. 2014:1-18.
228. Devine EB, Hansen RN, Wilson-Norton JL, Lawless NM, Fisk AW, Blough DK, et al. The impact of computerized provider order entry on medication errors in a multispecialty group practice. J Am Med Inform Assoc. 2010;17(1):78-84.
229. Hellstrom LM, Bondesson A, Hoglund P, Midlov P, Holmdahl L, Rickhag E, et al. Impact of the Lund Integrated Medicines Management (LIMM) model on medication appropriateness and drug-related hospital revisits. Eur J Clin Pharmacol. 2011;67(7):741-52.
230. Midlov P, Deierborg E, Holmdahl L, Hoglund P, Eriksson T. Clinical outcomes from the use of Medication Report when elderly patients are discharged from hospital. Pharm World Sci. 2008;30(6):840-5.
231. Hatoum HT, Hutchinson RA, Elliott LR, Kendzierski DL. Physicians' review of significant interventions by clinical pharmacists in inpatient care. Drug Intell Clin Pharm. 1988;22(12):980-2.
232. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The Incidence and Severity of Adverse Events Affecting Patients after Discharge from the Hospital. Ann Intern Med. 2003;138(3):161-7.
233. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of Adverse Events and Negligence in Hospitalized Patients. N Engl J Med. 1991;324(6):370-6.

234. Folli HL, Poole RL, Benitz WE, Russo JC. Medication error prevention by clinical pharmacists in two children's hospitals. Pediatrics. 1987;79(5):718-22.
235. National Coordinating Council for Medication Error Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors 2001 [Available from: http://www.nccmerp.org/types-medication-errors].
236. Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, Ma'luf N, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. The American Journal of Medicine. 1999;106(2):144-50.
237. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA. 1998;280(15):1311-6.
238. van Doormaal JE, van den Bemt PM, Zaal RJ, Egberts AC, Lenderink BW, Kosterink JG, et al. The influence that electronic prescribing has on medication errors and preventable adverse drug events: an interrupted time-series study. J Am Med Inform Assoc. 2009;16(6):816-25.
239. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA. 1997;277(4):307-11.
240. Hug BL, Keohane C, Seger DL, Yoon C, Bates DW. The costs of adverse drug events in community hospitals. Jt Comm J Qual Patient Saf. 2012;38(3):120-6.
241. Zsifkovits J, Zuba M, Geißler W, Lepuschütz L, Pertl D, Kernstock E, et al. Costs of unsafe care and cost-effectiveness of patient safety programmes. Gesundheit Österreich Forschungs-und Planungs GmbH and SOGETI; 2016.
242. Nuckols TK, Smith-Spangler C, Morton SC, Asch SM, Patel VM, Anderson LJ, et al. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. Systematic Reviews. 2014;3(1):56.
243. Jeffries M, Keers RN, Phipps DL, Williams R, Brown B, Avery AJ, et al. Developing a learning health system: Insights from a qualitative process evaluation of a pharmacist-led electronic audit and feedback intervention to improve medication safety in primary care. PLoS One. 2018;13(10):e0205419.
244. National Health Service (NHS England). Clinical Pharmacists 2019 [Available from: https://www.england.nhs.uk/gp/expanding-our-workforce/cp-gp/].
245. Jamtvedt G, Flottorp S, N. I. Audit and Feedback as a Quality Strategy. In: Busse R, Klazinga N, Panteli D, et al., editors. Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies [Internet]. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2019.
246. Jeffries M. The implementation of electronic audit and feedback systems for medicines optimisation in primary care: understandings from a sociotechnical perspective: University of Manchester; 2017.
247. Jeffries M, Gude WT, Keers RN, Phipps DL, Williams R, Kontopantelis E, et al. Understanding the utilisation of a novel interactive electronic medication safety dashboard in general practice: a mixed methods study. BMC Med Inform Decis Mak. 2020;20(1):69.
248. Coleman EA, Grothaus LC, Sandhu N, Wagner EH. Chronic care clinics: a randomized controlled trial of a new model of primary care for frail older adults. J Am Geriatr Soc. 1999;47(7):775-83.
249. Wasylewicz ATM, Scheepers-Hoeks AMJW. Clinical Decision Support Systems. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer International Publishing; 2019. p. 153-69.
250. U.S. National Library of Medicine. Medical Subject Headings (MeSH) - alert fatigue, health personnel 2017 [Available from: https://www.ncbi.nlm.nih.gov/mesh?term=alert%20fatigue].
251. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica; 2001.
252. PRIMIS. The PINCER tool [Available from: https://www.nottingham.ac.uk/primis/tools-audits/tools-audits/pincer/pincer.aspx].

253. Rodgers S, Salema N, Waring J, Armstrong S, Mehta R, Bell B, et al. Improving medication safety in general practices in the East Midlands through the PINCER intervention: Scaling Up PINCER. Report to the Health Foundation. 2018.
254. Cartwright N. What are randomised controlled trials good for? Philosophical Studies. 2009;147(1):59.
255. Schweizer ML, Braun BI, Milstone AM. Research Methods in Healthcare Epidemiology and Antimicrobial Stewardship-Quasi-Experimental Designs. Infect Control Hosp Epidemiol. 2016;37(10):1135-40.
256. White H, Sabarwal S. Quasi-Experimental Design and Methods. Florence; 2014.
257. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand S-L. Prospective Observational Studies to Assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report. Value Health. 2012;15(2):217-30.
258. Fredriksson A, Oliveira GMd. Impact evaluation using Difference-in-Differences. RAUSP Management Journal. 2019;54(4):519-32.
259. Bernal JL, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol. 2018;47(6):2082-93.
260. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011;46(3):399-424.
261. Abadie A, Diamond A, Hainmueller J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. J Amer Statistical Assoc. 2010;105(490):493-505.
262. Abadie A, Gardeazabal J. The Economic Costs of Conflict: A Case Study of the Basque Country. J American Economic Review. 2003;93(1):113-32.
263. Ewusie JE, Soobiah C, Blondal E, Beyene J, Thabane L, Hamid JS. Methods, Applications and Challenges in the Analysis of Interrupted Time Series Data: A Scoping Review. J Multidiscip Healthc. 2020;13:411-23.
264. Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. BMJ Glob Health. 2020;5(10).
265. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348-55.
266. Lechner M. The Estimation of Causal Effects by Difference-in-Difference Methods. Foundations and Trends(R) in Econometrics. 2011;4(3):165-224.
267. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract. 2000;17(suppl_1):S11-S6.
268. Victora CG, Habicht J-P, Bryce J. Evidence-Based Public Health: Moving Beyond Randomized Trials. 2004;94(3):400-5.
269. Lopez Bernal J, Cummins S, Gasparrini A. Difference in difference, controlled interrupted time series and synthetic controls. Int J Epidemiol. 2019;48(6):2062-3.
270. Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. J Epidemiol Community Health. 2011;65(7):582.
271. World Health Organization (WHO). WHO Guideline: Recommendations on digital interventions for health system strengthening. Geneva; 2019.
272. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and Undertaking a Health Economics Study of Digital Health Interventions. Am J Prev Med. 2016;51(5):852-60.
273. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. Implement Sci. 2017;12(1):61.
274. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: A systematic review. PLoS One. 2018;13(6):e0198112.
275. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev. 2018;6:CD012675.

276. Valenzuela Espinoza A, Steurbaut S, Dupont A, Cornu P, van Hooff RJ, Brouns R, et al. Health Economic Evaluations of Digital Health Interventions for Secondary Prevention in Stroke Patients: A Systematic Review. Cerebrovasc Dis Extra. 2019;9(1):1-8.
277. Grustam AS, Severens JL, van Nijnatten J, Koymans R, Vrijhoef HJ. Cost-effectiveness of telehealth interventions for chronic heart failure patients: a literature review. Int J Technol Assess Health Care. 2014;30(1):59-68.
278. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Value Health. 2013;16(2):e1-5.
279. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. PLoS One. 2013;8(7):e69930.
280. Bergmo TS. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. J Med Internet Res. 2015;17(11):e254-e.
281. Chapko MK, Liu CF, Perkins M, Li YF, Fortney JC, Maciejewski ML. Equivalence of two healthcare costing methods: bottom-up and top-down. Health Econ. 2009;18(10):1188-201.
282. Olsson TM. Comparing top-down and bottom-up costing approaches for economic evaluation within social welfare. Eur J Health Econ. 2011;12(5):445-53.
283. Smith P, Morrow R, Ross D, editors. Field Trials of Health Interventions: A Toolbox. 3rd ed. Oxford (UK)2015.
284. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med. 2009;38(1):40-5.
285. Curtis LA, Burns A. Unit Costs of Health and Social Care 2019. Kent, UK: Personal Social Services Research Unit (PSSRU); 2019.
286. Briggs A. Probabilistic Analysis of Cost-Effectiveness Models: Statistical Representation of Parameter Uncertainty. 2005;8(1):1-2.
287. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value Health. 2012;15(6):835-42.
288. Delignette-Muller ML, Pouillot Re, Denis J-B, Dutang C. Use of the package fitdistrplus to specify a distribution from non-censored or censored data. 2009.
289. Cohen DJ, Reynolds MR. Interpreting the results of cost-effectiveness studies. J Am Coll Cardiol. 2008;52(25):2119-26.
290. Risør BW, Lisby M, Sørensen J. Cost-Effectiveness Analysis of an Automated Medication System Implemented in a Danish Hospital Setting. Value Health. 2017;20(7):886-93.
291. University of Manchester. Ethics decision tool Manchester: University of Manchester; [Available from: http://www.training.itservices.manchester.ac.uk/uom/ERM/ethics_decision_tool/story_html5.html].
292. Medical Research Council (MRC) Regulatory Support Centre. Is my study research? 2017 [Available from: http://www.hra-decisiontools.org.uk/research/].
293. Campbell R. Practice Manager pay – Part 3: Should I have a pay rise? London: Practice Index Ltd; 2018 [Available from: https://practiceindex.co.uk/gp/blog/practice-manager-pay-part-3-pay-rise/].
294. Klok RM, Postma MJ. Four quadrants of the cost-effectiveness plane: some considerations on the south-west quadrant. Expert Rev Pharmacoecon Outcomes Res. 2014;4(6):599-601.
295. Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. 2015;32(7):507-8.
296. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. 2002;27(4):299-309.
297. Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. Stat Med. 2017;36(29):4660-76.
298. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care. 2003;19(4):613-23.

299. Abadie A. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. J Econ Lit. Forthcoming.
300. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. 2011;17(6):1231-8.
301. Franklin M, Wailoo A, Dayer MJ, Jones S, Prendergast B, Baddour LM, et al. The Cost-Effectiveness of Antibiotic Prophylaxis for Patients at Risk of Infective Endocarditis. Circulation. 2016;134(20):1568-78.
302. Nelson PA, Bradley F, Hogdgson D. Salford Primary Care Workforce Strategy: Contribution to a Safer Salford. University of Manchester 2018.
303. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care. 2008;17(3):216-23.
304. Hauck KD, Wang S, Vincent C, Smith PC. Healthy Life-Years Lost and Excess Bed-Days Due to 6 Patient Safety Incidents: Empirical Evidence From English Hospitals. Med Care. 2017;55(2):125-30.
305. Jorm L. Routinely collected data as a strategic resource for research: priorities for methods and workforce. Public Health Research & Practice. 2015;25(4).
306. Wongrakpanich S, Wongrakpanich A, Melhado K, Rangaswami J. A Comprehensive Review of Non-Steroidal Anti-Inflammatory Drug Use in The Elderly. Aging Dis. 2018;9(1):143-50.
307. Lamberts M, Lip GY, Hansen ML, Lindhardsen J, Olesen JB, Raunso J, et al. Relation of nonsteroidal anti-inflammatory drugs to serious bleeding and thromboembolism risk in patients with atrial fibrillation receiving antithrombotic therapy: a nationwide cohort study. Ann Intern Med. 2014;161(10):690-8.
308. Kent AP, Brueckmann M, Fraessdorf M, Connolly SJ, Yusuf S, Eikelboom JW, et al. Concomitant Oral Anticoagulant and Nonsteroidal Anti-Inflammatory Drug Therapy in Patients With Atrial Fibrillation. J Am Coll Cardiol. 2018;72(3):255-67.
309. Dalgaard F, Mulder H, Wojdyla DM, Lopes RD, Held C, Alexander JH, et al. Patients With Atrial Fibrillation Taking Nonsteroidal Anti-Inflammatory Drugs and Oral Anticoagulants in the ARISTOTLE Trial. Circulation. 2020;141(1):10-20.
310. Schjerning Olsen AM, McGettigan P, Gerds TA, Fosbol EL, Olesen JB, Sindet-Pedersen C, et al. Risk of gastrointestinal bleeding associated with oral anticoagulation and non-steroidal anti-inflammatory drugs in patients with atrial fibrillation: a nationwide study. Eur Heart J Cardiovasc Pharmacother. 2019.
311. Bayoumi I, Dolovich L, Hutchison B, Holbrook A. Medication-related emergency department visits and hospitalizations among older adults. Can Fam Physician. 2014;60(4):e217-22.
312. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National Surveillance of Emergency Department Visits for Outpatient Adverse Drug Events. JAMA. 2006;296(15):1858-66.
313. Warlé-van Herwaarden MF, Kramers C, Sturkenboom MC, van den Bemt PM, De Smet PA. Targeting outpatient drug safety: recommendations of the Dutch HARM-Wrestling Task Force. Drug Saf. 2012;35(3):245-59.
314. Wester K, Jönsson AK, Spigset O, Druid H, Hägg S. Incidence of fatal adverse drug reactions: a population based study. Br J Clin Pharmacol. 2008;65(4):573-9.
315. Ho KH, van Hove M, Leng G. Trends in anticoagulant prescribing: a review of local policies in English primary care. BMC Health Serv Res. 2020;20(1):279.
316. Johnson DA. Upper GI risks of NSAIDs and antiplatelet agents: key issues for the cardiologist. Rev Cardiovasc Med. 2005;6 Suppl 4:S15-22.
317. Lanas A. Gastrointestinal injury from NSAID therapy. How to reduce the risk of complications. Postgrad Med. 2005;117(6):23-8, 31.
318. de Abajo FJ, Gil MJ, Bryant V, Timoner J, Oliva B, García-Rodríguez LA. Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case–control study in a new general practice database. Eur J Clin Pharmacol. 2013;69(3):691-701.

319. Moore RA, Derry S, Phillips CJ, McQuay HJ. Nonsteroidal anti-inflammatory drugs (NSAIDs), cyxlooxygenase-2 selective inhibitors (coxibs) and gastrointestinal harm: review of clinical trials and clinical practice. BMC Musculoskelet Disord. 2006;7:79.
320. Lanas A, Sopena F. Nonsteroidal anti-inflammatory drugs and lower gastrointestinal complications. Gastroenterol Clin North Am. 2009;38(2):333-52.
321. Sikiric P, Seiwerth S, Rucman R, Turkovic B, Rokotov DS, Brcic L, et al. Toxicity by NSAIDs. Counteraction by stable gastric pentadecapeptide BPC 157. Curr Pharm Des. 2013;19(1):76-83.
322. Sostres C, Gargallo CJ, Lanas A. Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. Arthritis Res Ther. 2013;15 Suppl 3(Suppl 3):S3-S.
323. Goldstein JL, Cryer B. Gastrointestinal injury associated with NSAID use: a case study and review of risk factors and preventative strategies. Drug Healthc Patient Saf. 2015;7:31-41.
324. Bjarnason I, Scarpignato C, Holmgren E, Olszewski M, Rainsford KD, Lanas A. Mechanisms of Damage to the Gastrointestinal Tract From Nonsteroidal Anti-Inflammatory Drugs. Gastroenterology. 2018;154(3):500-14.
325. Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med. 2011;365(11):981-92.
326. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. N Engl J Med. 2009;361(12):1139-51.
327. Giercksky KE, Husby G, Rugstad HE, Revhaug A. Epidemiology of NSAID-induced gastrointestinal problems and the role of cimetidine in their prevention. Aliment Pharmacol Ther. 1988;2 Suppl 1:33-41.
328. Hearnshaw SA, Logan RF, Lowe D, Travis SP, Murphy MF, Palmer KR. Acute upper gastrointestinal bleeding in the UK: patient characteristics, diagnoses and outcomes in the 2007 UK audit. Gut. 2011;60(10):1327-35.
329. Oakland K. Changing epidemiology and etiology of upper and lower gastrointestinal bleeding. Best Pract Res Clin Gastroenterol. 2019;42-43:101610.
330. Staerk L, Lip GYH, Olesen JB, Fosbøl EL, Pallisgaard JL, Bonde AN, et al. Stroke and recurrent haemorrhage associated with antithrombotic treatment after gastrointestinal bleeding in patients with atrial fibrillation: nationwide cohort study. BMJ. 2015;351.
331. Battistella M, Mamdami MM, Juurlink DN, Rabeneck L, Laupacis A. Risk of upper gastrointestinal hemorrhage in warfarin users treated with nonselective NSAIDs or COX-2 inhibitors. Arch Intern Med. 2005;165(2):189-92.
332. Schjerning Olsen AM, Gislason GH, McGettigan P, Fosbol E, Sorensen R, Hansen ML, et al. Association of NSAID use with risk of bleeding and cardiovascular events in patients receiving antithrombotic therapy after myocardial infarction. JAMA. 2015;313(8):805-14.
333. Hansen ML, Sørensen R, Clausen MT, et al. Risk of bleeding with single, dual, or triple therapy with warfarin, aspirin, and clopidogrel in patients with atrial fibrillation. Arch Intern Med. 2010;170(16):1433-41.
334. Loo SY, Dell'Aniello S, Huiart L, Renoux C. Trends in the prescription of novel oral anticoagulants in UK primary care. Br J Clin Pharmacol. 2017;83(9):2096-106.
335. Davidson BL, Verheijen S, Lensing AA, et al. Bleeding risk of patients with acute venous thromboembolism taking nonsteroidal anti-inflammatory drugs or aspirin. JAMA Internal Medicine. 2014;174(6):947-53.
336. NHS Business Services Authority. NHS Medication Safety Dashboard 2018 [Available from: https://www.nhsbsa.nhs.uk/epact2/dashboards-and-specifications/medication-safety].
337. Varga Z, Sabzwari SRA, Vargova V. Cardiovascular Risk of Nonsteroidal Anti-Inflammatory Drugs: An Under-Recognized Public Health Issue. Cureus. 2017;9(4):e1144.
338. Sakuma I, Uchiyama S, Atarashi H, Inoue H, Kitazono T, Yamashita T, et al. Clinical risk factors of stroke and major bleeding in patients with non-valvular atrial fibrillation under rivaroxaban: the EXPAND Study sub-analysis. Heart Vessels. 2019;34(11):1839-51.
339. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ. 2018;363:k3532.

340. Patorno E, Schneeweiss S, Wang S. Transparency in real-world evidence (RWE) studies to build confidence for decision-making: Reporting RWE research in diabetes. Diabetes, Obesity and Metabolism. 2020;22:45-59.
341. Clinical Practice Research Datalink. Clinical Practice Research Datalink website [Available from: https://www.cprd.com/home/].
342. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.
343. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. Journal of Public Health. 2014;36(4):684-92.
344. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract. 2010;60(572):e128-36.
345. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010;69(1):4-14.
346. Office for National Statistics. Office for National Statistics website [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths].
347. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015;11(10):e1004525-e.
348. Ray WA. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. Am J Epidemiol. 2003;158(9):915-20.
349. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. Nature reviews Rheumatology. 2015;11(7):437-41.
350. Zalesak M, Siu K, Francis K, Yu C, Alvrtsyan H, Rao Y, et al. Higher persistence in newly diagnosed nonvalvular atrial fibrillation patients treated with dabigatran versus warfarin. Circ Cardiovasc Qual Outcomes. 2013;6(5):567-74.
351. Sorensen SV, Kansal AR, Connolly S, Peng S, Linnehan J, Bradley-Kennedy C, et al. Cost-effectiveness of dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation: a Canadian payer perspective. Thromb Haemost. 2011;105(5):908-19.
352. Gopi B, Sushmitha C, Nikitha Ksv, Monika M , Raghavendra Kumar G, Satanarayana V, et al. Cost-Effectiveness Analysis in the Management of Stroke. Asian Journal of Pharmaceutical and Clinical Research. 2017;10(7):127.
353. Wells G, Coyle D, Cameron C, Steiner S, Coyle K, Kelly S, et al. Safety, Effectiveness, and Cost-Effectiveness of New Oral Anticoagulants Compared with Warfarin in Preventing Stroke and Other Cardiovascular Events in Patients with Atrial Fibrillation. Ottawa (ON); 2012.
354. Vinogradova Y, Coupland C, Hill T, Hippisley-Cox J. Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care. BMJ. 2018:k2505.
355. 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.
356. Gallagher AM, van Staa TP, Murray-Thomas T, Schoof N, Clemens A, Ackermann D, et al. Population-based cohort study of warfarin-treated patients with atrial fibrillation: incidence of cardiovascular and bleeding outcomes. BMJ Open. 2014;4(1):e003839.
357. Spencer A. Hospital Episode Statistics (HES): Improving the quality and value of hospital data. 2011.
358. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Int J Epidemiol. 2017;46(4):1093-i.
359. Gallagher AM, Dedman D, Padmanabhan S, Leufkens HGM, de Vries F. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. Pharmacoepidemiol Drug Saf. 2019;28(5):563-9.
360. Salvo F, Fourrier-Réglat A, Bazin F, Robinson P, Riera-Guardia N, Haag M, et al. Cardiovascular and Gastrointestinal Safety of NSAIDs: A Systematic Review of Meta-Analyses of Randomized Clinical Trials. Clin Pharmacol Ther. 2011;89(6):855-66.

361. National Institute for Health and Care Excellence. Prescription writing [Available from: https://bnf.nice.org.uk/guidance/prescription-writing.html].
362. Kontopantelis E. PCDSEARCH: Stata module to extract code lists from primary care databases. Statistical Software Components: Boston College Department of Economics; 2015.
363. Springate DA, Kontopantelis E, Ashcroft DM, Olier I, Parisi R, Chamapiwa E, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. PLoS One. 2014;9(6):e99825.
364. van den Ham HA, Klungel OH, Singer DE, Leufkens HGM, van Staa TP. Comparative Performance of ATRIA, CHADS2, and CHA2DS2-VASc Risk Scores Predicting Stroke in Patients With Atrial Fibrillation: Results From a National Primary Care Database. J Am Coll Cardiol. 2015;66(17):1851-9.
365. Pye SR, Sheppard T, Joseph RM, Lunt M, Girard N, Haas JS, et al. Assumptions made when preparing drug exposure data for analysis have an impact on results: An unreported step in pharmacoepidemiology studies. Pharmacoepidemiol Drug Saf. 2018;27(7):781-8.
366. Joint Formulary Committee. British National Formulary (online) London [Available from: http://www.medicinescomplete.com].
367. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol. 2008;8:70.
368. Elwert F, Winship C. Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. Annu Rev Sociol. 2014;40:31-53.
369. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019;34(3):211-9.
370. Luque-Fernandez MA, Schomaker M, Redondo-Sanchez D, Jose Sanchez Perez M, Vaidya A, Schnitzer ME. Educational Note: Paradoxical collider effect in the analysis of non-communicable disease epidemiological data: a reproducible illustration and web application. Int J Epidemiol. 2019;48(2):640-53.
371. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest. 2010;137(2):263-72.
372. Olesen JB, Lip GYH, Hansen ML, Hansen PR, Tolstrup JS, Lindhardsen J, et al. Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. BMJ. 2011;342.
373. Saliba W, Gronich N, Barnett-Griness O, Rennert G. The role of CHADS2 and CHA2 DS2 -VASc scores in the prediction of stroke in individuals without atrial fibrillation: a population-based study. J Thromb Haemost. 2016;14(6):1155-62.
374. Xing Y, Sun Y, Li H, Tang M, Huang W, Zhang K, et al. CHA2DS2-VASc score as a predictor of long-term cardiac outcomes in elderly patients with or without atrial fibrillation. Clin Interv Aging. 2018;13:497-504.
375. Mitchell LB, Southern DA, Galbraith D, Ghali WA, Knudtson M, Wilton SB, et al. Prediction of stroke or TIA in patients without atrial fibrillation using CHADS2 and CHA2DS2-VASc scores. Heart. 2014;100(19):1524-30.
376. Hudzik B, Szkodzinski J, Hawranek M, Lekston A, Polonski L, Gasior M. CHA2DS2-VASc score is useful in predicting poor 12-month outcomes following myocardial infarction in diabetic patients without atrial fibrillation. Acta Diabetol. 2016;53(5):807-15.
377. Melgaard L, Gorst-Rasmussen A, Lane DA, Rasmussen LH, Larsen TB, Lip GY. Assessment of the CHA2DS2-VASc Score in Predicting Ischemic Stroke, Thromboembolism, and Death in Patients With Heart Failure With and Without Atrial Fibrillation. JAMA. 2015;314(10):1030-8.
378. Wu JT, Wang SL, Chu YJ, Long DY, Dong JZ, Fan XW, et al. CHADS2 and CHA2DS2-VASc Scores Predict the Risk of Ischemic Stroke Outcome in Patients with Interatrial Block without Atrial Fibrillation. J Atheroscler Thromb. 2017;24(2):176-84.
379. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest. 2010;138(5):1093-100.

380. Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N, et al. A new risk scheme to predict warfarin-associated hemorrhage: The ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) Study. J Am Coll Cardiol. 2011;58(4):395-401.
381. O'Brien EC, Simon DN, Thomas LE, Hylek EM, Gersh BJ, Ansell JE, et al. The ORBIT bleeding score: a simple bedside score to assess bleeding risk in atrial fibrillation. Eur Heart J. 2015;36(46):3258-64.
382. Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, et al. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). Am Heart J. 2006;151(3):713-9.
383. Apostolakis S, Lane DA, Guo Y, Buller H, Lip GYH. Performance of the HEMORR2HAGES, ATRIA, and HAS-BLED Bleeding Risk–Prediction Scores in Patients With Atrial Fibrillation Undergoing Anticoagulation. The AMADEUS (Evaluating the Use of SR34006 Compared to Warfarin or Acenocoumarol in Patients With Atrial Fibrillation) Study. 2012;60(9):861-7.
384. Roldan V, Marin F, Fernandez H, Manzano-Fernandez S, Gallego P, Valdes M, et al. Predictive value of the HAS-BLED and ATRIA bleeding scores for the risk of serious bleeding in a "real-world" population with atrial fibrillation receiving anticoagulant therapy. Chest. 2013;143(1):179-84.
385. Senoo K, Proietti M, Lane DA, Lip GY. Evaluation of the HAS-BLED, ATRIA, and ORBIT Bleeding Risk Scores in Patients with Atrial Fibrillation Taking Warfarin. Am J Med. 2016;129(6):600-7.
386. Zhu W, He W, Guo L, Wang X, Hong K. The HAS-BLED Score for Predicting Major Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation: A Systematic Review and Meta-analysis. Clin Cardiol. 2015;38(9):555-61.
387. Kooiman J, van Hagen N, Iglesias Del Sol A, Planken EV, Lip GY, van der Meer FJ, et al. The HAS-BLED Score Identifies Patients with Acute Venous Thromboembolism at High Risk of Major Bleeding Complications during the First Six Months of Anticoagulant Treatment. PLoS One. 2015;10(4):e0122520.
388. Smith JG, Wieloch M, Koul S, Braun O, Lumsden J, Rydell E, et al. Triple antithrombotic therapy following an acute coronary syndrome: prevalence, outcomes and prognostic utility of the HAS-BLED score. EuroIntervention. 2012;8(6):672-8.
389. Azoulay L, Dell'Aniello S, Simon T, Renoux C, Suissa S. The concurrent use of antithrombotic therapies and the risk of bleeding in patients with atrial fibrillation. Thromb Haemost. 2013;109(3):431-9.
390. Olesen JB, Lip GY, Lindhardsen J, Lane DA, Ahlehoff O, Hansen ML, et al. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: A net clinical benefit analysis using a 'real world' nationwide cohort study. Thromb Haemost. 2011;106(4):739-49.
391. Peacock WF, Tamayo S, Patel M, Sicignano N, Hopf KP, Yuan Z. CHA2DS2-VASc Scores and Major Bleeding in Patients With Nonvalvular Atrial Fibrillation Who Are Receiving Rivaroxaban. Ann Emerg Med. 2017;69(5):541-50 e1.
392. Alikhan R, Lefevre C, Menown I, Lister S, Bird A, You M, et al. Risk of Recurrent Bleeding Events in Nonvalvular Atrial Fibrillation Treated with Vitamin K Antagonists: A Clinical Practice Research Datalink Study. TH Open. 2019;3(4):e316-e24.
393. Lacoin L, Lumley M, Ridha E, Pereira M, McDonald L, Ramagopalan S, et al. Evolving landscape of stroke prevention in atrial fibrillation within the UK between 2012 and 2016: a cross-sectional analysis study using CPRD. BMJ open. 2017;7(9):e015363-e.
394. Tennant PWG, Harrison WJ, Murray EJ, Arnold KF, Berrie L, Fox MP, et al. Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. medRxiv. 2019:2019.12.20.19015511.
395. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48.
396. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016;45(6):1887-94.

397. Zanutto EL. A comparison of propensity score and linear regression analysis of complex survey data. Journal of data Science. 2006;4(1):67-91.
398. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services and Outcomes Research Methodology. 2001;2(3):169-88.
399. Cochran WG, Rubin DB. Controlling Bias in Observational Studies: A Review. The Indian Journal of Statistics. 1973;35(4):417-46.
400. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.
401. Austin PC. The Relative Ability of Different Propensity Score Methods to Balance Measured Covariates Between Treated and Untreated Subjects in Observational Studies. Med Decis Making. 2009;29(6):661-77.
402. Bergstra SA, Sepriano A, Ramiro S, Landewe R. Three handy tips and a practical guide to improve your propensity score models. RMD Open. 2019;5(1):e000953.
403. Wyss R, Girman CJ, LoCasale RJ, Brookhart AM, Sturmer T. Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. Pharmacoepidemiol Drug Saf. 2013;22(1):77-85.
404. Zhang Z, Kim HJ, Lonjon G, Zhu Y, written on behalf of AMEB-DCTCG. Balance diagnostics after propensity score matching. Annals of translational medicine. 2019;7(1):16-.
405. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. J Roy Stat Soc Ser A (Stat Soc). 2008;171(2):481-502.
406. Kirchhof P, Haas S, Amarenco P, Hess S, Lambelet M, van Eickels M, et al. Impact of Modifiable Bleeding Risk Factors on Major Bleeding in Patients With Atrial Fibrillation Anticoagulated With Rivaroxaban. Journal of the American Heart Association. 2020;9(5):e009530.
407. Rydberg DM, Linder M, Malmstrom RE, Andersen M. Risk factors for severe bleeding events during warfarin treatment: the influence of sex, age, comorbidity and co-medication. Eur J Clin Pharmacol. 2020;76(6):867-76.
408. Albertsen IE, Rasmussen LH, Overvad TF, Graungaard T, Larsen TB, Lip GY. Risk of stroke or systemic embolism in atrial fibrillation patients treated with warfarin: a systematic review and meta-analysis. Stroke. 2013;44(5):1329-36.
409. Kjerpeseth LJ, Ellekjær H, Selmer R, Ariansen I, Furu K, Skovlund E. Risk factors for stroke and choice of oral anticoagulant in atrial fibrillation. Eur J Clin Pharmacol. 2018;74(12):1653-62.
410. Paciaroni M, Agnelli G, Caso V, Silvestrelli G, Seiffge DJ, Engelter S, et al. Causes and Risk Factors of Cerebral Ischemic Events in Patients With Atrial Fibrillation Treated With Non-Vitamin K Antagonist Oral Anticoagulants for Stroke Prevention. 2019;50(8):2168-74.
411. Leung K-M, Elashoff RM, Afifi AA. Censoring issues in survival analysis. 1997;18(1):83-104.
412. Loo SY, Coulombe J, Dell'Aniello S, Brophy JM, Suissa S, Renoux C. Comparative effectiveness of novel oral anticoagulants in UK patients with non-valvular atrial fibrillation and chronic kidney disease: a matched cohort study. BMJ Open. 2018;8(1).
413. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. Ann Transl Med. 2018;6(7):121.
414. Austin PC, Latouche A, Fine JP. A review of the use of time-varying covariates in the Fine-Gray subdistribution hazard competing risk regression model. Stat Med. 2020;39(2):103-13.
415. Poguntke I, Schumacher M, Beyersmann J, Wolkewitz M. Simulation shows undesirable results for competing risks analysis with time-dependent covariates for clinical outcomes. BMC Med Res Methodol. 2018;18(1):79.
416. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016;133(6):601-9.
417. Cleves M, Gould W, Gutierrez R, Marchenko Y. An Introduction to Survival Analysis Using Stata: Stata Press; 2010.
418. Cortese G, Andersen PK. Competing risks and time-dependent covariates. Biom J. 2010;52(1):138-58.
419. Lesko CR, Lau B. Bias Due to Confounders for the Exposure-Competing Risk Relationship. Epidemiology. 2017;28(1):20-7.

420. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology (Cambridge, Mass). 2009;20(4):512-22.
421. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals. Value Health. 2010;13(2):273-7.
422. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med. 2016;35(30):5642-55.
423. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. JAMA. 2019;321(6):602-3.
424. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-74.
425. Blum MR, Tan YJ, Ioannidis JPA. Use of E-values for addressing confounding in observational studies-an empirical assessment of the literature. Int J Epidemiol. 2020;49(5):1482-94.
426. Sebag-Montefiore D. ARISTOTLE - A phase Ill trial comparing standard versus novel CRT as pre-operative treatment for MRI defined locally advanced rectal cancer - Protocol. London: University College London; 2015.
427. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician. Adv Ther. 2018;35(11):1763-74.
428. Fewell Z, Davey Smith G, Sterne JAC. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. Am J Epidemiol. 2007;166(6):646-55.
429. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. The Lancet. 2010;376(9735):112-23.
430. Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. BMC Health Serv Res. 2012;12:392.
431. McDonald L, Sammon CJ, Samnaliev M, Ramagopalan S. Under-recording of hospital bleeding events in UK primary care: a linked Clinical Practice Research Datalink and Hospital Episode Statistics study. Clin Epidemiol. 2018;10:1155-68.
432. Shehab N, Ziemba R, Campbell KN, Geller AI, Moro RN, Gage BF, et al. Assessment of ICD-10-CM code assignment validity for case finding of outpatient anticoagulant-related bleeding among Medicare beneficiaries. Pharmacoepidemiol Drug Saf. 2019;28(7):951-64.
433. Kivimäki M, Batty GD, Singh-Manoux A, Britton A, Brunner EJ, Shipley MJ. Validity of Cardiovascular Disease Event Ascertainment Using Linkage to UK Hospital Records. Epidemiology (Cambridge, Mass). 2017;28(5):735-9.
434. Mansfield K, Crellin E, Denholm R, Quint JK, Smeeth L, Cook S, et al. Completeness and validity of alcohol recording in general practice within the UK: a cross-sectional study. BMJ Open. 2019;9(11):e031537.
435. Ali MS, Prieto-Alhambra D, Lopes LC, Ramos D, Bispo N, Ichihara MY, et al. Propensity Score Methods in Health Technology Assessment: Principles, Extended Applications, and Recent Advances. Front Pharmacol. 2019;10:973-.
436. Johnson ES, Bartman BA, Briesacher BA, Fleming NS, Gerhard T, Kornegay CJ, et al. The incident user design in comparative effectiveness research. Pharmacoepidemiol Drug Saf. 2013;22(1):1-6.
437. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value Health. 2011;14(4):539-45.
438. Curtis LA, Burns A. Unit Cost of Health and Social Care 2017. University of Kent; 2017. Report No.: https://doi.org/10.22024/UniKent/01.02/65559.
439. Department of Health and Social Care. NHS Reference Costs 2015 to 2016. London: Department of Health; 2016.

440. Saramago P, Manca A, Sutton AJ. Deriving Input Parameters for Cost-Effectiveness Modeling: Taxonomy of Data Types and Approaches to Their Statistical Synthesis. Value Health. 2012;15(5):639-49.
441. de Rezende BA, Or Z, Com-Ruelle L, Michel P. Economic evaluation in patient safety: a literature review of methods. BMJ Qual Saf. 2012;21(6):457-65.
442. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. Pharmacoeconomics. 2016;34(4):349-61.
443. NHS Specialist Pharmacy Service. Community Pharmacy NSAID Safety Audit 2018-19. 2019.
444. Agah S, Akbari A, Heshmati J, Sepidarkish M, Morvaridzadeh M, Adibi P, et al. Systematic review with meta-analysis: Effects of probiotic supplementation on symptoms in functional dyspepsia. J Funct Foods. 2020;68:103902.
445. Tian W, Jiang W, Yao J, Nicholson CJ, Li RH, Sigurslid HH, et al. Predictors of mortality in hospitalized COVID-19 patients: A systematic review and meta-analysis. J Med Virol. 2020.
446. Chau SH, Sluiter RL, Kievit W, Wensing M, Teichert M, Hugtenburg JG. Cost Effectiveness of Gastroprotection with Proton Pump Inhibitors in Older Low-Dose Acetylsalicylic Acid Users in the Netherlands. Drugs Aging. 2017;34(5):375-86.
447. Chau SH, Sluiter RL, Hugtenburg JG, Wensing M, Kievit W, Teichert M. Cost–Utility and Budget Impact Analysis for Stopping the Inappropriate Use of Proton Pump Inhibitors After Cessation of NSAID or Low-Dose Acetylsalicylic Acid Treatment. Drugs Aging. 2019:1-8.
448. Saini S, Fendrick A, Scheiman J. Cost-effectiveness analysis: cardiovascular benefits of proton pump inhibitor co-therapy in patients using aspirin for secondary prevention. Aliment Pharmacol Ther. 2011;34(2):243-51.
449. Takabayashi N, Murata K, Tanaka S, Kawakami K. Cost-effectiveness of proton pump inhibitor co-therapy in patients taking aspirin for secondary prevention of ischemic stroke. Pharmacoeconomics. 2015;33(10):1091-100.
450. Latimer N, Lord J, Grant RL, O'Mahony R, Dickson J, Conaghan PG. Cost effectiveness of COX 2 selective inhibitors and traditional NSAIDs alone or in combination with a proton pump inhibitor for people with osteoarthritis. BMJ. 2009;339:b2538.
451. Latimer N, Lord J, Grant RL, O'Mahony R, Dickson J, Conaghan PG. Value of information in the osteoarthritis setting: cost effectiveness of COX-2 selective inhibitors, traditional NSAIDs and proton pump inhibitors. Pharmacoeconomics. 2011;29(3):225-37.
452. de Groot NL, Spiegel BM, van Haalen HG, de Wit NJ, Siersema PD, van Oijen MG. Gastroprotective strategies in chronic NSAID users: a cost-effectiveness analysis comparing single-tablet formulations with individual components. Value Health. 2013;16(5):769-77.
453. de Groot NL, van Haalen HG, Spiegel BM, Laine L, Lanas A, Focks JJ, et al. Gastroprotection in low-dose aspirin users for primary and secondary prevention of ACS: results of a cost-effectiveness analysis including compliance. Cardiovasc Drugs Ther. 2013;27(4):341-57.
454. Wielage R, Bansal M, Wilson K, Klein R, Happich M. Cost-effectiveness of duloxetine in chronic low back pain: a Quebec societal perspective. Spine (Phila Pa 1976). 2013;38(11):936-46.
455. Wielage RC, Myers JA, Klein RW, Happich M. Cost-effectiveness analyses of osteoarthritis oral therapies: A systematic review. Applied Health Economics and Health Policy. 2013;11(6):593-618.
456. Brereton N, Pennington B, Ekelund M, Akehurst R. A cost-effectiveness analysis of celecoxib compared with diclofenac in the treatment of pain in osteoarthritis (OA) within the Swedish health system using an adaptation of the NICE OA model. J Med Econ. 2014;17(9):677-84.
457. Brereton N, Winn B, Akehurst R. The cost-effectiveness of celecoxib vs diclofenac in the treatment of osteoarthritis in the UK; an update to the NICE model using data from the CONDOR trial. J Med Econ. 2012;15(3):465-72.
458. Nasef SA, Shaaban AA, Mould-Quevedo J, Ismail TA. The cost-effectiveness of celecoxib versus non-steroidal anti-inflammatory drugs plus proton-pump inhibitors in the treatment of osteoarthritis in Saudi Arabia. Health Econ Rev. 2015;5(1):53.

459. Chung S-J, Park H-J, Park M-C. Cost-effectiveness of Non-steroidal Anti-inflammatory Drugs Adjusting for Upper and Lower Gastrointestinal Toxicities in Rheumatoid Arthritis Patients. Journal of Rheumatic Diseases. 2017;24(1).
460. Lee SP, Sung IK, Kim JH, Lee SY, Park HS, Shim CS. Risk Factors for the Presence of Symptoms in Peptic Ulcer Disease. Clin Endosc. 2017;50(6):578-84.
461. Wielage RC, Bansal M, Andrews JS, Klein RW, Happich M. Cost-utility analysis of duloxetine in osteoarthritis: a US private payer perspective. Appl Health Econ Health Policy. 2013;11(3):219-36.
462. Sorensen SV, Dewilde S, Singer DE, Goldhaber SZ, Monz BU, Plumb JM. Cost-effectiveness of warfarin: trial versus "real-world" stroke prevention in atrial fibrillation. Am Heart J. 2009;157(6):1064-73.
463. Spackman E, Burch J, Faria R, Corbacho B, Fox D, Woolacott N. Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation (TA 249). 2011.
464. González-Juanatey JR, Álvarez-Sabin J, Lobos JM, Martínez-Rubio A, Reverter JC, Oyagüez I, et al. Cost-effectiveness of Dabigatran for Stroke Prevention in Non-valvular Atrial Fibrillation in Spain. Revista Española de Cardiología (English Edition). 2012;65(10):901-10.
465. Kansal AR, Sharma M, Bradley-Kennedy C, Clemens A, Monz BU, Peng S, et al. Dabigatran versus rivaroxaban for the prevention of stroke and systemic embolism in atrial fibrillation in Canada. Comparative efficacy and cost-effectiveness. Thromb Haemost. 2012;108(4):672-82.
466. Kansal AR, Sorensen SV, Gani R, Robinson P, Pan F, Plum JM, et al. Cost-effectiveness of dabigatran etexilate for the prevention of stroke and systemic embolism in UK patients with atrial fibrillation. Heart. 2012;98(7):573-8.
467. Janzic A, Kos M. Cost effectiveness of novel oral anticoagulants for stroke prevention in atrial fibrillation depending on the quality of warfarin anticoagulation control. Pharmacoeconomics. 2015;33(4):395-408.
468. Langkilde LK, Bergholdt Asmussen M, Overgaard M. Cost-effectiveness of dabigatran etexilate for stroke prevention in non-valvular atrial fibrillation. Applying RE-LY to clinical practice in Denmark. J Med Econ. 2012;15(4):695-703.
469. Zheng Y, Sorensen SV, Gonschior AK, Noack H, Heinrich-Nols J, Sunderland T, et al. Comparison of the cost-effectiveness of new oral anticoagulants for the prevention of stroke and systemic embolism in atrial fibrillation in a UK setting. Clin Ther. 2014;36(12):2015-28.e2.
470. Andrikopoulos GK, Fragoulakis V, Maniadakis N. Economic evaluation of dabigatran etexilate in the management of atrial fibrillation in Greece. Hellenic J Cardiol. 2013;54(4):289-300.
471. Bergh M, Marais CA, Miller-Janson H, Salie F, Stander MP. Economic appraisal of dabigatran as first-line therapy for stroke prevention in atrial fibrillation. S Afr Med J. 2013;103(4):241-5.
472. Dorian P, Kongnakorn T, Phatak H, Rublee DA, Kuznik A, Lanitis T, et al. Cost-effectiveness of apixaban vs. current standard of care for stroke prevention in patients with atrial fibrillation. Eur Heart J. 2014;35(28):1897-906.
473. Giorgi MA, Caroli C, Giglio ND, Micone P, Aiello E, Vulcano C, et al. Estimation of the cost-effectiveness of apixaban versus vitamin K antagonists in the management of atrial fibrillation in Argentina. Health Economics Review. 2015;5 (1)(17).
474. Athanasakis K, Boubouchairopoulou N, Karampli E, Tarantilis F, Savvari P, Bilitou A, et al. Cost Effectiveness of Apixaban versus Warfarin or Aspirin for Stroke Prevention in Patients with Atrial Fibrillation: A Greek Perspective. Am J Cardiovasc Drugs. 2017;17(2):123-33.
475. Baron Esquivias G, Escolar Albaladejo G, Zamorano JL, Betegon Nicolas L, Canal Fontcuberta C, de Salas-Cansado M, et al. Cost-effectiveness Analysis Comparing Apixaban and Acenocoumarol in the Prevention of Stroke in Patients With Nonvalvular Atrial Fibrillation in Spain. Rev Esp Cardiol (Engl Ed). 2015;68(8):680-90.
476. Kongnakorn T, Lanitis T, Lieven A, Thijs V, Marbaix S. Cost effectiveness of apixaban versus aspirin for stroke prevention in patients with non-valvular atrial fibrillation in Belgium. Clin Drug Investig. 2014;34(10):709-21.

477. Lip GY, Kongnakorn T, Phatak H, Kuznik A, Lanitis T, Liu LZ, et al. Cost-effectiveness of apixaban versus other new oral anticoagulants for stroke prevention in atrial fibrillation. Clin Ther. 2014;36(2):192-210.e20.
478. Pradelli L, Calandriello M, Di Virgilio R, Bellone M, Tubaro M. Comparative pharmacoeconomic assessment of apixaban vs. standard of care for the prevention of stroke in Italian atrial fibrillation patients. Farmeconomia Health economics and therapeutic pathways. 2014;15(1S):15.
479. Lanitis T, Kongnakorn T, Jacobson L, De Geer A. Cost-effectiveness of apixaban versus warfarin and aspirin in Sweden for stroke prevention in patients with atrial fibrillation. Thromb Res. 2014;134(2):278-87.
480. Hallinen T, Soini EJ, Linna M, Saarni SI. Cost-effectiveness of apixaban and warfarin in the prevention of thromboembolic complications among atrial fibrillation patients. Springerplus. 2016;5(1):1354.
481. Gage BF, Cardinalli AB, Albers GW, Owens DK. Cost-effectiveness of warfarin and aspirin for prophylaxis of stroke in patients with nonvalvular atrial fibrillation. JAMA. 1995;274(23):1839-45.
482. Lee S, Anglade MW, Meng J, Hagstrom K, Kluger J, Coleman CI. Cost-effectiveness of apixaban compared with aspirin for stroke prevention in atrial fibrillation among patients unsuitable for warfarin. Circ Cardiovasc Qual Outcomes. 2012;5(4):472-9.
483. Lee S, Anglade MW, Pham D, Pisacane R, Kluger J, Coleman CI. Cost-effectiveness of rivaroxaban compared to warfarin for stroke prevention in atrial fibrillation. Am J Cardiol. 2012;110(6):845-51.
484. Lee S, Mullin R, Blazawski J, Coleman CI. Cost-effectiveness of apixaban compared with warfarin for stroke prevention in atrial fibrillation. PLoS One. 2012;7(10):e47473.
485. Lopez-Lopez JA, Sterne JAC, Thom HHZ, Higgins JPT, Hingorani AD, Okoli GN, et al. Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. BMJ. 2017;359:j5058.
486. Singh SM, Micieli A, Wijeysundera HC. Economic evaluation of percutaneous left atrial appendage occlusion, dabigatran, and warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. Circulation. 2013;127(24):2414-23.
487. Ademi Z, Pasupathi K, Liew D. Cost-effectiveness of apixaban compared to warfarin in the management of atrial fibrillation in Australia. European Journal of Preventive Cardiology. 2015;22(3):344-53.
488. Nshimyumukiza L, Duplantie J, Gagnon M, Douville X, Fournier D, Lindsay C, et al. Dabigatran versus warfarin under standard or pharmacogenetic-guided management for the prevention of stroke and systemic thromboembolism in patients with atrial fibrillation: A cost/utility analysis using an analytic decision model. Thrombosis Journal. 2013;11 (1) (no pagination)(14).
489. Ademi Z, Pasupathi K, Liew D. Clinical and Cost Effectiveness of Apixaban Compared to Aspirin in Patients with Atrial Fibrillation: An Australian Perspective. Applied Health Economics and Health Policy. 2016:1-12.
490. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. Value Health. 2012;15(6):812-20.
491. Edwards S, Hamilton V, Trevor N, Nherera L, Karnar C, Thurgar E. Apixaban (Eliquis®)for the prevention of stroke and systemic embolism in people with non-valvular atrial fibrillation (TA 275). 2012.
492. National Collaborating Centre for Chronic Conditions. Osteoarthritis: National Clinical Guideline for Care and Management in Adults. UK; 2008.
493. Staerk L, Gislason GH, Lip GY, Fosbol EL, Hansen ML, Lamberts M, et al. Risk of gastrointestinal adverse effects of dabigatran compared with warfarin among patients with atrial fibrillation: a nationwide cohort study. Europace. 2015;17(8):1215-22.
494. Office for National Statistics. Deaths registered in England and Wales (2018). 2019.
495. Lip GHY, Kongnakorn T, Phatak H, Kuznik A, Rublee D, Lanitis T, et al. Cost-effectiveness of apixaban against other novel oral anticoagulants (NOACs) for stroke prevention in atrial fibrillation patients. Eur Heart J. 2012;1):54.

496. Rosenstock S, Jørgensen T, Bonnevie O, Andersen L. Risk factors for peptic ulcer disease: a population based prospective cohort study comprising 2416 Danish adults. Gut. 2003;52(2):186-93.
497. Cai S, Garcia Rodriguez LA, Masso-Gonzalez EL, Hernandez-Diaz S. Uncomplicated peptic ulcer in the UK: trends from 1997 to 2005. Aliment Pharmacol Ther. 2009;30(10):1039-48.
498. Lin KJ, Garcia Rodriguez LA, Hernandez-Diaz S. Systematic review of peptic ulcer disease incidence rates: do studies without validation provide reliable estimates? Pharmacoepidemiol Drug Saf. 2011;20(7):718-28.
499. Lassen A, Hallas J, de Muckadell OBS. Complicated and uncomplicated peptic ulcers in a Danish county 1993–2002: a population-based cohort study. Am J Gastroenterol. 2006;101(5):945-53.
500. Malmi H, Kautiainen H, Virta LJ, Farkkila MA. Increased short- and long-term mortality in 8146 hospitalised peptic ulcer patients. Aliment Pharmacol Ther. 2016;44(3):234-45.
501. Yoon H, Kim SG, Jung HC, Song IS. High Recurrence Rate of Idiopathic Peptic Ulcers in Long-Term Follow-up. Gut Liver. 2013;7(2):175-81.
502. Malmi H, Kautiainen H, Virta LJ, Farkkila N, Koskenpato J, Farkkila MA. Incidence and complications of peptic ulcer disease requiring hospitalisation have markedly decreased in Finland. Aliment Pharmacol Ther. 2014;39(5):496-506.
503. Malmi H, Färkkilä MA, Virta LJ. Peptic ulcer disease - incidence, associated morbidity and mortality. Helsinki: University of Helsinki; 2018.
504. Tapaskar N, Pang A, Werner DA, Sengupta N. Resuming Anticoagulation Following Hospitalization for Gastrointestinal Bleeding Is Associated with Reduced Thromboembolic Events and Improved Mortality: Results from a Systematic Review and Meta-Analysis. Dig Dis Sci. 2020.
505. Ariesen MJ, Claus SP, Rinkel GJ, Algra A. Risk factors for intracerebral hemorrhage in the general population: a systematic review. Stroke. 2003;34(8):2060-5.
506. Komen JJ, Forslund T, Mantel-Teeuwisse AK, Klungel OH, von Euler M, Braunschweig F, et al. Association of Preceding Antithrombotic Therapy in Atrial Fibrillation Patients With Ischemic Stroke, Intracranial Hemorrhage, or Gastrointestinal Bleed and Mortality. Eur Heart J Cardiovasc Pharmacother. 2019.
507. Lip GY, Clementy N, Pericart L, Banerjee A, Fauchier L. Stroke and major bleeding risk in elderly patients aged >/=75 years with atrial fibrillation: the Loire Valley atrial fibrillation project. Stroke. 2015;46(1):143-50.
508. van Walraven C, Hart RG, Connolly S, Austin PC, Mant J, Hobbs FD, et al. Effect of age on stroke prevention therapy in patients with atrial fibrillation: the atrial fibrillation investigators. Stroke. 2009;40(4):1410-6.
509. Hillen T, Coshall C, Tilling K, Rudd AG, McGovern R, Wolfe CDA. Cause of Stroke Recurrence Is Multifactorial. 2003;34(6):1457-63.
510. Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Burvill PW, Anderson CS, et al. Long-term risk of first recurrent stroke in the Perth Community Stroke Study. Stroke. 1998;29(12):2491-500.
511. Flach C, Muruet W, Wolfe CDA, Bhalla A, Douiri A. Risk and Secondary Prevention of Stroke Recurrence: A Population-Base Cohort Study. Stroke. 2020;51(8):2435-44.
512. Rasmussen LH, Larsen TB, Graungaard T, Skjøth F, Lip GYH. Primary and secondary prevention with new oral anticoagulant drugs for stroke prevention in atrial fibrillation: indirect comparison analysis. BMJ : British Medical Journal. 2012;345:e7097.
513. Hankey GJ, Patel MR, Stevens SR, Becker RC, Breithardt G, Carolei A, et al. Rivaroxaban compared with warfarin in patients with atrial fibrillation and previous stroke or transient ischaemic attack: a subgroup analysis of ROCKET AF. The Lancet Neurology. 2012;11(4):315-22.
514. Easton JD, Lopes RD, Bahit MC, Wojdyla DM, Granger CB, Wallentin L, et al. Apixaban compared with warfarin in patients with atrial fibrillation and previous stroke or transient ischaemic attack: a subgroup analysis of the ARISTOTLE trial. The Lancet Neurology. 2012;11(6):503-11.
515. Pamela J B, Joseph H, Matthew K, Thomas G B, Lee N, Judith M K, et al. Warfarin Use and Mortality, Stroke, and Bleeding Outcomes in a Cohort of Elderly Patients with non-Valvular Atrial Fibrillation. Journal of atrial fibrillation. 2019;12(1):2155-.

516. Gomez-Outes A, Lagunar-Ruiz J, Terleira-Fernandez AI, Calvo-Rojas G, Suarez-Gea ML, Vargas-Castrillon E. Causes of Death in Anticoagulated Patients With Atrial Fibrillation. J Am Coll Cardiol. 2016;68(23):2508-21.
517. Hippisley-Cox J, Coupland C, Logan R. Risk of adverse gastrointestinal outcomes in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ. 2005;331(7528):1310-6.
518. Gonzalez-Perez A, Saez ME, Johansson S, Nagy P, Garcia Rodriguez LA. Risk factors associated with uncomplicated peptic ulcer and changes in medication use after diagnosis. PLoS One. 2014;9(7):e101768.
519. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Inform Prim Care. 2011;19(4):251-5.
520. WHO Collaborating Centre for Drug Statistics Methodology. ATC/DD Index 2020: Norwegian Institute of Public Health; 2020 [Available from: https://www.whocc.no/atc_ddd_index/].
521. Elliott RA, Hooper L, Payne K, Brown TJ, Roberts C, Symmons D. Preventing non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: are older strategies more cost-effective in the general population? Rheumatology. 2006;45(5):606-13.
522. National Institute for Health and Care Excellence. NICE guideline 184-Gastro-oesophageal reflux disease and dyspepsia in adults: indyspepsia in adults: investigation and management. 2014.
523. Campbell HE, Stokes EA, Bargo D, Logan RF, Mora A, Hodge R, et al. Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial. BMJ Open. 2015;5(4):e007230.
524. Sterne JA, Bodalia PN, Bryden PA, Davies PA, Lopez-Lopez JA, Okoli GN, et al. Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis. Health Technol Assess. 2017;21(9):1-386.
525. Department of Health and Social Care. National Schedule of Reference Costs 2017-2018. London; 2018.
526. Ali AN, Abdelhafiz A. Clinical and Economic Implications of AF Related Stroke. Journal of atrial fibrillation. 2016;8(5):1279-.
527. Luengo-Fernandez R, Yiin GS, Gray AM, Rothwell PM. Population-based study of acute- and long-term care costs after stroke in patients with AF. Int J Stroke. 2013;8(5):308-14.
528. NHS Prescription Services. NHS Electronic Drug Tariff 2019 [Available from: http://www.drugtariff.nhsbsa.nhs.uk/].
529. EuroQol Group:EuroQol. A new facility for the measurement of health related quality of life. Health Policy. 1990;16:199-208.
530. van Reenen M, Janssen, B. . EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. 2015.
531. National Institute for Health and Care Excellence. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018) 2018 [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l].
532. Berg J, Lindgren P, Nieuwlaat R, Bouin O, Crijns H. Factors determining utility measured with the EQ-5D in patients with atrial fibrillation. Qual Life Res. 2010;19(3):381-90.
533. Thom HHZ, Hollingworth W, Sofat R, Wang Z, Fang W, Bodalia PN, et al. Directly Acting Oral Anticoagulants for the Prevention of Stroke in Atrial Fibrillation in England and Wales: Cost-Effectiveness Model and Value of Information Analysis. MDM policy & practice. 2019;4(2):2381468319866828-.
534. Edwards S, Hamilton V, Nherera L, Trevor N, Barton S. Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation (TA 256). 2011.
535. Delaney BC, Qume M, Moayyedi P, Logan RF, Ford AC, Elliott C, et al. Helicobacter pylori test and treat versus proton pump inhibitor in initial management of dyspepsia in primary care: multicentre randomised controlled trial (MRC-CUBE trial). BMJ. 2008;336(7645):651-4.

536. Tai M-LS, Norhatta N, Goh KJ, Moy FM, Sujarita R, Asraff AA, et al. The impact of dyspepsia on symptom severity and quality of life in adults with headache. PLoS One. 2015;10(1).
537. Groeneveld PW, Lieu TA, Fendrick AM, Hurley LB, Ackerson LM, Levin TR, et al. Quality of life measurement clarifies the cost-effectiveness of Helicobacter pylori eradication in peptic ulcer disease and uninvestigated dyspepsia. The American journal of gastroenterology. 2001;96(2):338-47.
538. Maetzel A, Krahn M, Naglie G. The cost effectiveness of rofecoxib and celecoxib in patients with osteoarthritis or rheumatoid arthritis. Arthritis Rheum. 2003;49(3):283-92.
539. Groeneveld PW, Lieu TA, Fendrick AM, Hurley LB, Ackerson LM, Levin TR, et al. Quality of life measurement clarifies the cost-effectiveness of Helicobacter pylori eradication in peptic ulcer disease and uninvestigated dyspepsia11This work was conducted at the Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA; University of California, San Francisco; and VA Medical Center, San Francisco, CA. The American Journal of Gastroenterology. 2001;96(2):338-47.
540. Doble B, Pufulete M, Harris JM, Johnson T, Lasserson D, Reeves BC, et al. Health-related quality of life impact of minor and major bleeding events during dual antiplatelet therapy: a systematic literature review and patient preference elicitation study. Health & Quality of Life Outcomes. 2018;16(1):191.
541. Campbell H, Stokes E, Bargo D, Logan R, Mora A, Hodge R, et al. Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial. BMJ open. 2015;5(4):e007230.
542. Wang K, Li H, Kwong WJ, Antman EM, Ruff CT, Giugliano RP, et al. Impact of Spontaneous Extracranial Bleeding Events on Health State Utility in Patients with Atrial Fibrillation: Results from the ENGAGE AF-TIMI 48 Trial. J Am Heart Assoc. 2017;6(8).
543. Spiegel BM, Chiou CF, Ofman JJ. Minimizing complications from nonsteroidal antiinflammatory drugs: cost-effectiveness of competing strategies in varying risk groups. Arthritis Rheum. 2005;53(2):185-97.
544. Luengo-Fernandez R, Gray AM, Bull L, Welch S, Cuthbertson F, Rothwell PM, et al. Quality of life after TIA and stroke: ten-year results of the Oxford Vascular Study. Neurology. 2013;81(18):1588-95.
545. Haacke C, Althaus A, Spottke A, Siebert U, Back T, Dodel R. Long-Term Outcome After Stroke. Stroke. 2006;37(1):193-8.
546. TreeAge Pro Healthcare 2021. [computer software]. Williamstown, MA; 2020.
547. Badeaux J, Bonanno L, Au H. Effectiveness of ondansetron as an adjunct to lidocaine intravenous regional anesthesia on tourniquet pain and postoperative pain in patients undergoing elective hand surgery: a systematic review protocol. JBI Database Of Systematic Reviews And Implementation Reports. 2015;13(1):27-38.
548. Straus WL, Ofman JJ, MacLean C, Morton S, Berger ML, Roth EA, et al. Do NSAIDs cause dyspepsia? A meta-analysis evaluating alternative dyspepsia definitions. Am J Gastroenterol. 2002;97(8):1951-8.
549. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation modeling for health decision sciences using R: A Tutorial. Med Decis Making. 2018;38(3):400-22.
550. Freeman M, Groom B, Spackman M. Social discount rates for cost–benefit analysis: a report for HM treasury 2018 [Available from: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent].
551. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18(2 Suppl):S68-80.
552. Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on Medical Decision Analysis: Part 4-Analyzing the Model and Interpreting the Results. Med Decis Making. 1997;17(2):142-51.
553. Schwappach DL. Risk factors for patient-reported medical errors in eleven countries. Health Expect. 2014;17(3):321-31.
554. Patel I, Balkrishnan R. Medication Error Management around the Globe: An Overview. Indian J Pharm Sci. 2010;72(5):539-45.

555. Price DB, Román-Rodríguez M, McQueen RB, Bosnic-Anticevich S, Carter V, Gruffydd-Jones K, et al. Inhaler Errors in the CRITIKAL Study: Type, Frequency, and Association with Asthma Outcomes. The Journal of Allergy and Clinical Immunology: In Practice. 2017;5(4):1071-81.e9.
556. Batura N, Pulkki-Brännström A-M, Agrawal P, Bagra A, Haghparast- Bidgoli H, Bozzani F, et al. Collecting and analysing cost data for complex public health trials: reflections on practice. Global Health Action. 2014;7(1):23257.
557. Franklin M, Lomas J, Richardson G. Conducting Value for Money Analyses for Non-randomised Interventional Studies Including Service Evaluations: An Educational Review with Recommendations. Pharmacoeconomics. 2020;38(7):665-81.
558. Detsky AS, Laupacis A. Relevance of Cost-effectiveness Analysis to Clinicians and Policy Makers. JAMA. 2007;298(2):221-4.
559. Chen L-C, Ashcroft DM, Elliott RA. Do economic evaluations have a role in decision-making in Medicine Management Committees? A qualitative study. Pharm World Sci. 2007;29(6):661-70.
560. Garattini L, van de Vooren K. Budget impact analysis in economic evaluation: a proposal for a clearer definition. Eur J Health Econ. 2011;12(6):499-502.
561. Trueman P, Drummond M, Hutton J. Developing guidance for budget impact analysis. Pharmacoeconomics. 2001;19(6):609-21.
562. Moore RA, Derry S, Wiffen PJ, Straube S, Aldington DJ. Overview review: Comparative efficacy of oral ibuprofen and paracetamol (acetaminophen) across acute and chronic pain conditions. European Journal of Pain. 2015;19(9):1213-23.
563. Srivastava K, Arora A, Kataria A, Cappelleri JC, Sadosky A, Peterson AM. Impact of reducing dosing frequency on adherence to oral therapies: a literature review and meta-analysis. Patient preference and adherence. 2013;7:419-34.
564. Morales DR, Dreischulte T, Lipworth BJ, Donnan PT, Jackson C, Guthrie B. Respiratory effect of beta-blocker eye drops in asthma: population based study and meta-analysis of clinical trials. Br J Clin Pharmacol. 2016.
565. Morales D. Quantifying the risk of beta-blockers and non-steroidal anti-inflammatory drugs in asthma: University of Dundee; 2014.
566. Morales DR, Lipworth BJ, Donnan PT, Jackson C, Guthrie B. Respiratory effect of beta-blockers in people with asthma and cardiovascular disease: population-based nested case control study. BMC Med. 2017;15(1):18.
567. Scheiman JM. Strategies to reduce the GI risks of antiplatelet therapy. Rev Cardiovasc Med. 2005;6 Suppl 4:S23-31.
568. Serebruany VL, Dinicolantonio JJ, Can MM, Pershukov IV, Kuliczkowski W. Gastrointestinal adverse events after dual antiplatelet therapy: clopidogrel is safer than ticagrelor, but prasugrel data are lacking or inconclusive. Cardiology. 2013;126(1):35-40.
569. Abraham NS, Hartman C, Castillo D, Richardson P, Smalley W. Effectiveness of national provider prescription of PPI gastroprotection among elderly NSAID users. Am J Gastroenterol. 2008;103(2):323-32.
570. Al-Taee AM, Ghoulam E, Lee P, Sbertoli R, Hachem C. Under-utilization of proton pump inhibitors in the prevention of NSAID-related peptic ulcer disease in the elderly. Am J Gastroenterol. 2019;114 (Supplement):S684-S5.
571. Anonymous. Nonsteroidal anti-inflammatory drugs: add an anti-ulcer drug for patients at high risk only. Always limit the dose and duration of treatment with NSAIDs. Prescrire Int. 2011;20(119):216-9.
572. Arora G, Singh G, Triadafilopoulos G. Proton pump inhibitors for gastroduodenal damage related to nonsteroidal anti-inflammatory drugs or aspirin: twelve important questions for clinical practice. Clin Gastroenterol Hepatol. 2009;7(7):725-35.
573. Blandizzi C, Tuccori M, Colucci R, Gori G, Fornai M, Antonioli L, et al. Clinical efficacy of esomeprazole in the prevention and healing of gastrointestinal toxicity associated with NSAIDs in elderly patients. Drugs Aging. 2008;25(3):197-208.
574. Payne K, McAllister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. Health Econ. 2013;22(3):258-71.

575. Lorgelly PK. Choice of Outcome Measure in an Economic Evaluation: A Potential Role for the Capability Approach. Pharmacoeconomics. 2015;33(8):849-55.
576. Cousens S, Hargreaves J, Bonell C, Armstrong B, Thomas J, Kirkwood BR, et al. Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. J Epidemiol Community Health. 2011;65(7):576.
577. Degli Esposti M, Spreckelsen T, Gasparrini A, Wiebe DJ, Bonander C, Yakubovich AR, et al. Can synthetic controls improve causal inference in interrupted time series evaluations of public health interventions? Int J Epidemiol. 2020.
578. National Institute for Health and Care Excellence. Prophylaxis against infective endocarditis (CG64.1). 2015.
579. Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. 2 ed: Houghton Mifflin; 2002.
580. Jaksa A, Wu J, Jónsson P, Eichler H-G, Vititoe S, Gatto NM. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. Journal of Comparative Effectiveness Research. 2021;10(9):711-31.
581. Ghosh RE, Crellin E, Beatty S, Donegan K, Myles P, Williams R. How Clinical Practice Research Datalink data are used to support pharmacovigilance. Ther Adv Drug Saf. 2019;10:2042098619854010.
582. Williams R, Jenkins DA, Ashcroft DM, Brown B, Campbell S, Carr MJ, et al. Diagnosis of physical and mental health conditions in primary care during the COVID-19 pandemic: a retrospective cohort study. The Lancet Public Health. 2020;5(10):e543-e50.
583. Morgan CJ. Reducing bias using propensity score matching. J Nucl Cardiol. 2018;25(2):404-6.
584. Landewe R, van der Heijde D. Primer: challenges in randomized and observational studies. Nat Clin Pract Rheumatol. 2007;3(11):661-6.
585. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1):64-74.
586. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. BMJ. 1997;315(7116):1151.
587. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. Pharmacoepidemiol Drug Saf. 2003;12(7):551-8.
588. McMahon AD. Observation and Experiment with the Efficacy of Drugs: A Warning Example from a Cohort of Nonsteroidal Anti-inflammatory and Ulcer-healing Drug Users. Am J Epidemiol. 2001;154(6):557-62.
589. Caro JJ, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, et al. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014;17(2):174-82.
590. Linden A. Conducting Interrupted Time-series Analysis for Single- and Multiple-group Comparisons. 2015;15(2):480-500.
591. Stevens S, Valderas JM, Doran T, Perera R, Kontopantelis E. Analysing indicators of performance, satisfaction, or safety using empirical logit transformation. 2016;352:i1114.
592. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Acad Pediatr. 2013;13(6 Suppl):S38-44.
593. Wooldridge JM. Introductory econometrics: a modern approach. 5th ed. Mason, Ohio: South-Western Cengage Learning; 2012.
594. Baum CF, Schaffer ME. ACTEST: Stata module to perform Cumby-Huizinga general test for autocorrelation in time series. Statistical Software Components. Boston College Department of Economics2013.
595. Prais S, Winsten C. Trend estimators and serial correlation. Commission Discussion Papers, Statistics. Chicago: Cowles Foundation for Research in Economics; 1954.
596. Nelson BK. Statistical methodology: V. Time series analysis using autoregressive integrated moving average (ARIMA) models. Acad Emerg Med. 1998;5(7):739-44.
597. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.

598. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.
599. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006;11(2):193-206.
600. Petropoulou M, Mavridis D. A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: a simulation study. Stat Med. 2017;36(27):4266-80.
601. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials. 2015;45(Pt A):139-45.
602. Dumic I, Nordin T, Jecmenica M, Stojkovic Lalosevic M, Milosavljevic T, Milovanovic T. Gastrointestinal Tract Disorders in Older Age. Can J Gastroenterol Hepatol. 2019;2019:6757524-.
603. Lallana MJ, Feja C, Aguilar-Palacio I, Malo S, Rabanaque MJ. Use of Non-Steroidal Anti-Inflammatory Drugs and Associated Gastroprotection in a Cohort of Workers. Int J Environ Res Public Health. 2018;15(9):1836.
604. Cullen DJE, Seager J, Holmes S, Doherty M, Wilson J, Garrud P, et al. Pharmacoepidemiology of non-steroidal anti-inflammatory drug use in Nottingham general practices. Aliment Pharmacol Ther. 2000;14:177-85.
605. Crooks CJ, West J, Card TR. Comorbidities affect risk of nonvariceal upper gastrointestinal bleeding. Gastroenterology. 2013;144(7):1384-e19.
606. Eastwood GL. The role of smoking in peptic ulcer disease. J Clin Gastroenterol. 1988;10 Suppl 1:S19-23.
607. Strate LL, Singh P, Boylan MR, Piawah S, Cao Y, Chan AT. A Prospective Study of Alcohol Consumption and Smoking and the Risk of Major Gastrointestinal Bleeding in Men. PLoS One. 2016;11(11):e0165278.
608. Braekkan SK, van der Graaf Y, Visseren FL, Algra A. Obesity and risk of bleeding: the SMART study. J Thromb Haemost. 2016;14(1):65-72.
609. Molnar AO, Bota SE, Garg AX, Harel Z, Lam N, McArthur E, et al. The Risk of Major Hemorrhage with CKD. 2016;27(9):2825-32.
610. Saeed F, Agrawal N, Greenberg E, Holley JL. Lower Gastrointestinal Bleeding in Chronic Hemodialysis Patients. International Journal of Nephrology. 2011;2011:272535.
611. Brown JD, Goodin AJ, Lip GYH, Adams VR. Risk Stratification for Bleeding Complications in Patients With Venous Thromboembolism: Application of the HAS-BLED Bleeding Score During the First 6 Months of Anticoagulant Treatment. J Am Heart Assoc. 2018;7(6).
612. Yuet WC, Derasari D, Sivoravong J, Mason D, Jann M. Selective Serotonin Reuptake Inhibitor Use and Risk of Gastrointestinal and Intracranial Bleeding. J Am Osteopath Assoc. 2019;119(2):102-11.
613. Narum S, Westergren T, Klemp M. Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis. 2014;4(5):e004587.
614. Jahromi SR, Togha M, Fesharaki SH, Najafi M, Moghadam NB, Kheradmand JA, et al. Gastrointestinal adverse effects of antiepileptic drugs in intractable epileptic patients. Seizure. 2011;20(4):343-6.
615. De Leon J. A practitioner's guide to prescribing antiepileptics and mood stabilizers for adults with intellectual disabilities. New York: Springer; 2012.
616. van den Ham HA, Klungel OH, Leufkens HGM, van Staa TP. The patterns of anticoagulation control and the risk of stroke, bleeding and mortality in patients with non-valvular atrial fibrillation. J Thromb Haemost. 2013;11(1):107-15.
617. Lue A, Lanas A. Protons pump inhibitor treatment and lower gastrointestinal bleeding: Balancing risks and benefits. World J Gastroenterol. 2016;22(48):10477-81.
618. Parisi R, Rutter MK, Lunt M, Young HS, Symmons DPM, Griffiths CEM, et al. Psoriasis and the Risk of Major Cardiovascular Events: Cohort Study Using the Clinical Practice Research Datalink. J Invest Dermatol. 2015;135(9):2189-97.
619. Lamberts M, Olesen JB, Ruwald MH, Hansen CM, Karasoy D, Kristensen SL, et al. Bleeding after initiation of multiple antithrombotic drugs, including triple therapy, in atrial fibrillation patients following myocardial infarction and coronary intervention: a nationwide cohort study. Circulation. 2012;126(10):1185-93.

620. Nielen JTH, Dagnelie PC, Emans PJ, Veldhorst-Janssen N, Lalmohamed A, van Staa T-P, et al. Safety and efficacy of new oral anticoagulants and low-molecular-weight heparins compared with aspirin in patients undergoing total knee and hip replacements. Pharmacoepidemiol Drug Saf. 2016;25(11):1245-52.
621. Komen JJ, Hjemdahl P, Mantel-Teeuwisse AK, Klungel OH, Wettermark B, Forslund T. Concomitant Anticoagulant and Antidepressant Therapy in Atrial Fibrillation Patients and Risk of Stroke and Bleeding. Clin Pharmacol Ther. 2019.
622. Wright AK, Kontopantelis E, Emsley R, Buchan I, Sattar N, Rutter MK, et al. Life Expectancy and Cause-Specific Mortality in Type 2 Diabetes: A Population-Based Cohort Study Quantifying Relationships in Ethnic Subgroups. Diabetes Care. 2017;40(3):338.
623. Iwagami M, Tomlinson LA, Mansfield KE, Douglas IJ, Smeeth L, Nitsch D. Gastrointestinal bleeding risk of selective serotonin reuptake inhibitors by level of kidney function: A population-based cohort study. Br J Clin Pharmacol. 2018;84(9):2142-51.
624. Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, et al. A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: A systematic review with economic modelling. Health Technol Assess. 2006;10(38):iii-183.
625. Halvorsen S, Storey RF, Rocca B, Sibbing D, Ten Berg J, Grove EL, et al. Management of antithrombotic therapy after bleeding in patients with coronary artery disease and/or atrial fibrillation: expert consensus paper of the European Society of Cardiology Working Group on Thrombosis. Eur Heart J. 2017;38(19):1455-62.
626. Staerk L, Fosbol EL, Lamberts M, Bonde AN, Gadsboll K, Sindet-Pedersen C, et al. Resumption of oral anticoagulation following traumatic injury and risk of stroke and bleeding in patients with atrial fibrillation: A nationwide cohort study. Eur Heart J. 2018;39(19):1698-705.
627. Baker CL, Dhamane AD, Mardekian J, Dina O, Russ C, Rosenblatt L, et al. Comparison of Drug Switching and Discontinuation Rates in Patients with Nonvalvular Atrial Fibrillation Treated with Direct Oral Anticoagulants in the United States. Adv Ther. 2019;36(1):162-74.
628. Coleman CI, Tangirala M, Evers T. Treatment Persistence and Discontinuation with Rivaroxaban, Dabigatran, and Warfarin for Stroke Prevention in Patients with Non-Valvular Atrial Fibrillation in the United States. PLoS One. 2016;11(6):e0157769.
629. Groot AE, Vermeij JM, Westendorp WF, Nederkoorn PJ, van de Beek D, Coutinho JM. Continuation or Discontinuation of Anticoagulation in the Early Phase After Acute Ischemic Stroke. Stroke. 2018;49(7):1762-5.
630. Hellfritzsch M, Grove EL, Husted SE, Rasmussen L, Poulsen BK, Johnsen SP, et al. Clinical events preceding switching and discontinuation of oral anticoagulant treatment in patients with atrial fibrillation. Europace. 2017;19(7):1091-5.

Appendix

Appendix A – Search criteria for the systematic review of economic evaluations on reducing hazardous prescribing in Chapter Two

Table A.1: Search terms for 'health economics'- Concept 1

Medline		Embase	
CRD	**NICE**	**CRD**	**NICE**
1 Economics/	economics/	economics/	health economics/
2 exp 'costs and cost analysis'/	exp 'costs and cost analysis'/	exp 'costs and cost analysis'/	exp economic evaluation/
3 Economics, Dental/			exp Health Care Cost/
4 exp economics, hospital/	exp economics, hospital/	exp economics, hospital/	(economic* or pharmaco?economic*).ti.
5 Economics, Medical/	exp economics, medical/	exp economics, medical/	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
6 Economics, Nursing/	economics, nursing/	economics, nursing/	cost*.ti
7 Economics, Pharmaceutical/	economics, pharmaceutical/	economics, pharmaceutical/	budget/
8 (economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab.	cost*.ti.	cost*.ti.	(value adj2 (money or monetary)).ti,ab.
9 (expenditure$ not energy).ti,ab.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.	budget*.ti,ab.
10 value for money.ti,ab.	(value adj2 (money or monetary)).ti,ab.	(value adj2 (money or monetary)).ti,ab.	exp fee/
11 budget$.ti,ab.	budget*.ti,ab.	budget*.ti,ab.	funding/
12 or/1-11	exp budgets/	exp budgets/	(price* or pricing*).ti,ab.
13 ((energy or oxygen) adj cost).ti,ab.	value of life/	value of life/	(financ* or fee or fees).ti,ab.
14 (metabolic adj cost).ti,ab.	exp 'fees and charges'/	exp 'fees and charges'/	
15 ((energy or oxygen) adj expenditure).ti,ab.	(financ* or fee or fees).ti,ab.	(financ* or fee or fees).ti,ab.	
16 or/13-15	(economic* or pharmaco?economic*).ti.	(economic* or pharmaco?economic*).ti.	
17 12 not 16	(price* or pricing*).ti,ab.	(price* or pricing*).ti,ab.	
18 letter.pt.			
19 editorial.pt.			
20 historical article.pt.			
21 or/18-20			
22 17 not 21			

23 exp animals/ not humans/
24 22 not 23
25 bmj.jn.
26 'cochrane database of systematic reviews'.jn.
27 health technology assessment winchester england.jn.
28 or/25-27
29 24 not 28
30 limit 29 to yr='2010 - Current'

Table A.2: Search terms for 'Decision-analytic model' – Concept 2

Medline	Embase
exp models, economic/	1 statistical model/
*models, theoretical/	2 exp economic aspect/
*models, organizational/	3*theoretical model/
markov chains/	4*nonbiological model/
monte carlo method/	5 stochastic model/
exp decision theory/	6 decision theory/
(markov* or monte carlo).ti,ab.	7 decision tree/
econom* model*.ti,ab.	8 monte carlo method/
(decision* adj2 (tree* or analy* or model*)).ti,ab.	9 (markov* or monte carlo).ti,ab.
	10 econom* model*.ti,ab.
	(1and2)or 3-10

Table A.3: Search terms for 'Medication error' – Concept 3

Embase and Medline	Source for search term
(medication and error).mp.	Elliott 2018
(inappropriate and prescribing).mp.	Elliott 2018
inappropriate medication.mp.	Elliott 2018
(preventable and adverse and drug and event*).mp.	Elliott 2018
(preventable and adverse and drug and reaction*).mp.	Elliott 2018
(prescribing and error*).mp.	Elliott 2018
(transcription and error*).mp.	Elliott 2018
(medication and discrep*).mp.	Elliott 2018
(medication and omission*).mp.	Elliott 2018
inappropriate prescribing/	added Mesh
exp medical error/	added Mesh
exp ME/	added Mesh
(hazardous and prescr*).mp.	added free text

Appendix B – Prescribing safety indicators defining HPE types in SMASH

Table B.1: Overview of prescribing safety indicators (PSIs) used in SMASH to identify hazardous prescribing events (HPE) in the at-risk population (excluding monitoring errors)

PSI	Group at risk of HPE (denominator)	Group with HPE identified (numerator)
1	Patients aged ≥18 years with a history of peptic ulcer or GI bleed and not prescribed a GPA	Prescription of an antiplatelet drug without co-prescription of a GPA, to a patient with a history of peptic ulcer or GI bleed
2	Patients aged ≥18 years prescribed an oral anticoagulant (OAC)	Prescription an OAC in combination with an oral NSAID
3	Patients aged ≥18 years prescribed an OAC without co-prescription of a GPA	Prescription of an OAC and an antiplatelet drug in combination without co-prescription of a GPA
4	Patients aged ≥18 years prescribed aspirin without co-prescription of a GPA	Prescription of aspirin in combination with another antiplatelet drug (without co-prescription of a GPA)
5	Patients aged ≥65 years without co-prescription of a GPA	Prescription of an oral NSAID to patients aged ≥65 years without co-prescription of a GPA
6	Patients aged ≥18 years with a history of peptic ulcer without co-prescription of a GPA	Prescription of an oral NSAID without co-prescription of a GPA to a patient with a history of peptic ulcer
7	Patients aged ≥18 years with asthma (unresolved) prescribed a LABA	Prescription of a long-acting beta-2 agonist inhaler to a patient with asthma (unresolved) who is not also prescribed an inhaled corticosteroid
8	Patients aged ≥18 years with asthma (unresolved)	Prescription of a non-selective beta-blocker to a patient with asthma (unresolved)
9	Patients aged ≥18 years who have a diagnosis of heart failure	Prescription of an oral NSAID to a patient with a history of HF
10	Patients aged ≥18 years with an eGFR <45	Prescription of an oral NSAID to a patient with eGFR <45

GI bleed: gastro-intestinal bleed; GPA: gastroprotective agent; HPE: hazardous prescribing event; NSAID: non-steroidal anti-inflammatory drug; OAC: oral anticoagulant including vitamin K antagonist and direct oral anticoagulants

Appendix C – CHEERS checklist for economic evaluations

Table C.1: CHEERS reporting checklist with key sections reporting recommended criteria of each item relevant to the economic evaluation in Chapter Three and Chapter Six (278)

Item	Item No	Recommendation	Chapter Three	Chapter Six
Title and abstract				
Title	1	Identify the study as an economic evaluation or use more specific terms, such as 'cost-effectiveness analysis', and describe the interventions compared.	N/A	N/A
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	N/A	N/A
Introduction				
Background and objectives	3	Provide an explicit statement of the broader context for the study.	3.1	6.1
		Present the study question and its relevance for health policy or practice decisions.	3.1	6.1
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	3.2.1.1	6.2.1.1
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	2.9; 3.2.1	2.9; 3.2.1
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	3.2.1.3	6.2.1.3
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	2.9; 3.2.1.2	2.9; 6.2.1.2
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	3.2.1	5.2.2; 6.2.1.4
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	3.2.1.3	6.2.1.3; 5.2.2
Choice of health outcome	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	3.2.1.3; 3.2.2; 3.1	6.1; 6.2.1.3
Measurement of effectiveness	11a	Single study-based estimates: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	3.2.2	3.2.2; 6.2.2
	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of data.	N/A	N/A
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	N/A	N/A
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	3.2.3	3.2.3; 6.2.3
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	N/A	5.2.4.2; 6.2.4

Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	3.2.1.3	6.2.1.3
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	3.2.1.4	6.2.1.4
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	3.2.1	Appendix M
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	3.2.4; 3.2.5	6.2.5; 6.2.6; 6.2.7
Results				
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	3.3.2	3.3.2; 5.2.4; 6.3.1
Incremental cost and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	3.3.4	6.3.4
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental costs and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	3.3.4	6.3.4
	20b	Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	N/A	6.3.4
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	N/A	6.3.4
Discussion				
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	1.1	6.4
Other				
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	N/A	N/A
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	N/A	N/A

Appendix D – Detailed description of methods used in SMASH effectiveness study

Appendix D.1. The Interrupted times series analysis (ITSA) [Peek et al. 2020]

The interrupted times series analysis (ITSA) was carried out by researchers from the SMASH analysis team using the `itsa` command in *Stata* version 15 as described by Linden et al. (2015) (590). The `itsa` command incorporates ordinary least square (OLS) regression-based models, sometimes referred to as segmented regression. Because OLS regression assumes a linear trend and the descriptive statistics showed a ceiling effect after intervention was implemented, the observed HPE rates (in percentages) were log transformed (591). The ITSA results were back transformed from the log scale after they were meta-analysed. The underlying regression model of the ITSA denoted as:

$$Y_t = \beta_0 + \beta_1 T_1 + \beta_2 X_t + \beta_3 X_t T_t + \epsilon_t$$

Where Y_t was the aggregated HPE rate in percent at time *t* (logit transformed); *t* was the time since start of the observations where each *T* was one of these observations; X_t was a dummy variable for the intervention (before intervention start *X=0*, after intervention start *X=1*). The intercept, β_0, shows the HPE rate at the start of the observations at *t=0*. While β_1 explains the slope of the pre-intervention trend, β_2 and β_3 describe the effect of the intervention. The immediate effect of the intervention (level change) was indicated by β_2 and β_3 was the effect of the intervention over time.

In general with OLS, it is important to check for correlated values or seasonality that often occur in time series data (265, 592). The observed values in the SMASH intervention did not suggest any seasonality. However, serial correlation between observed HPE rates at subsequent months could not be ruled out.

To account for potential serial correlation, the analysts followed suggestions from Linden et al. (2015) fitting the OLS using the `newey` command that is included as default in the `itsa` command in *Stata* version 15. The `newey` command fits an OLS but generates Newey-West standard errors instead of the usual OLS standard errors. The Newey-west method to generate standard errors accounts for autocorrelation up to a specified lag and for heteroscedasticity (593). However, it was required to test how well the model fitted by testing the autocorrelation in the error term. The analysis team used `actest`, as recommended by Linden et al. (2015) to perform the Cumby-Huizinga general test for

autocorrelation (594). The Cumby-Huizinga tests the null hypothesis that the time series is a moving average of known order *q*, which could be zero or a positive value. In this analysis, it was tested if autocorrelation was present up to the lag specified in the `newey` command. A lag of 12 was used. If this was not the case further modelling approaches with different lags, would have been necessary or other methods, such as the Prais–Winsten method (595) that is also incorporated in `itsa` command, or autoregressive integrated moving average (ARIMA) (596).

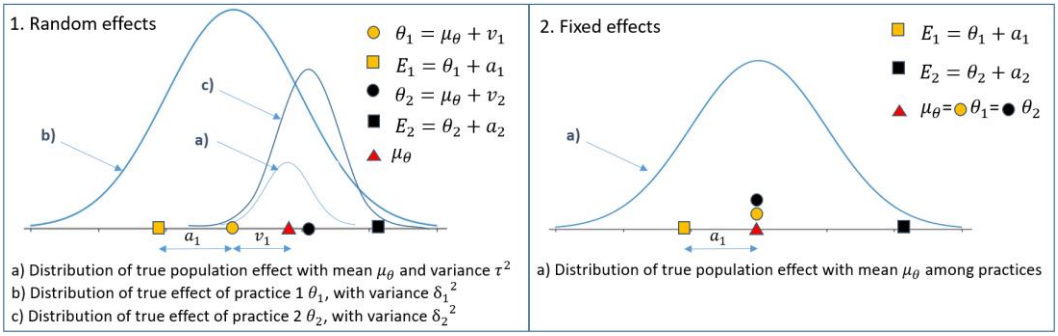
The reported estimates from `itsa` command ($\beta_1, \beta_2, \beta_3$) on the log scale were used in linear post estimation commands to identify the HPE rate without the SMASH intervention based on the extrapolated pre-intervention trend and the difference between intervention (observed) and extrapolated pre-intervention trend at different time points post intervention (4, 12, 24 and 52 weeks).

The ITSA was carried out at practice level for the two primary outcomes of the study, the prevalence of exposure to (i) any potentially hazardous prescribing (10 indicators, Appendix B) and (ii) any inadequate blood-test monitoring (2 indicators) among patients with risk factors for such prescribing and monitoring (as defined by individual indicators). For the individual indicators an overall ITSA for all practices was conducted with pooled HPE rates from all practices.

Appendix D.2 – Meta-analysis [Peek et al. 2020]

The random effects meta-analysis chosen for the SMASH intervention by researchers from the SMASH analysis team accounted for practice heterogeneity if compared with a fixed effects meta-analysis. Figure:2 shows that the observed effect E in a fixed effect model is drawn from a distribution around the true effect θ. The true effect is 'fixed' for all practices. In random effects meta-analysis, the population effect size θ is assumed to be a continuous random variable, not fixed, where each practices true effect θ_i is drawn from a distribution of θ (Figure:1a). For the variable θ the width of the distribution of the true effect is defined by the variance τ^2, the practice heterogeneity, and the centre is the average intervention effect among all practices μ_θ. The random effects meta-analysis is based on the assumption that there is a between practice error ν, additionally to the within practice error α (Figure:1).

Figure D.1: Visualisation of basic assumptions in (1) random effects and (2) fixed effects models



In conclusion, the advantage of the random effects model is the acknowledgment of between-practice heterogeneity. To apply a random effects model, it is necessary to estimate the heterogeneity between studies before any form of weighting can be performed for the meta-analysis. There are various methods that can be used in order to assess this between-study variance that is required to estimate the true population average effect μ_θ. The most common approach is the DerSimonian-Laird method (279, 597).

The DerSimonian-Laird study heterogeneity estimator $\hat{\tau}^2_{DL}$ denotes as

Equation 1: $\hat{\tau}^2_{DL} = \frac{Q_{\hat{w}} - (k-1)}{\left(\sum_i \hat{w}_i - \frac{\sum_i \hat{w}_i{}^2}{\sum_i \hat{w}_i}\right)}$

where $Q_{\hat{w}}$ is the Q statistic, (k -1) are the degrees of freedom, and $\hat{w}_i = \frac{1}{\hat{\delta}_i^2}$ the inverse within study variance for the Q statistic of a fixed effect model. The Q statistic is a measure of total variation among the studies and is a weighted sum of squares of the observed effects E_i over their mean $\bar{E}_.$,

Equation 2: $Q_{\hat{w}} = \sum_{i=1}^{k} w_i (E_i - \bar{E}_.)^2$

The degrees of freedom, df, represent the Q statistic assuming only within study variance for the k number of practices. Hence, $\hat{\tau}^2_{DL}$is estimated as the difference between total observed variation expressed as $Q_{\hat{w}}$and the expected variation df assuming homogeneity($\hat{\tau}^2 = 0$).

The heterogeneity can also be expressed using the I^2 index (598), which is comparable to a percentage of the between study variance from the total variation $Q_{\hat{w}}$.

Equation 3: $I^2 = \frac{Q_{\hat{w}} - (k-1)}{Q_{\hat{w}}}$

Huedo-Medina et al. (2006) found the I^2 index to be a good addition to the normal Q statistic with the advantage of quantifying the heterogeneity in a way easy to interpret (599).

For the SMASH intervention a bootstrap version of the DerSimonian-Laird method was carried out. This method was found to be the most precise to identify between study variance and to estimate the effect size in a re-analysis of almost 60,000 meta-analyses identified from the Cochrane database (279). This is supported by findings from Petropoulou et al. (2017) (600). The bootstrap version of the DerSimonian-Laird method is based on $\hat{\tau}^2_{DL}$(597, 601) and can be performed as part of the STATA version 15 `metaan` command. The bootstrap version of the DerSimonian-Laird method selects random samples with replacement (1000 repetitions) from all practices. Each sample contained 43 practices equal to the number of practices in the study. Within each sampling set practice heterogeneity is calculated using the formula for $\hat{\tau}^2_{DL}$ above. The mean of the different $\hat{\tau}^2_{DL}$ estimaters of each sampled set is generated and referred to as $\hat{\tau}^2_{DLb}$ (279).

Once between study heterogeneity τ^2 is estimated, this can be used to apply a weighting factor for each practice. For the SMASH intervention inverse variance weighting was chosen. Weights v_i are applied for each practice according to the practice specific within study variance $\delta_i^2$and the general between study variance τ^2among all practices using the inverse of the total variance.

Equation 4: $v_i = \frac{1}{\delta_i^2 + \tau^2}$

Applying the weights estimated using a bootstrap version of the DerSimonian-Laird method the average population effect μ_θ can be estimated after the log transformed estimates are back transformed. The between study heterogeneity is quantified using the I^2 index.

Appendix E – Semi-structured interviews to inform costing of SMASH

Aim of this costing analysis is to assess the cost of the intervention for a typical practice in Salford. According to the qualitative work that has been done the variability between practices in how the intervention was implemented was significant. However, to assess the cost-effectiveness we need to base our assumptions on costs for a typical practice. Variability around this is relevant for sensitivity analysis. Please, feel free to give ranges around the average you would estimate.

1. How was the interviewee involved in the study?
2. Ask for details on the training pharmacists received as part of the intervention (How long, where, provider, expenses, number of attendees)
3. Ask for details on the initial staff meetings (How long, where, provider, expenses, number of attendees)
4. Ask for quantifiable estimates on managing HPEs:
    a. Time pharmacists spend with HPE
    b. Time GPs spend with HPE
    c. Percentage of HPE requiring further actions
5. What did the IT service support entail? How often was it used and who provided it?
6. Where there any costs associated with using and maintaining the dashboard, such as server costs?
7. Ask for details on salary bands of staff involved in the intervention.
8. Did we miss any relevant resources required to deliver the intervention?

Appendix F – Checklists for reporting in observational cohort studies

Table F.1: The RECORD, STROBE and RECORD-PE statements and sections the items are referred to in Chapter Four

Item No	STROBE items	RECORD items	RECORD-PE items	Section
Title and abstract				
1	(a) Indicate the study's design with a commonly used term in the title or the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found.	1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. 1.2: If applicable, the geographical region and timeframe within which the study took place should be reported in the title or abstract. 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.		N/A
Introduction				
Background rationale				
2	Explain the scientific background and rationale for the investigation being			Chapter 4; 4.2
Objectives				
3	State specific objectives, including any prespecified hypotheses.			4.3
Methods				
Study design				
4	Present key elements of study design early in the paper		4.a: Include details of the specific study design (and its features) and report the use of multiple designs if used. 4.b: The use of a diagram(s) is recommended to illustrate key aspects of the study design(s), including exposure, washout, lag and observation periods, and covariate definitions as relevant	4.3.1; 4.3.3; Figure 4.1; Table 4.2
Setting				
5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.			4.3.3; Figure 4.1; Table 4.2
Participants				

Item No	STROBE items	RECORD items	RECORD-PE items	Section
6	(a) Cohort study—give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study—give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross sectional study—give the eligibility criteria, and the sources and methods of selection of participants. (b) Cohort study—for matched studies, give matching criteria and number of exposed and unexposed. Case-control study—for matched studies, give matching criteria and the number of controls per case.	6.1: The methods of study population selection (such as codes or algorithms used to identify participants) should be listed in detail. If this is not possible, an explanation should be provided. 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1.a: Describe the study entry criteria and the order in which these criteria were applied to identify the study population. Specify whether only users with a specific indication were included and whether patients were allowed to enter the study population once or if multiple entries were permitted. See explanatory document for guidance related to matched designs.	4.3.3; Figure 4.1; Table 4.2; 4.3.6
Variables				
7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria if applicable.	7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1.a: Describe how the drug exposure definition was developed. 7.1.b: Specify the data sources from which drug exposure information for individuals was obtained. 7.1.c: Describe the time window(s) during which an individual is considered exposed to the drug(s). The rationale for selecting a particular time window should be provided. The extent of potential left truncation or left censoring should be specified. 7.1.d: Justify how events are attributed to current, prior, ever, or cumulative drug exposure. 7.1.e: When examining drug dose and risk attribution, describe how current, historical or time on therapy are considered. 7.1.f: Use of any comparator groups should be outlined and justified. 7.1.g: Outline the approach used to handle individuals with	4.3.4; 4.3.7; Appendix H

Item No	STROBE items	RECORD items	RECORD-PE items	Section
			more than one relevant drug exposure during the study period.	
Data sources/measurement				
8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.		8.a: Describe the healthcare system and mechanisms for generating the drug exposure records. Specify the care setting in which the drug(s) of interest was prescribed.	0; 4.3.6.2
Bias				
9	Describe any efforts to address potential sources of bias.			4.3.3; 4.3.7.3; 4.3.8
Study size				
10	Explain how the study size was arrived at.			Figure 4.3
Quantitative variables				
11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.			Appendix H
Statistical methods				
12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) Cohort study—if applicable, explain how loss to follow-up was addressed. Case-control study—if applicable, explain how matching of cases and controls was addressed. Cross sectional study—if applicable, describe analytical methods taking account of sampling strategy. (e) Describe any sensitivity analyses.		12.1.a: Describe the methods used to evaluate whether the assumptions have been met. 12.1.b: Describe and justify the use of multiple designs, design features, or analytical approaches.	4.3.7.3; 4.3.8
Data access and cleaning methods				
12		12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. 12.2: Authors should provide information on the data cleaning methods used in the study.		4.3.6
Linkage				
12		12.3: State whether the study included person level,		0

Item No	STROBE items	RECORD items	RECORD-PE items	Section
		institutional level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.		
Results				
Participants				
13	(a) Report the numbers of individuals at each stage of the study (e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed). (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram.	13.1: Describe in detail the selection of the individuals included in the study (that is, study population selection) including filtering based on data quality, data availability, and linkage. The selection of included individuals can be described in the text or by means of the study flow diagram.		4.4.1; Figure 4.3
Descriptive Data				
14	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders. (b) Indicate the number of participants with missing data for each variable of interest. (c) Cohort study—summarise follow-up time (e.g., average and total amount).			4.4.1; Appendix G**Error! Reference source not found.**
Outcome data				
15	Cohort study—report numbers of outcome events or summary measures over time. Case-control study—report numbers in each exposure category, or summary measures of exposure. Cross sectional study—report numbers of outcome events or summary measures.			
Main results				
16	(a) Give unadjusted estimates and if applicable, confounder adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included. (b) Report category boundaries when continuous variables are categorised. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.			4.4.2

Item No	STROBE items	RECORD items	RECORD-PE items	Section
Other analyses				
17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses.			4.4.2; 4.4.3
Discussion				
Key results				
18	Summarise key results with reference to study objectives.			4.5.1
Limitations				
19	Discuss limitations of the study, considering sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	19.1.a: Describe the degree to which the chosen database(s) adequately captures the drug exposure(s) of interest.	4.5.4; 4.5.5
Interpretation				
20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	20.a: Discuss the potential for confounding by indication, contraindication or disease severity or selection bias (healthy adherer/sick stopper) as alternative explanations for the study findings when relevant.		4.5.1
Generalisability				
21	Discuss the generalisability (external validity) of the study results.			4.5.5
Other Information				
Funding				
22	Give the source of funding and the role of the funders for the present study and if applicable, for the original study on which the present article is based.			N/A
Accessibility of protocol, raw data and programming code				
23		22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.		N/A

RECORD: reporting of studies conducted using observational routinely collected data; RECORD-PE=RECORD for pharmacoepidemiological research; STROBE: strengthening the reporting of observational studies in epidemiology. This checklist has been duplicated from table 1 in BMJ 2018;363:k3532, as a standalone document for readers to print out or fill in electronically.

Appendix G – Considerations on confounders in cohort study Chapter Four

Table G.1: Reasoning behind grouping of variables as confounders, collider, indirect/direct risk factors and those without any effect on exposure or outcome variable given by GPs and sources from the literature to support their assessment

Covariate definition	Link to exposure and outcome [source from literature to support GPs decisions]
Patient specific information	
Sex: male or female	GI risk factor (355) [CHA2DS2-VASc]
Socio-economic status: index of multiple deprivation (IMD). Level 1 (least deprived) to level 5 (most deprived)	GI risk factor (355)
Ethnicity: subdivided into white, Asian, Black, other and unknown	GI risk factor (355)
Age: age in the year of index date	Confounder [CHA2DS2-VASc, HAS-BLED] GI bleeding events more common with age due to thinner stomach lining, reduced blood flow and secretion of protective prostaglandins or mucin (602). NSAID use is associated with age due to increased incidence of conditions with age requiring NSAID treatment (for example: osteoarthritis and gout) (603, 604). Relationship probably not linear. From 65 years on NSAID use is not recommended anymore and GPs might be more hesitant to prescribe NSAIDs [NICE guidelines].
Smoking: smoking status was subdivided into current smoker, ex-smoker, never smoker and missing smoking status.	GI risk factor [indirect (605)] Smoking affects gastric acid secretion, inhibits protective bicarbonate secretion, reduces mucosal blood flow, which increases the risk of peptic ulcers, which could increase GI bleeding risk (606). However, other sources report no significant effect on GI bleeding (607).
Alcohol dependence: includes heavy drinkers and excludes moderate alcohol consumption.	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. High alcohol consumption can cause exfoliation of the gastric epithelium, necrosis of tissue and microvascular erosions resulting in increased GI bleeding risk (607).
High BMI: BMI >30 kg/m^2 (obese)	GI risk factor [indirect] High BMI does not cause GI bleeds, but could increase risk of comorbidities associated with bleeding, such as oesophageal varices (608).
Diet	No effect No effect on bleeding risk found that would not be covered by other risk factors already included.
Comorbidities	
Renal insufficiency: chronic kidney disease stage 4 or worse and renal insufficiencies with similar severity, chronic dialysis, and transplant	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Severe renal disease is considered an independent risk factor of GIB and other bleeding by causing albuminuria, platelet dysfunction (609) or anaemia in haemodialysis patients(610). NSAID treatment can cause renal damage.
Liver disease: severe liver disease [code list was reviewed by GPs and found to affect bleeding risk]	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Liver disease and the increased risk of varices and the potentially affected haemostasis can increase GI bleeding risk(605). The metabolism of OACs in the liver might also be affected. NSAID treatment can cause liver damage.

Atrial fibrillation (AF)	GI risk factor [indirect] (355) In AF OACs are a common treatment that can increase bleeding risk (41). An impaired blood flow, can increase the workload for the kidneys and result in renal insufficiency, which is then linked with an increased bleeding risk.
Coronary heart disease (CHD): diagnosis of CHD, heart failure (HF), myocardial infarction or angina	GI risk factor [indirect] (355) [CHA2DS2-VASc] Common treatment of coronary heart disease in primary and secondary prophylaxis are aspirin and antiplatelets, which increase GI bleeding risk.
Cerebro-vascular disease: diagnosis of stroke or transient ischaemic attack	GI risk factor [indirect] (355) [CHA2DS2-VASc, HAS-BLED]
Peripheral artery disease (PAD) diagnosis	GI risk factor [indirect] [CHA2DS2-VASc] Common treatment of peripheral artery disease are OACs that can increase bleeding risk. PAD is also associated with an increased risk of CHD.
Venous thrombo-embolism (VTE): diagnosis of pulmonary embolism or deep vein thrombosis	GI risk factor [indirect] (355, 611) [CHA2DS2-VASc] Common treatment of peripheral artery disease are OACs that can increase bleeding risk.
Valvular heart disease diagnosis	GI risk factor [indirect] Can result in heart failure or prosthetic valve replacements, which often require OAC or aspirin treatment.
Hypertension diagnosis (controlled or uncontrolled)	GI risk factor [indirect] (355) [CHA2DS2-VASc]
Uncontrolled hypertension: Blood pressure measurements >160mmHg	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Strong predictor of bleeding risk (379) and can be a cause of NSAID treatment.
Labile International Normalized ratio (INR) as defined in HAS-BLED(379)	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Strong predictor of bleeding risk (379) and can be a cause of NSAID treatment.
Diabetes diagnosis (type 1 and type 1)	GI risk factor [indirect] (355)
COPD diagnosis	GI risk factor [indirect] Common treatment of COPD with corticosteroids, which increase GI bleeding risk.
Cancer: diagnosis of the 12 most common cancer types (includes most GI cancers)	GI risk factor [indirect] (355, 605)
Bleeding event: diagnosis of GI bleeding events (including perforated ulcers), intracranial haemorrhages (ICH) and rectal bleeds (haematuria and haemoptysis were considered too minor to be included when recorded in CPRD but were included when identified as primary diagnosis in hospital records)	Confounder (355, 605) [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Any previous bleeding event is a strong predicter of a recurrent event (379).
Peptic ulcer: diagnosis of peptic ulcer disease (excluded were perforated and haemorrhagic ulcers, which were included as bleeding events)	Confounder Prescribing of NSAIDs negatively influenced by presence of disease. Peptic ulcer is a strong predictor of GI bleeding risk (355, 605) and can be a cause of NSAID treatment.
Diagnosis of dyspepsia and heartburn	GI risk factor [indirect] Can be symptoms of ulcers and can be caused by NSAID treatment.
GI inflammation: diagnosis of gastritis, duodenitis and oesophagitis	GI risk factor [indirect] Increases the risk of ulcers and erosions that are associated with an increased bleeding risk. Can be caused by NSAID treatment.

H. pylori: diagnosis of H. pylori infection or positive test	GI risk factor [indirect] Infections with H. pylori are associated with an increased risk of ulcers.
Oesophageal varices diagnosis	Confounder Prescribing of NSAIDs negatively influenced by presence of disease. Varices can rupture and bleed.
Anaemia diagnosis	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by presence of disease. Anaemia is a strong predictor of bleeding events (379) and a result of bleeding events. NSAID use can also cause anaemia.
Medications	
Prescriptions of systemic aspirin	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by prevalent treatment with drug.
Prescriptions of systemic antiplatelets	Confounder [HAS-BLED] Prescribing of NSAIDs negatively influenced by prevalent treatment with drug.
Prescriptions of systemic antidepressants (SSRI and TCAs)	Confounder Prescribing of NSAIDs negatively influenced by prevalent treatment with drug. SSRI increase gastric acid and down regulate 5HT- receptors, which are also on platelets and are required for various mechanisms to activate platelets. This down regulation could increases bleeding (612). TCA inhibit warfarin metabolism and could increase bleeding risk in warfarin patients.
Prescriptions of systemic corticosteroids (excluding inhaled corticosteroids)	Confounder Prescribing of NSAIDs negatively influenced by prevalent treatment with drug. Cortisone can increase bleeding risk especially when used in combination with Aspirin. (613)
Prescription of antibiotics (macrolides)	No clear evidence found that it is associated with exposure or outcome
Prescription of antiepileptic drug (phenytoin or carbamazepine)	GI risk factor [indirect] (355) Carbamazepine, phenytoin and valproic acid can cause thrombocytopenia and are associated with an increased risk of dyspepsia and heartburn (614, 615)
Prescriptions for systemic GPAs (H2-receptor antagonists, proton-pump-inhibitors (PPIs) and misoprostol)	Collider GPAs are commonly used to reduce bleeding risk of NSAIDs and to treat GI adverse events and GI bleedings.
Prescriptions for systemic nitrates	No effect No evidence could be found suggesting an association between nitrates and bleedings or NSAID treatment.
Prescriptions for systemic statins	No effect Controversial evidence.

BMI: body mass index; BNF: British National Formulary; COPD: chronic obstructive pulmonary disease; CPRD: Clinical Practice Research Datalink; HAS-BLED: bleeding risk score for patients using anticoagulants GI: gastro-intestinal; GPA: gastroprotective agent; NSAID: non-steroidal anti-inflammatory agent; SSRI: selective serotonin re-uptake inhibitor; TCA: tricyclic antidepressants

Table G.2: List of potential covariates with details on where code lists were derived from and where the records were extracted from

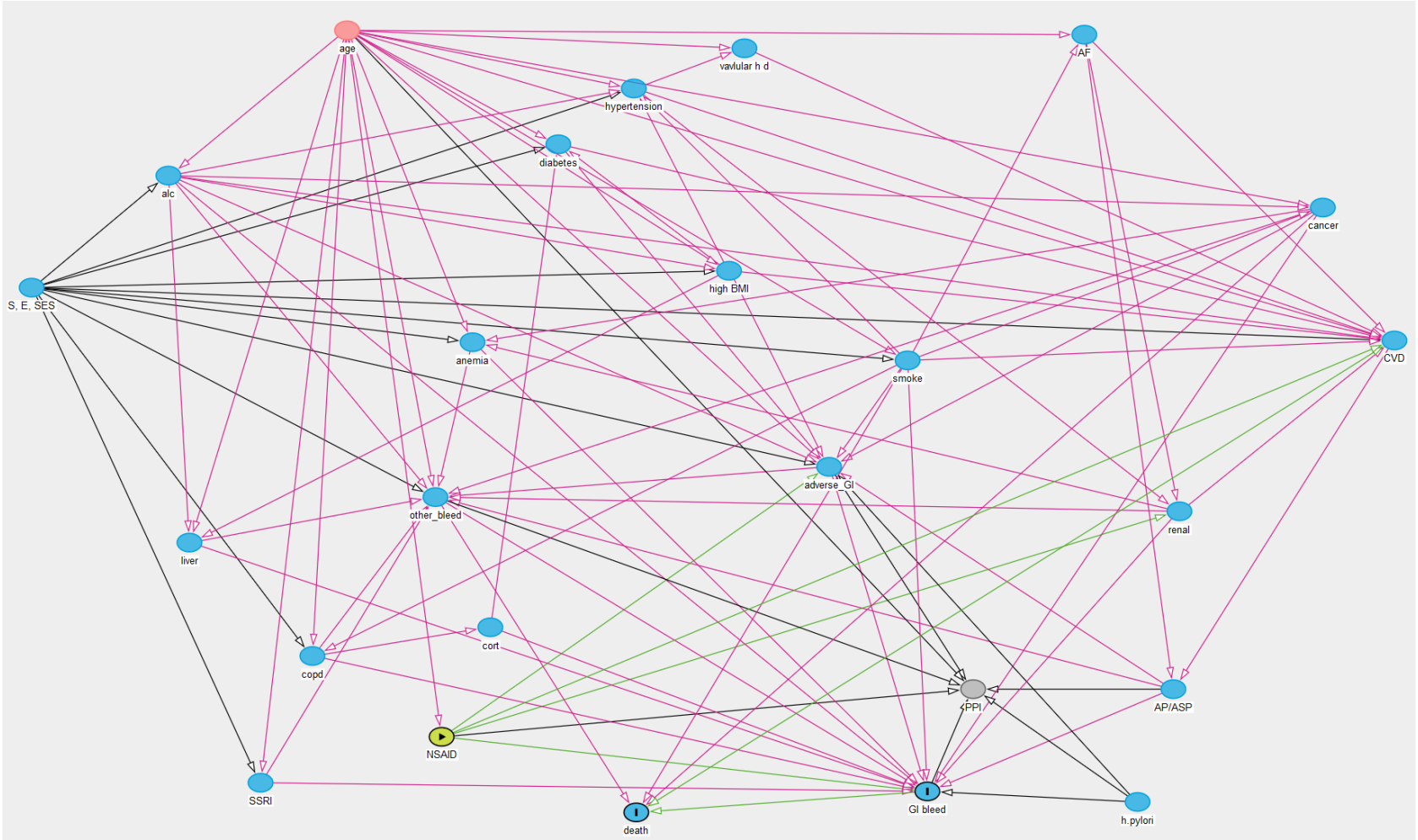
Covariate	Definition	Source of code list	Source of data
Patient specific information			
Sex	Gender (male or female)	N/A	From CPRD records at baseline
Socio-economic status	Index of multiple deprivation (IMD). Level 1 (least deprived) to level 5 (most deprived)	N/A	From linked patient level IMD records
Ethnicity	White, Asian, Black, other and unknown		From HES and CPRD records
Age	Age in the year of index date		From year of birth record in CPRD
Smoking	Smoking status was subdivided into current smoker, ex-smoker, never smoker and missing smoking status.	Code lists linking records with smoking status were derived from colleagues and entity type 4 and 23 were used.	From additional, clinical and referral file records in CPRD.
Alcohol dependence	Alcohol dependence includes heavy drinkers and excludes moderate alcohol consumption.	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
High BMI	BMI >30 kg/m^2 (obese)	Entity type 13 and 14 were used	From additional and clinical records in CPRD.
Comorbidities			
Renal insufficiency	Chronic kidney disease stage 4 or worse and renal insufficiencies with similar severity, chronic dialysis or transplant	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
Liver disease	Severe liver disease [code list was reviewed by GPs and found to affect bleeding risk]	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
Atrial fibrillation (AF)	Diagnosis of AF	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
Coronary heart disease (CHD)	Diagnosis of CHD, heart failure (HF), myocardial infarction or angina	Code lists were previously used in (354) and provided by the authors. Codes for HF were used as identified by PRIMIS.	From clinical CPRD records.
Cerebro-vascular disease	Diagnosis of stroke and transient ischaemic attack	Codes were used as identified by PRIMIS.	From clinical CPRD records.

Peripheral artery disease (PAD)	Diagnosis of PAD	Code list was previously published (616)	From clinical CPRD records.
Venous thrombo-embolism (VTE)	Diagnosis of pulmonary embolism or deep vein thrombosis	Codes were used as identified by PRIMIS	From clinical CPRD records.
Valvular heart disease	Diagnosis of valvular heart disease	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records
Hypertension	Diagnosis of hypertension	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
Uncontrolled hypertension	Blood pressure measurements >160mmHg	Entity type 1 was used [The closest record within 12 months before the index date]	From additional and clinical records in CPRD.
Labile INR		Entity type 232 was used	From test records in CPRD
Diabetes	Diagnosis of diabetes (type 1 and type 1)	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
COPD	Diagnosis of COPD	Codes were used as identified by PRIMIS.	From clinical CPRD records.
Cancer	Diagnosis of the 12 most common cancer types (includes most GI cancers)	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records
Bleeding event	Diagnosis of GI bleeding events (including perforated ulcers), intracranial haemorrhages (ICH) and rectal bleeds (haematuria and haemoptysis were considered too minor to be included when recorded in CPRD but were included when identified as primary diagnosis in hospital records)	Codes were used as identified by PRIMIS and combined with codes previously used in (354) and provided by the authors. From HES records the outcome codes for major bleeding events were used.	From clinical CPRD records.
Peptic ulcer	Diagnosis of peptic ulcer disease (excluded were perforated and haemorrhagic ulcers, which were included as bleeding events)	Codes were used as identified by PRIMIS.	From clinical CPRD records.
Dyspepsia	Diagnosis of dyspepsia and heartburn	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
GI inflammation	Diagnosis of gastritis, duodenitis and oesophagitis	Code list was developed using the code browser and PCD search	From clinical CPRD records.

		(keywords: *gastrit*, *duodeni*, *oesophagi*)	
H. pylori		Code list was developed using the code browser and PCD search (keywords: *pylori* *helicob*)	From clinical CPRD records.
Oesophageal varices	Diagnosis of oesophageal varices	Code lists were previously used in (354) and provided by the authors.	From clinical CPRD records.
Anaemia	Diagnosis of anaemia	Code list was developed using the code browser and PCD search (keywords: anemia, anaemia, *anaem*)	From clinical CPRD records.
Medications			
NSAID	Prescriptions of systemic NSAIDs (excluding aspirin)	Code list was developed using the code browser and PCD search (keywords: all drug names of this group found in BNF)	From therapy CPRD records.
Aspirin	Prescriptions of systemic aspirin	Code list was developed using the code browser and PCD search (keywords: all drug names of this group found in BNF)	From therapy CPRD records.
Antiplatelet	Prescriptions of systemic antiplatelets	Code list was developed using the code browser and PCD search (keywords: all drug names of this group found in BNF)	From therapy CPRD records.
OAC	Prescriptions of systemic warfarin, phenprocoumon, phenindione and DOACs.	Code list was developed using the code browser and PCD search (keywords: all drug names of this group found in BNF)	From therapy CPRD records.
Anti-depressants	Prescriptions of systemic SSRI and TCAs	Code lists were previously used in (354) and provided by the authors.	From therapy CPRD records.
Corticosteroid	Prescriptions of systemic corticosteroids (excluding inhaled corticosteroids)	Code lists were previously used in (354) and provided by the authors.	From therapy CPRD records.
Anti-convulsants	Prescriptions for phenytoin or carbamazepine	Code lists were previously used in (354) and provided by the authors.	From therapy CPRD records.
GPA	Prescriptions for systemic formulations of H2-receptor antagonists, proton-pump-inhibitors (PPIs) and misoprostol	Code list was developed using the code browser and PCD search (keywords: all drug names of this group found in BNF)	From therapy CPRD records.

BMI: body mass index; BNF: British National Formulary; CHA2DS2-VASc: stroke risk score for patients with atrial fibrillation; CPRD: Clinical Practice Research Datalink; GI: gastro-intestinal; GPA: gastroprotective agent; HES: Hospital Episodes Statistics

Figure G.1: DAG showing simplified structure of potential variables influencing exposure (NSAID use) and the outcomes (GI bleed/death). The confounder 'age' is shown in red.



Appendix H – Tests performed as part of the statistical analysis of the cohort study

Figure H.1: Distribution of the propensity score across NSAID users and non-NSAID users in the matched cohort. The area of common support between NSAID users and never NSAID users

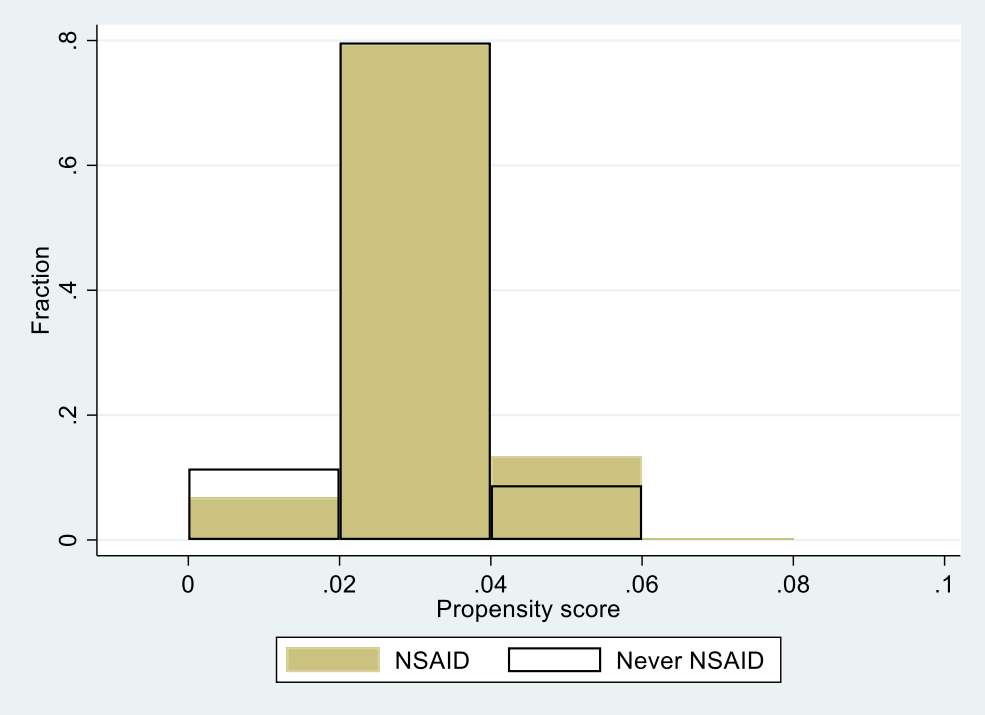


Figure H.2: Standardised difference between Never NSAID users and NSAID users for each variable listed in the baseline characteristics in section 4.1 before [unmatched] and after matching [matched]. A standardised difference greater than 10% is considered to indicate imbalance of this variable.

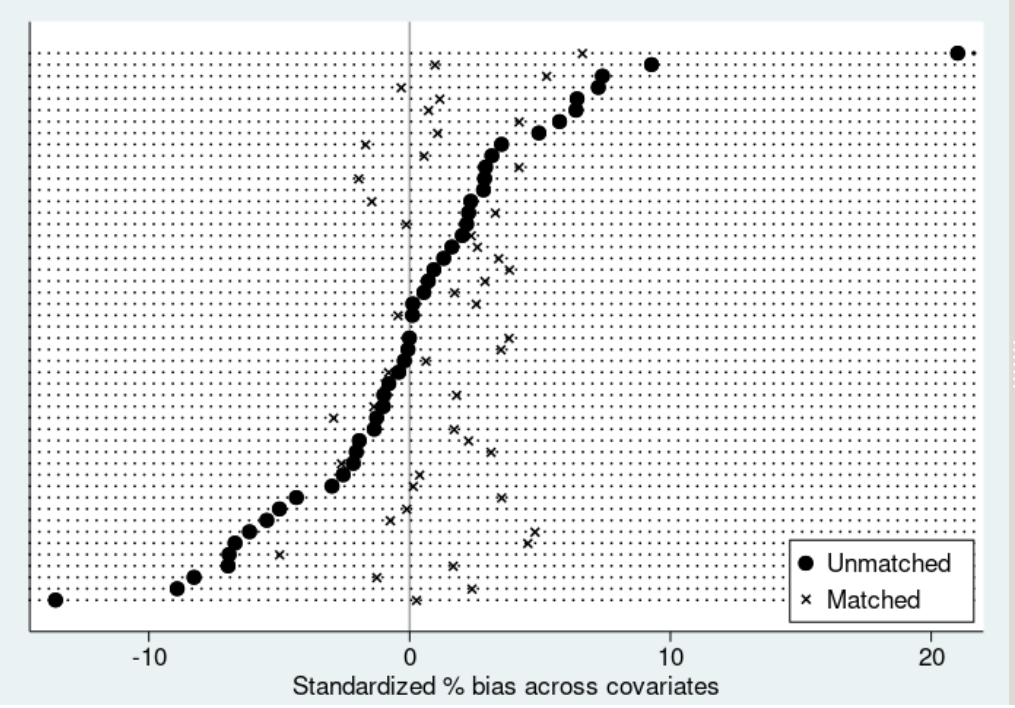


Figure H.3: Log-Log plot of the survival probability estimated by the GI bleeding Cox proportional hazard model over the analysis time to test if the proportional hazard assumption has been violated. Parallel graphs indicate no violation.

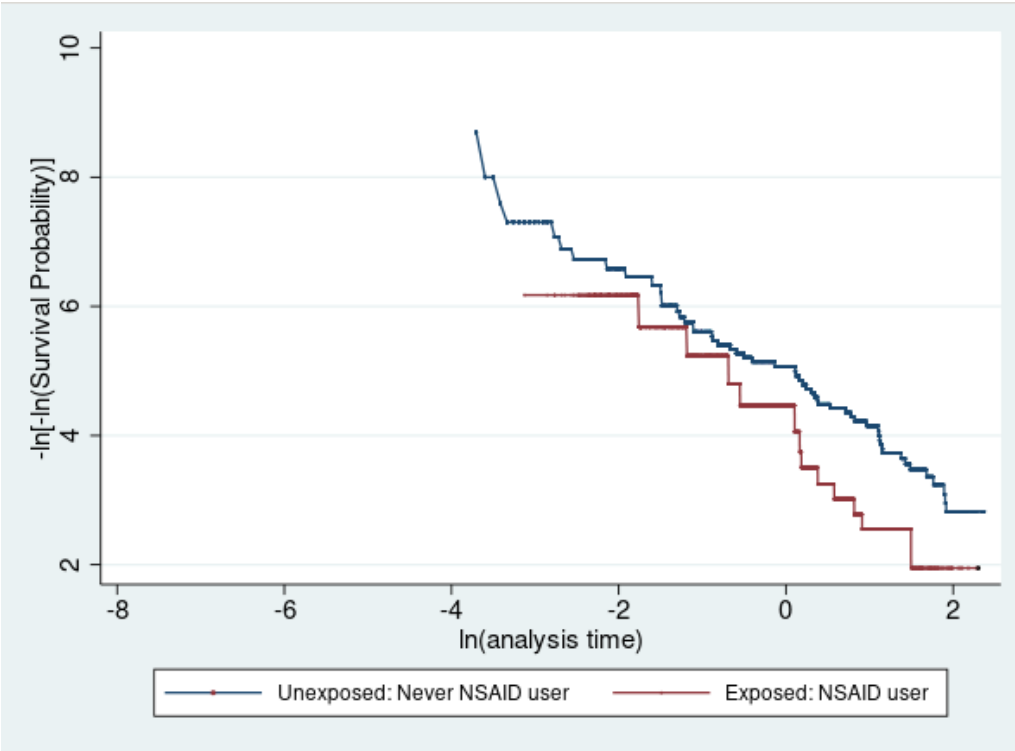


Figure H.4: Log-Log plot of the survival probability estimated by the GI bleeding Cox proportional hazard model over the analysis time to test if the proportional hazard assumption has been violated. Parallel graphs indicate no violation.

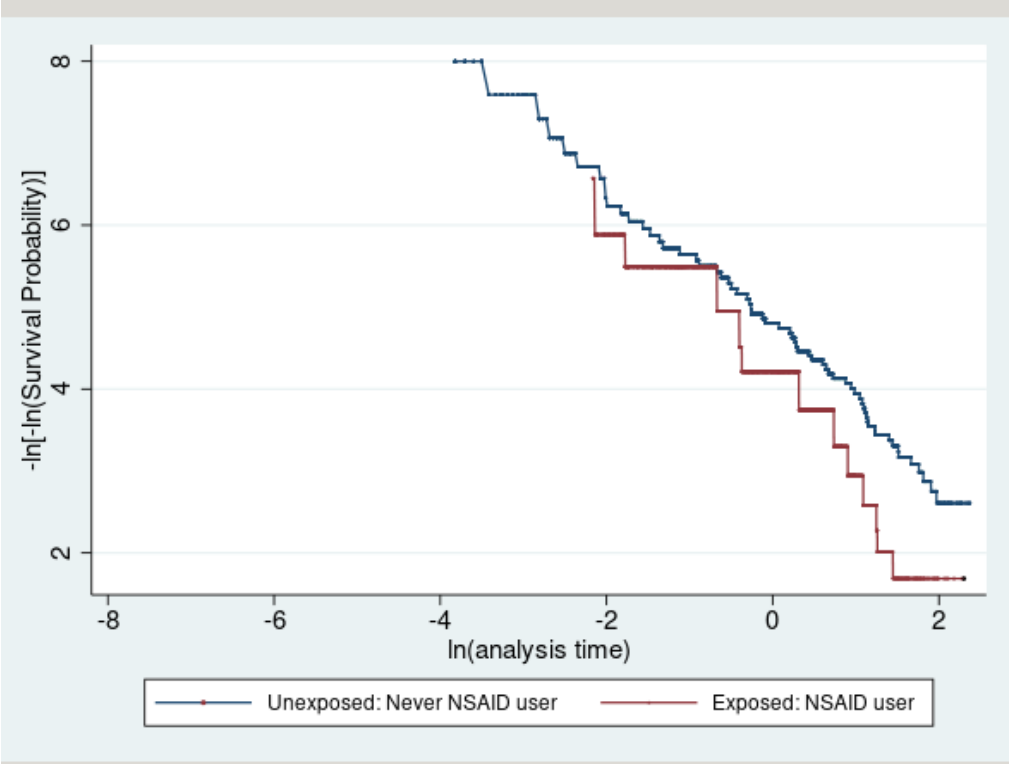


Table H.1: E-values for the HRs calculated for serious GI events and stroke compared with the impact of risk factors for each outcome conditional on NSAID exposure from Cox proportional hazard model in the matched cohort in the base case analysis

Serious GI events		Stroke	
E-value (mean, lower bound)	5.25, 2.58	E-value (mean, lower bound)	4.23, 2.06
Variable[a]	Impact on outcome, HR (95% CI)	Variable[b]	Impact on outcome, HR (95% CI)
Renal disease	0.52 (0.07; 3.73)	Heart failure	1.13 (0.62; 2.05)
Chronic liver disease	N/A	Stroke/Transient ischaemic attack	3.81 (2.42; 5.99)
Age	1.01 (0.99; 1.03)	Age	1.06 (1.03; 1.09)
Uncontrolled blood pressure (>160mmHg)	0.67 (0.16; 2.75)	Hypertension	2.77 (1.60; 4.82)
Alcohol dependence	1.31 (0.32; 5.35)	peripheral artery disease	2.24 (0.91; 5.56)
Bleeding	2.04 (1.18; 3.55)	Coronary heart disease	2.46 (1.57; 3.85)
Peptic ulcer	2.25 (0.97; 5.21)	Diabetes	1.17 (0.68; 1.04)
Anaemia	1.23 (0.61; 2.49)	Female	1.53 (0.98; 2.39)
Aspirin	1.30 (0.75; 2.23)		
Antiplatelet	1.38 (0.50; 3.8)		
Antidepressants (SSRI, TCA)	1.58 (0.91; 2.74)		
Corticosteroids	0.76 (0.30; 1.89)		

[a]Confounder identified by clinicians for serious GI events; [b]main risk factor for stroke from CHA2DS2-VASc

Appendix I – List of assumptions used in cohort analysis

Table I.1: List of assumptions made to prepare the electronic health record data, generate code lists, extract relevant records, prepare the population and extract information on baseline characteristics

Data preparation	
Identifying continuous medication use	
Days, where prescriptions of the same drug were overlapping, were not added to the end of the consecutive treatment	This did not yield reasonable results. Overlapping days were sometimes years in total for OAC patients. It was assumed that medication surplus through overlaps was covered by the grace period.
Code list generation	
Keywords identified for OACs	Warfarin Xarelto rivaroxaban Lixiana edoxaban apixaban Eliquis phenindione Pradaxa dabigatran acenocoumarol Dindevan Marevan
Keywords identified for NSAIDs	Aceclofenac acemetacin celecoxib dexibuprofen dexketoprofen diclofenac diflunisal etodolac Etoricoxib Fenbufen fluribuprofen Ketoprofen ibuprofen indomethacin meloxicam nabumetone naproxen phenylbutazone piroxicam sulindac tenoxicam tiaprofenic tolfenamic
Perforated ulcers were included as serious GI events in the ICD10 code list	NSAIDs can also cause perforations of ulcers. This assumption was supported by the 3 clinicians that validated the code list
Excluded from the ICD-10 code list were diagnoses on lower GI bleeding events.	Even though lower GI bleedings were associated with NSAID and OAC use, the clinicians recommended not to include them. While for the OAC/NSAID HPE an inclusion would have been reasonable according to the clinicians, they suggested to exclude them because they were less relevant for the other CV HPE. All other CV HPE were resolved by adding a GPA, which do not lower the risk of lower GI bleeding events, only upper GI bleeding events (617)
Cleaning of data on baseline characteristics	
BMI records	
BMI was used as recorded in the additional files and if missing calculated from height and weight values in entity 13 and 14.	This approach was chosen to minimise missingness. From height and weight values 124 new BMIs could be calculated.
Extreme BMI (>70), height (<1.2m or >2.4m) and weight (<30 or>300) values were dropped.	The applied limits were based on assumptions for adults in the UK. The limits were approved by a physician.
Smoking records	
Smoking status was categorised as never smoker, current smoker, ex-smoker or missing smoking status	Colleagues provided a list of medcodes on smoking categorised by smoking status that have been used previously (618)

To inform the smoking status at baseline records from CPRD therapy, referral, clinical and additional files were used.	All these flat files can potentially contain relevant information to assess smoking status. Therapy files were searched for prodcodes of smoking cessation therapy, such as nicotine patches. Clinical and additional files contained information on smoking status and cigarettes smoked from recorded medcodes and entities (see entity assumption), related to smoking
Entities 4 and 23 were considered relevant entities to identify smoking status	Entity describes what the observation in the additional file relates to. Entity 4 describes records of additional files on smoking status [data1] or cigarettes smoked per day [data2]. Entity 23 describes additional information on given life-style advice that can be associated with current or ex-smoker.
If duplicate smoking records in merged clinical and additional files on same day with different smoking status [ex-smoker and current smoker], the status recorded before this was used and if it was the first recorded smoking status the subsequent record was used. If smoking records existed only on one day that were different, it was kept if data1 records were in agreement with smoking status and if not, the patient was dropped	It was decided that it is not possible to identify the 'correct' record if multiple different records were identified for the same patient at the same day. The previous value was assumed to be more precise and if this was not available the subsequent record was seen to represent the status at that time the best. The data 1 record from the additional files provided further information. Data1=1 indicated a current smoker. Data3=3 indicated an ex-smoker. This was used to validate and identify the correct smoking status among the duplicates. If not available or data1 records and smoking status were inconsistent, no decision could be made and the patient had to be dropped
Never smoker records that were recorded after a patient had a current or ex-smoker record were replaced by ex-smoker records if the patient's additional information suggests the patients is or has been a smoker.	It was assumed that never smoking records that were recorded after a current or ex-smoker record, were potentially mis-specified. If the additional information [record of entity 23 together with smoking medcode, referral with smoking medcode, nicotine replacement therapy prodcode or entity 4 data1 entry that suggests the patient smokes more than 0 cigarettes per day] suggests the patient is/has been a smoker, the assumption that this patient is rather an ex-smoker than a never smoker was made.
Blood pressure records Uncontrolled blood pressure was identified by extracting all patients with blood pressure measurements recorded in the additional files that had a systolic blood pressure >160 mmHg [data2]	Entity 1 records were used to link the observations in the clinical files with the additional files. The HAS-BLED bleeding risk score identifies uncontrolled blood pressure [>160mmHg] as a key bleeding risk indicator. In the clinical validation, the GPs questioned agreed that 160mmHg was a good indicator and that additional files should be sufficient.

If multiple records existed that were recorded on the same day, the mean blood pressure was used.	It is not rare that multiple measurements are taken. Blood pressure measurements are very sensitive. They can be elevated if the patient just walked up the stairs or is nervous to see the doctor or nurse. Measurements can also vary due to measurement error. Multiple measurements give healthcare professionals the option to take a mean value that is more representative of the real blood pressure.
HES data cleaning	
Bleeding events were only extracted if recorded as primary diagnosis	It is common not to include subsequent diagnosis in HES records (354-356). Even though there is a chance the primary diagnosis does not identify all outcome events, the chance to include a false positive or incorrect outcome event when including subsequent diagnosis is higher. Including more not real cases, could dilute the effect.
The recorded episode start date was seen as the date the event happened	The episode start date was considered to be closest to the day the event occurred because it describes the day of admission due to this event. Episodes that started before the start of follow-up of the patient and end after start of follow-up were not considered relevant because these would not have been cause by the drugs (OAC/NSAID) under investigation
If multiple episode start dates are in the same spell with the same date, only the episode with the lowest episode identifier (epikey) was used, even if the episodes were from different spells	All events recorded at the same day were assumed to describe the same event and therefore only one event was kept. The first entry, hence the one with the lowest epikey, was assumed to be most accurate.
If multiple discharges with discharge method 'died', and no ONS/CPRD death record was available, the last entry was used.	ONS and CPRD records were considered more accurate. If none of these was available, it was assumed that if a patient was readmitted to hospital after discharged as dead, he did not die. Hence, the latest record was used. (8/117,563 OAC patients were affected)
Population	
Base line characteristics	
Covariate assessment at baseline for conditions ever before the index date	Diagnosis usually do not change over time. The DAG suggested that new conditions after exposure are not confounders. They either lie on the causal pathway between exposure and outcome or are not related to the exposure.
Covariate assessment at baseline for drugs in 6 months prior to index	Followed approach from other studies reporting characteristics from observational data of OAC cohorts that used 6 month as an assessment window (619-621). This was considered a reasonable time window to capture current use of drugs at baseline.

Information on ethnicity were extracted from CPRD and HES records	In order to identify the most appropriate ethnicity estimate in case of varying information in the dataset or between dataset, a flow chart developed by Alison Wright was followed (622)
BMI was assessed 12 months before and 12 months after index date	BMI measurements further away from the index date, were not considered relevant, as the weight might have changed. The BMI measurement closest to the index date was considered most appropriate. Missingness was high. That's why we also looked at BMI measurements 12 months after index date and considered this to be the baseline value.
Smoking status closest to index date was considered baseline smoking status	Smoking status before index date did not entail a smoking status record for 990 of 117,563 patients. Including records after baseline yielded smoking status records closer to the index date and only 440 patients with no record of smoking status. This approach was also used previously when bleeding risk in OAC patients and SSRI was analysed in CPRD (623). If the smoking status record before and after index date had the same distance to the index date, the smoking status of the earlier record was used [122 patients].
Blood pressure measurements were considered within 12 months of the index date	Blood pressure was considered, similar to the assumptions for BMI measurements, to vary with time. Measurements more than 12 months before or after the index date were not considered relevant to define baseline blood pressure measurements.

BMI: body mass index; CHA2DS2-VASc: stroke risk score for patients with atrial fibrillation; CPRD: Clinical Practice Research Datalink; HAS-BLED: bleeding risk score for patients using anticoagulants GI: gastro-intestinal; GPA: gastroprotective agent; NSAID: non-steroidal anti-inflammatory agent; OAC: oral anticoagulants; ONS: Office for National Statistics; SSRI: selective serotonin re-uptake inhibitor

Appendix J – ICD-10 code lists for outcomes in Chapter Four

This section contains ICD-10 codes used to identify outcomes. Read codes used to identify patients on treatment with oral anticoagulants and NSAIDs are not reported here but available from the authors on request. Read codes for baseline characteristics are also available on request.

Table J.1: ICD-10 codes used to identify outcomes in Hospital Episodes Statistics (HES)

ICD-10	Description	Type
I61	Intracerebral haemorrhage	Stroke, MB
I61.1	Intracerebral haemorrhage in hemisphere, cortical	Stroke, MB
I61.2	Intracerebral haemorrhage in hemisphere, unspecified	Stroke, MB
I61.3	Intracerebral haemorrhage in brain stem	Stroke, MB
I61.4	Intracerebral haemorrhage in cerebellum	Stroke, MB
I61.5	Intracerebral haemorrhage, intraventricular	Stroke, MB
I61.6	Intracerebral haemorrhage, multiple localized	Stroke, MB
I61.8	Other intracerebral haemorrhage	Stroke, MB
I61.9	Intracerebral haemorrhage, unspecified	Stroke, MB
I63	Cerebral infarction	Stroke
I63.0	Cerebral infarction due to thrombosis of precerebral arteries	Stroke
I63.1	Cerebral infarction due to embolism of precerebral arteries	Stroke
I63.2	Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries	Stroke
I63.3	Cerebral infarction due to thrombosis of cerebral arteries	Stroke
I63.4	Cerebral infarction due to embolism of cerebral arteries	Stroke
I63.5	Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries	Stroke
I63.8	Other cerebral infarction	Stroke
I63.9	Cerebral infarction, unspecified	Stroke
I64	Stroke, not specified as haemorrhage or infarction	Stroke
G45	Transient cerebral ischaemic attacks and related syndromes	Stroke
G45.0	Vertebro-basilar artery syndrome	Stroke
G45.1	Carotif artery syndrome (hemispheric)	Stroke
G45.2	Multiple and bilateral precerebral artery syndromes	Stroke
G45.3	Amaurosis fugax	Stroke
G45.8	Other transient cerebral ischaemic attacks and related syndromes	Stroke
G45.9	Transient cerebral ischaemic attack, unspecified	Stroke
G46	Vascular syndromes of brain in cerebrovascular diseases	Stroke
G46.0	Middle cerebral artery syndrome	Stroke
G46.1	Anterior cerebral artery syndrome	Stroke
G46.2	Posterior cerebral artery syndrome	Stroke
G46.3	Brain stem stroke syndrome	Stroke
G46.4	Cerebellar stroke syndrome	Stroke
G46.5	Pure motor lacunar syndrome	Stroke
G46.6	Pure sensory lacunar syndrome	Stroke
G46.7	Other lacunar syndromes	Stroke
G46.8	Other vascular syndromes of brain in cerebrovascular disease	Stroke
I26.0	Pulmonary embolism with acute cor pulmonale	SE
I26.9	Pulmonary embolism without acute cor pulmonale	SE
I74.0	Embolism and thrombosis of abdominal aorta	SE
I74.1	Embolism and thrombosis of other and unspecified parts of aorta	SE
I74.2	I74.2 Embolism and thrombosis of arteries of the upper extremities	SE
I74.3	I74.3 Embolism and thrombosis of arteries of the lower extremities	SE
I74.4	I74.4 Embolism and thrombosis of arteries of extremities, unspecified	SE
I74.5	I74.5 Embolism and thrombosis of iliac artery	SE
I74.8	I74.8 Embolism and thrombosis of other arteries	SE
I74.9	I74.9 Embolism and thrombosis of unspecified artery	SE

I85.0	Oesophageal varices with bleeding	GIB, MB
K22.8	Haemorrhage of oesophagus	GIB, MB
K25.0	Gastric ulcer, acute with haemorrhage	GIB, MB
K25.1	Gastric ulcer - Acute with perforation	GIB, MB
K25.2	Gastric ulcer, acute with both haemorrhage and perforation	GIB, MB
K25.4	Gastric ulcer, chronic or unspecified with haemorrhage	GIB, MB
K25.5	Duodenal ulcer - Chronic or unspecified with perforation	GIB, MB
K25.6	Chronic or unspecified with both haemorrhage and perforation	GIB, MB
K26.0	Duodenal ulcer, acute with haemorrhage	GIB, MB
K26.1	Duodenal ulcer - Acute with perforation	GIB, MB
K26.2	Duodenal ulcer, acute with both haemorrhage and perforation	GIB, MB
K26.4	Duodenal ulcer, chronic or unspecified with haemorrhage	GIB, MB
K26.5	Duodenal ulcer - Chronic or unspecified with perforation	GIB, MB
K26.6	Chronic or unspecified with both haemorrhage and perforation	GIB, MB
K27.0	Peptic ulcer, acute with haemorrhage	GIB, MB
K27.1	Peptic ulcer, site unspecified - Acute with perforation	GIB, MB
K27.2	Peptic ulcer, acute with both haemorrhage and perforation	GIB, MB
K27.4	Peptic ulcer, chronic or unspecified with haemorrhage	GIB, MB
K27.5	Peptic ulcer, site unspecified - Chronic or unspecified with perforation	GIB, MB
K27.6	Chronic or unspecified duodenal ulcer with both haemorrhage and perforation	GIB, MB
K28.0	Gastrojejunal ulcer, acute with haemorrhage	GIB, MB
K28.1	Gastrojejunal ulcer - Acute with perforation	GIB, MB
K28.2	Acute gastrojejunal ulcer with both haemorrhage and perforation	GIB, MB
K28.4	Gastrojejunal ulcer, chronic or unspecified with haemorrhage	GIB, MB
K28.5	Gastrojejunal ulcer - Chronic or unspecified with perforation	GIB, MB
K28.6	Chronic or unspecified ulcer with both haemorrhage and perforation	GIB, MB
K29.0	Acute haemorrhagic gastritis	GIB, MB
K66.1	Haemoperitoneum	GIB, MB
K92.0	Haematemesis	GIB, MB
K92.1	Melaena	GIB, MB
K92.2	Gastrointestinal bleed, unspecified	GIB, MB
J94.2	Haemopneumothorax	MB
H31.3	Choroidal haemorrhage and rupture	MB
H43.1	Vitreous haemorrhage	MB
H45.0	Vitreous haemorrhage in diseases classified elsewhere	MB
R04	haemorrhage from respiratory passages (epistaxis, throat, cough with haemorraghe)	MB
R04.1	Haemorrhage from throat	MB
R04.2	Haemoptysis	MB
R04.8	Haemorrhage from other sites in respiratory passages	MB
R04.9	Haemorrhage from respiratory passages, unspecified	MB
R31	Unspecified haematuria	MB
R58	Haemorrhage, not elsewhere classified	MB
M25.0	Haemarthrosis (bleeding into joint spaces)	MB
N02	Recurrent and persistent haematuria	MB
K62.5	Haemorrhage of anus and rectum	MB
K55.21	Angiodysplasia of colon with bleed	MB

GIB: gastro-intestinal bleeds including ulcer perforations referred to as serious GI events; MB: major bleeding events; SE: systemic embolism

Appendix K – Search terms and results of literature searches informing conceptualisation of the state-transition model in Chapter Five

Titles and abstracts of identified studies were screened, and full texts of potentially relevant records used to inform model structure or identify potential model parameters. Studies were considered preferable if they were large, UK-based/relevant, and recent (i.e., published since 2009). Table K.1 reports the search terms and Table K.2 the results of the searches.

Table K.1: Literature searches conducted

Searches	Databases	Search terms
State-transition models comparing NSAID use in anticoagulated patients	• Medline (1946-February Week 1 2021) and Embase (1974-10th February 2021) • HTA database	(((NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor) or (meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib)).ab. or Anti-Inflammatory Agents, Non-Steroidal/ or Cyclooxygenase 2 Inhibitors/ or Cyclooxygenase Inhibitors/ or Meloxicam/ or Naproxen/ or Ibuprofen/ or Diclofenac/ or Mefenamic acid/ or Celecoxib/ or Etoricoxib/) AND (((eliquis or pradaxa or lixiana or xarelto) or (doac or NOAC or oral anticoagulant or atrial fibrillation) or (rivaroxaban or dabigatran or apixaban or edoxaban or warfarin)).ab. or Factor Xa Inhibitors/ or Atrial Fribrillation/ or Anticoagulants/ or anticoagulant agent/ or rivaroxaban/ or dabigatran/ or warfarin/) AND ((Modeling or modelling or Model or Markov or transition or markov chains). Mp) AND ((Economics, Pharmaceutical/ or Costs and cost analysis/ or budgets /or cost-benefit analysis/ or (cost or pharmacoeconomic or value for money or budget or cost-effectiveness or cost-effectiveness analysis).mp)) HTA database: ((rivaroxaban or dabigatran or apixaban or edoxaban or warfarin) OR (doac or NOAC or oral anticoagulant or atrial fibrillation)) AND ((meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib) OR (NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor))
State-transition models of NSAIDs	• Medline, Embase Jan 2008 to Jan 2020	(Modeling or modelling or Model or Markov or transition or markov chains). Mp AND (Economics, Pharmaceutical/ or Costs and cost analysis/ or budgets /or cost-benefit analysis/ or (cost or pharmacoeconomic or value for money or budget or cost-effectiveness or cost-effectiveness analysis).mp) AND

	• HTA database		AND ((meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib) OR (NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor)) HTA database: (meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib) OR (NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor)
State-transition models of OACs	•	Medline (1946-Dec week 4 2018) and Embase (1974-3rd Jan 2019)	((Modeling or modelling or Model or Markov or transition).mp or markov chains) AND (Economics, Pharmaceutical/ or Costs and cost analysis/ or budgets /or cost-benefit analysis/ or (cost or pharmacoeconomic or value for money or budget or cost-effectiveness or cost-effectiveness analysis).mp) AND (Atrial Fibrillation/ or Anticoagulants/ or Warfarin/ or Factor Xa Inhibitors/ or rivaroxaban/ or dabigatran/ (oral anticoagulant or doac or NOAC or edoxaban or rivaroxaban or apixaban or dabigatran or Eliquis or Pradaxa or Xarelto or lixiana).mp
Risks (impact of NSAIDs in OAC patients-GI events)	•	Medline (1946-26th November) and Embase (1974-26th November)	(((NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor) or (meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib)).ab. or Anti-Inflammatory Agents, Non-Steroidal/ or Cyclooxygenase 2 Inhibitors/ or Cyclooxygenase Inhibitors/ or Meloxicam/ or Naproxen/ or Ibuprofen/ or Diclofenac/ or Mefenamic acid/ or Celecoxib/ or Etoricoxib/) AND (((eliquis or pradaxa or lixiana or xarelto) or (doac or NOAC or oral anticoagulant or atrial fibrillation) or (rivaroxaban or dabigatran or apixaban or edoxaban or warfarin)).ab. or Factor Xa Inhibitors/ or Atrial Fribrillation/ or Anticoagulants/ or anticoagulant agent/ or rivaroxaban/ or dabigatran/ or warfarin/) AND ((((gastrointestinal or gi) and (symptom or diseases or disorder or adverse or diagnosis)) or (dyspepsia or diarrhoea or flatulence or vomiting or nausea or abdominal pain or epigastric pain)).ab. OR (Gastrointestinal diseases/ or Dyspepsia/ or Abdominal pain/ or gastrointestinal symptoms/ or epigastric pain/) OR (((peptic or digestive system or duodenum or gastric or stomach or esophagus or jejunum or colon or symptomatic) and (ulcer)).ab. or peptic ulcer bleeding/ or ulcer incidence/ or digestive system ulcer/ or duodenum ulcer/ or acetic acid-induced gastric ulcer/ or gastric ulcer bleeding/ or stomach ulcer/ or duodenal ulcer bleeding/ or esophagus ulcer hemorrhage/ or esophagus ulcer/ or jejunum ulcer/ or ulcer perforation/ or peptic ulcer/ or recurrent peptic ulcer/ or indomethacin-induced gastric ulcer/ or colon ulcer/ or ulcer/) OR ((death OR died OR fatal OR mortality OR deceased OR non survivor OR non Survival).ab. or Death/ or Hospital Mortality/ or Mortality/ or
	• HTA database*		

		Fatal Outcome/ or Survivors/ or Survivor/ or Survival Rate/ or Survival/ or mortality rate/ or standardized mortality ratio/))) HTA database: ((rivaroxaban or dabigatran or apixaban or edoxaban or warfarin) OR (doac or NOAC or oral anticoagulant or atrial fibrillation)) AND ((meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib) OR (NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor))
Risks (impact of NSAIDs in OAC patients - cardiovascular events)	Medline (1946-week 3 Feb 2020) Embase (Embase 1974-week 3 February 2020)	(((NSAID or anti-inflammatory drug or coxibe or cyclooxygenase inhibitor or cyclooxygenase 2 inhibitor) or (meloxicam or naproxen or ibuprofen or diclofenac or mefenamic or celecoxib or etoricoxib)).ab. or Anti-Inflammatory Agents, Non-Steroidal/ Cyclooxygenase 2 Inhibitors/ or Cyclooxygenase Inhibitors/ Meloxicam/ or Naproxen/ or Ibuprofen/ or Diclofenac/ or Mefenamic acid/ or Celecoxib/ or Etoricoxib/) AND (Stroke or myocardial or heart failure or cardiovascular or systemic embolism or thromboembolic).ab AND (Atrial Fibrillation/ or Anticoagulants/ or Warfarin/ or Factor Xa Inhibitors/ or rivaroxaban/ or dabigatran/ (oral anticoagulant or doac or NOAC or edoxaban or rivaroxaban or apixaban or dabigatran or Eliquis or Pradaxa or Xarelto or lixiana).mp
Utility values	• Medline Embase (2011-November 2020) • Systematic review until 2011 in TA249(463) and TA275 (491) and in PINCER economic evaluation review of utilities until 2010 (149)	((EQ-5D*) or (SF-12) or (QALY*) or (quality-adjusted*) or (qol) or quality of life) or (HRQoL)) .ab. AND(((((gastrointestinal or gi) and (symptom or diseases or disorder or adverse or diagnosis)) or (dyspepsia or diarrhoea or flatulence or vomiting or nausea or abdominal pain or epigastric pain)).ab. or Gastrointestinal diseases/ or Dyspepsia/ or Abdominal pain/ or gastrointestinal symptoms/ or epigastric pain/) OR (((peptic or digestive system or duodenum or gastric or stomach or esophagus or jejunum or colon or symptomatic) and (ulcer)).ab. or peptic ulcer bleeding/ or ulcer incidence/ or digestive system ulcer/ or duodenum ulcer/ or acetic acid-induced gastric ulcer/ or gastric ulcer bleeding/ or stomach ulcer/ or duodenal ulcer bleeding/ or esophagus ulcer hemorrhage/ or esophagus ulcer/ or jejunum ulcer/ or ulcer perforation/ or peptic ulcer/ or recurrent peptic ulcer/ or indomethacin-induced gastric ulcer/ or colon ulcer/ or ulcer/) OR (stroke.mp. or Stroke/) OR (Gastrointestinal Hemorrhage/ or gastro intestinal bleeding.mp. or Peptic Ulcer Hemorrhage/ or bleeding.mp. or Hemorrhage/ or Peptic Ulcer Hemorrhage/ Or Duodenal Ulcer/ Or Peptic Ulcer/ Or Peptic Ulcer Perforation/ Or Ulcer/ Or Stomach Ulcer/ Or Ulcer.mp.)

*HTA = health technology assessment; database updated until 31st March 2018, therefore this search was limited to that date

Table K.2. Results of literature searches conducted to inform the state-transition model

Searches	Identified studies[a]	Relevant studies identified
State-transition models on NSAID use in OAC population	46	No decision-analytic models were identified that explored the effect of NSAIDs in an OAC population [low dose aspirin was not considered relevant as an NSAID]
State-transition models of NSAIDs	70 (Medline and Embase), hand search	Various models identified with 4 unique model structures (149, 151, 227, 450-454, 456-459, 492): (1) de Groots model (151, 452, 453), (2) NICE model developed by Latimer et al. (2009) (450, 451, 454, 456-458, 461), (3) the PINCER model (149, 227, 492, 624) and (4) a model structure by Moriarty et al. (2019) (151)
State-transition models of OACs	113 studies (Medline and Embase and HTA after duplicates were removed)	In 13 systematic reviews identified 20 models were reported that were relevant for OAC populations in AF [13] and VTE [7]. 8 unique model structure were identified. Main models identified that were repeatedly used were a state-transition model by Sorensen et al. (2009) used in HTA 249 (351, 463-471) and the Dorian model used in TA 275 (472) and multiple other studies (473-480) and a structure by Gage from 1995 (481) that was adapted by Lee et al. (2012) (482-484). Five model structures were identified that used a different base model (485-488).
Risks (impact of NSAIDs in OAC patients, serious GI events)	717 studies (Medline, Embase, HTA database)	Studies were included if they investigated the effect of NSAIDs on peptic ulcer, adverse GI effects or mortality in patients on OAC treatment. Excluded were studies solely looking at GI bleeding or cardiovascular events because these were searched for separately. Reviews identified were screened for potentially relevant studies missed in the literature search. The systematic review identified 701 studies from Embase and Medline on 16th July 2020 after duplicates were removed. Abstracts of the articles were screened and 41 studies were selected for a full text review. None of the studies provided estimates in the specified patient group. There were no studies assessing risk of peptic ulcers, GI adverse effects or mortality associated with NSAID use in patients with OAC treatment.
Risks (impact of NSAIDs in OAC patients, cardiovascular events)	1410 studies (Embase, Medline, HTA after duplicates were removed), 2 from hand search	4 studies were identified that analysed the effect of NSAIDs on cardiovascular risk in patients with OACs (307-309, 338)
Utility values	8357 studies (Embase, Medline after duplicates were removed)	For each health state the results of this review were discussed in the 'utility' section of the input parameters if they fulfilled the requirements.

[a]Abstracts of identified studies were screened

Appendix L – AdViSHE (Assessment of the Validation Status of Health-Economic decision models)

AdViSHE is a questionnaire that modellers can complete to report on the efforts performed to improve the validation status of their health-economic decision model. It is not intended to replace validation by model users but rather to inform the direction of validation efforts and to provide a baseline for replication of the results. In addition to using it after a model is finished, the modellers can use AdViSHE to guide validation efforts during the modelling process.

The modellers are asked to comment on the validation efforts performed while building the underlying HE decision model and afterwards. Many of the questions simply refer to the model documentation. AdViSHE is divided into five parts, each covering an aspect of validation:

- Part A: Validation of the conceptual model (2 questions)
- Part B: Input data validation (2 questions)
- Part C: Validation of the computerized model (4 questions)
- Part D: Operational validation (4 questions)
- Part E: Other validation techniques (1 question)

No final validation score is calculated, as the assessment of the answers and the overall validation effort is left to the model users. It is assumed that the model has been built according to prevailing modelling and reporting guidelines. Some questions may not be applicable to a particular model. If this is the case, the model builder should take the opt-out option and provide a justification of why this item is not deemed applicable. Results reported in Table L.1.

Table L.1: Author responses to the Advishe Model validation tool

	AdViSHE tool	Author response/section reference
	Part A: Validation of the conceptual model (2 questions) Part A discusses techniques for validating the conceptual model. A conceptual model describes the underlying system (e.g., progression of disease) using a mathematical, logical, verbal, or graphical representation. Please indicate where the conceptual model and its underlying assumptions are described and justified.	5.2.1; 5.2.2
A1	Face validity testing (conceptual model): Have experts been asked to judge the appropriateness of the conceptual model? If yes, please provide information on the following aspects: - Who are these experts? - What is your justification for considering them experts? - To what extent do they agree that the conceptual model is appropriate? If no, please indicate why not. Aspects to judge include: appropriateness to represent the underlying clinical process/disease (disease stages, physiological processes, etc.); and appropriateness for economic evaluation (comparators, perspective, costs covered, etc.).	5.2.1: Expert consultation
A2	Cross validity testing (conceptual model): Has this model been compared to other conceptual models found in the literature or clinical textbooks? If yes, please indicate where this comparison is reported. If no, please indicate why not.	5.2.1
	Part B: Input data validation (2 questions) Part B discusses techniques to validate the data serving as input in the model. These techniques are applicable to all types of models commonly used in Health economic modelling. Please indicate where the description and justification of the following aspects are given: - search strategy; - data sources, including descriptive statistics; - reasons for inclusion of these data sources; - reasons for exclusion of other available data sources; - assumptions that have been made to assign values to parameters for which no data was available; - distributions and parameters to represent uncertainty; - data adjustments: mathematical transformations (e.g., logarithms, squares); treatment of outliers; treatment of missing data; data synthesis (indirect treatment comparison, network meta-analysis); calibration; etc.	Details regarding search strategy; data sources and assumptions that have been made to assign values to parameters for which no data was available have been provided in the appendix. Distributions and parameters to represent uncertainty are described in 3.3.2; 5.2.4; 6.3.1. The other points on data adjustments were not applicable.
B1	Face validity testing (input data): Have experts been asked to judge the appropriateness of the input data? If yes, please provide information on the following aspects: - Who are these experts? - What is your justification for considering them experts?	5.2.1: Expert consultation; the experts agreed that chosen input

	- To what extent do they agree that appropriate data has been used? If no, please indicate why not. Aspects to judge may include but are not limited to: potential for bias; generalizability to the target population; availability of alternative data sources; any adjustments made to the data.	parameters represent the best available data
B2	Model fit testing: When input parameters are based on regression models, have statistical tests been performed? If yes, please indicate where the description, the justification and the outcomes of these tests are reported. If no, please indicate why not Examples of regression models include but are not limited to: disease progression based on survival curves; risk profiles using regression analysis on a cohort; local cost estimates based on multi-level models; meta-regression; quality-of-life weights estimated using discrete choice analysis; mapping of disease-specific quality-of-life weights to utility values. Examples of tests include but are not limited to: comparing model fit parameters (R2, AIC, BIC); comparing alternative model specifications (covariates, distributional assumptions); comparing alternative distributions for survival curves (Weibull, lognormal, logit); testing the numerical stability of the outcomes (sufficient number of iterations); testing the convergence of the regression model; visually testing model fit and/or regression residuals.	Various test were conducted to generate the hazard ratios for the increased likelihood of stroke and serious GI events with NSAID treatment 4.4.3; Appendix H
	Part C: Validation of the computerized model (4 questions) Part C discusses various techniques for validating the model as it is implemented in a software program. If there are any differences between the conceptual model (Part A) and the final computerized model, please indicate where these differences are reported and justified.	The mathematical model is as per the conceptual model.
C1	External review: Has the computerized model been examined by modelling experts? If yes, please provide information on the following aspects: - Who are these experts? - What is your justification for considering them experts? - Can these experts be qualified as independent? - Please indicate where the results of this review are reported, including a discussion of any unresolved issues. If no, please indicate why not. Aspects to judge may include but are not limited to: absence of apparent bugs; logical code structure optimized for speed and accuracy; appropriate translation of the conceptual model.	Appendix A
C2	Extreme value testing: Has the model been run for specific, extreme sets of parameter values in order to detect any coding errors? If yes, please indicate where these tests and their outcomes are reported. If no, please indicate why not. Examples include but are not limited to: zero and extremely high (background) mortality; extremely beneficial, extremely detrimental, or no treatment effect; zero or extremely high treatment or healthcare costs.	Yes, extreme value testing has been done and have been reported in this supplement

C3	Testing of traces: Have patients been tracked through the model to determine whether its logic is correct? If yes, please indicate where these tests and their outcomes are reported. If no, please indicate why not. In cohort models, this would involve listing the number of patients in each disease stage at one, several, or all time points (e.g., Markov traces). In individual patient simulation models, this would involve following several patients throughout their natural disease progression.	Yes, the number of patients in the death state were tracked through the model and logic has been found to be correct. In TreeAge Pro Healthcare 2021, the number of patients in each health state can be tracked in the individual Markov models at each cycle. The first 10 cycles were checked and were found to show anticipated changes.
C4	Unit testing: Have individual sub-modules of the computerized model been tested? If yes, please provide information on the following aspects: - Was a protocol that describes the tests, criteria, and acceptance norms defined beforehand? - Please indicate where these tests and their outcomes are reported. If no, please indicate why not. Examples include but are not limited to: turning sub-modules of the program on and off; altering global parameters; testing messages (e.g., warning against illegal or illogical inputs), drop-down menus, named areas, switches, labelling, formulas and macros; removing redundant elements.	In TreeAge Pro Healthcare 2021 different methods for discounting were tested [global discounting function imbedded in the software and manual discounting of cost and outcomes in each health state]. They all produced the same estimates with and without discounting.
	Part D: Operational validation (4 questions) Part D discusses techniques used to validate the model outcomes.	
D1	Face validity testing (model outcomes): Have experts been asked to judge the appropriateness of the model outcomes? If yes, please provide information on the following aspects: - Who are these experts? - What is your justification for considering them experts? - To what extent did they conclude that the model outcomes are reasonable? If no, please indicate why not. Outcomes may include but are not limited to: (quality-adjusted) life years; deaths; hospitalizations; total costs.	The final model and the results were reported to the PROTECT team [5.2.1: Expert consultation], the results were also presented at a conference [PRIMM 2021: Manchester] to health policy researchers and experts in patient safety. The estimated economic burden was found to be reasonable.
D2	Cross validation testing (model outcomes): Have the model outcomes been compared to the outcomes of other models that address similar problems? If yes, please provide information on the following aspects:	5.4.2

	- Are these comparisons based on published outcomes only, or did you have access to the alternative model? - Can the differences in outcomes between your model and other models be explained? - Please indicate where this comparison is reported, including a discussion of the comparability with your model. If no, please indicate why not. Other models may include models that describe the same disease, the same intervention, and/or the same population.	
D3	Validation against outcomes using alternative input data: Have the model outcomes been compared to the outcomes obtained when using alternative input data? If yes, please indicate where these tests and their outcomes are reported. If no, please indicate why not. Alternative input data can be obtained by using different literature sources or datasets, but can also be constructed by splitting the original dataset in two parts, and using one part to calculate the model outcomes and the other part to validate against.	5.3.2; 6.2.6
D4	Validation against empirical data: Have the model outcomes been compared to empirical data? If yes, please provide information on the following aspects: - Are these comparisons based on summary statistics, or patient-level datasets? - Have you been able to explain any difference between the model outcomes and empirical data? - Please indicate where this comparison is reported. If no, please indicate why not.	No, not applicable.
D4A	Comparison against the data sources on which the model is based (dependent validation).	No, not applicable.
D4B	Comparison against a data source that was not used to build the model (independent validation).	No, not applicable.
	Part E: Other validation techniques (1 question)	
E1	Other validation techniques: Have any other validation techniques been performed? If yes, indicate where the application and outcomes are reported, or else provide a short summary here. Examples of other validation techniques: structured 'walk-throughs' (guiding others through the conceptual model or computerized program step-by-step); naïve benchmarking ('back-of-the-envelope' calculations); heterogeneity tests; double programming (two model developers program components independently and/or the model is programmed in two different software packages to determine if the same results are obtained).	Conducted as part of internal validation

Appendix M – Key structural assumptions and validation of assumptions by clinicians for the state-transition model in Chapter Five

This appendix provides details on model assumptions [Table M.1], reports feedback from clinicians on how the HPE is resolved in practice [Table M.2] and on general model assumptions [Table M.3]

Table 1: Summary of key model assumptions

Model assumption	Justification	Approved by clinicians*
Only one type of event can occur per three-months cycle.	In a state-transition model some form of simplification has to be done. A three-monthss cycle was considered appropriate and has been used previously in NSAID (45, 452) and AF state-transition model for NICE TAs (463, 491, 524)	Yes
All patients experiencing an adverse event stay on the OAC therapy they used prior to the event	Guidelines recommend continuing OAC treatment after stroke or serious GI events (41, 625)	Yes
No long-term effects of GI adverse events expected on cost, utilities and transition probabilities expected. All patients not experiencing a recurrent event or die move to the post-GI event state	There is no sufficient data to support the assumptions that there is a prolonged effect of the adverse GI events after the cycle in which the event occurred. The impact was always highest in the first… months after the event and considered negligible thereafter (this was validated by clinicians in the team)	Yes
NSAID is removed after any of the adverse events.	Clinicians suggested that each of the adverse events, would trigger a review of the medication and discontinuation of the NSAID.	Yes [Table 2, Table 3]
No disutility associated with removing NSAID	It was assumed that alternative treatments, such as paracetamol have the same utility because there was no appropriate data to suggest otherwise. One model that included a disutility with paracetamol use conducted a meta-analysis of arthritis index in Ontario (492). However, it was not clear how this was achieved.	Yes [Table 3]
OAC is not discontinued or interrupted after any of the adverse events	Cost of interruptions of OAC treatment were considered negligible and is recommended for 7 days. Discontinuation rates associated with GI bleeding were very diverse (626-630). It was assumed that discontinuation rates were the same in the HPE and non HPE model and were therefore not considered to impact the overall results.	N/A
Only one recurrence per adverse GI event after initial event	Observational data on serious GI events and that the majority of recurrences occur in the first 3 months (330). It was assumed that this was the same for the other adverse GI events	Yes [Table 3]

*Specific questions are reported in Table M.2 and Table M.3; HPE: hazardous prescribing event; NSAID: non-steroidal anti-inflammatory drug; OAC: oral anticoagulant; TA: NICE technology appraisal

Table M.2: Feedback on how the HPE is resolved in practice from clinicians [GPs: TA, BG; pharmacists: TD, GG

Assumption	Reviewer initials	Agree (Y/N)	Comment
Only options to correct HPE are removing NSAID or OAC	TD	Y	
	TA	Y	
	BG	N	Adding gastroprotection is an option but not preferred.
	GG	N	In some cases, after discussion with the patient, we have not been able to stop the NSAID but have lowered the dose and added a decent dose of GPA. Agree not ideal outcome and rare cases.
Removing NSAID is the is the preferred choice after flagged up by intervention	TD	Y	
	TA	Y	Strong preference
	BG	Y	NSAID will rarely be indicated, so should usually be removed. Sometimes it turns out that the OAC isn't indicated (although I'd guess in SMASH most just came to the end of a short treatment course). Gastroprotection is the best you can do if you have no choice but is not first choice.
	GG	Y	
Removing NSAID is the preferred choice after serious harm outcome occurred	TD	y	
	TA	Y	
	BG	Y	Same logic.
	GG	Y	

HPE: hazardous prescribing event; NSAID: non-steroidal anti-inflammatory drug; OAC: oral anticoagulant

Table M.3: Expert consultation with GPs (TA, BG)

Assumptions	GP1	GP2
The model looks at the query with patients with oral anticoagulation and the hazardous prescription of a NSAID. Do you think that the following health states cover the potential adverse events associated with NSAID use in this anticoagulated population: Serious GI events (GI bleeds, ulcer bleeding, ulcer perforation), symptomatic ulcer (non-bleeding ulcer with symptoms), GI discomfort (abdominal pain, dyspepsia etc.), stroke? The risk of MI was not found to be increased in OAC patients taking NSAIDs compared to non-NSAID patients with OAC (307, 308).	That looks fine to me, and interesting that you have the evidence around risk of MI not being increased (presumably because the OAC reduces risk of coronary thrombosis)	Ditto
Do you think that after three months GI adverse events (GI discomfort, symptomatic ulcer and serious GI events) on average do not have an impact on healthcare resource use, utilities or mortality rates anymore? Or that if no recurrence occurs, the impact on quality of life, costs or mortality diminishes after 3 months?	I think that 3 months is reasonable for the length of the effects of GI averse events.	Yes. Simple GI discomfort probably lasts less long, but I assume that 3 months is your cycle time? If not, then one month for that. A few people with really serious GI events will have longer term harm (massive bleed, in ITU, long hospital stay, long rehab) but rare (and I've no idea *how* rare…)
We could not find any studies looking at the likelihood of recurring symptomatic ulcer. For serious GI events (bleeding and perforations) a systematic review and meta-analysis from 2019 reported a 90-day rebleeding probability of 10.1% for anticoagulated patients. Is this a reasonable recurrence risk for symptomatic ulcer?	Based on clinical impression that would seem a reasonable and plausible risk of recurrence.	Ditto. I have no quantitative idea, but feels not unreasonable.
There is some evidence in existing literature which suggests that people are at a higher risk of a second GI bleed (health state: serious GI event) in first month following the event. In the model, we are assuming that there is only a higher risk of a recurrent (or an unresolved) serious GI event in the first 3 months after the event (as this is the cycle length used in the model), after which the risk decreases. Is this a reasonable assumption?	Sounds reasonable.	Sounds reasonable
In the model we assume that the NSAID is stopped after GI discomfort. Is it reasonable to assume that the NSAID prescription is stopped after patient is diagnosed with GI discomfort (diarrhoea, abdominal pain, dyspepsia, nausea, vomiting, flatulence)?	Yes	Yes

In the model we assume that the NSAID is stopped after symptomatic ulcer. Is it reasonable to assume that the NSAID prescription is stopped after patient is diagnosed with symptomatic ulcer (People have a clinical diagnosis of a peptic ulcer following adverse GI symptoms, which may include endoscopic confirmation)?	Yes	Yes
In the model we assume that the NSAID is stopped after stroke. Is it reasonable to assume that the NSAID prescription is stopped after patient is diagnosed with stroke, which is both associated with the NSAID use?	This is less certain as some doctors may not be aware of the association. It would be a complete guess if I were to suggest what proportion of patients would have their NSAID stopped. If you want me to guess, I would say 50%, but do you have anything in CPRD that might help?	I agree. I don't think I would naturally assume that stroke or systemic embolism was associated with the NSAID unless it was a haemorrhagic stroke (although I might well take the opportunity to stop the NSAID since I really don't like the combination). Some data would be better, but 50% is a reasonable guess.
Do you think GI discomfort has an impact on mortality? In the model the same mortality is used as for the general OAC population.	On its own I would not expect this to have an impact on mortality.	Agree.
If event specific mortality rates were not available by age group, it was assumed that the mortality increased with age as the mortality of the general population would change. Is this a reasonable approach to adjust death rates to age by assuming the risk of death increases with increasing age as it would in the general population?	Took me a bit of time for me to get my head round this when you presented this to the team, but I now agree with the approach.	Agree (assuming that you are describing GI bleeding etc mortality not competing risk mortality). It's definitely wrong to assume a flat distribution of mortality by age, so adjusting for general population distribution is arbitrary but reasonable. What about sex though?!? Not saying you should adjust, but if you did or are, then I don't think the

According to the NHS patient information website there are 3 different types of test for H. Pylori (blood, stool, breath). Which is most commonly used in routine practice?	Very unlikely to be breath routinely (trip to a hospital). In the two health boards I've worked in, it's stool (blood is less good for reinfection).	Most often I think GPs do serology (detects previous/current infection) as it is easy to do. NICE Clinical Knowledge Service advice is here (favouring breath/stool for initial detection):https://cks.nice.org.uk/dyspepsia-unidentifiedcause#!scenario As GP1 says, breath testing is not easily available from general practice.
In some work conducted by Rachel in the early 2000s the resource use associated with GI discomfort was described. I have combined this with current NICE guidance for the model. Does the following reflect current treatment of GI discomfort in your experience? a) GP visit, no secondary care investigation b) Prescription of lansoprazole	Yes. Prescription of PPI rather than lansoprazole (although that's very common, probably the most common).	Agree: PPI, which is most commonly lansoprazole or omeprazole
The same work also looked at the resource use associated with symptomatic ulcer. Does the following reflect current treatment of symptomatic ulcers in your experience? a) 1x diagnostic endoscopy, 1x therapeutic endoscopy, 2x GP visits, 2x outpatient visits, 1x H. pylori test b) Prescription of lansoprazole	Don't think a therapeutic endoscopy is normal. Doubt most would get two outpatient visits. Locally GP visits + single endoscopy and treatment recommendation with no follow up would be pretty normal. Gastric ulcer is the only one that gets follow up to confirm healing.	Agree.

CPRD: Clinical Practice Research Datalink; GP: general practitioner; HPE: hazardous prescribing event; MI: myocardial infarction; NSAID: non-steroidal anti-inflammatory drug; OAC: oral anticoagulant; PPI: proton pump inhibitor

Appendix N – Internal validation of the state-transition model in Chapter Five

Table N.1: Internal model validation by EC according to a predefined checklist used in PROTECT

Item	Error checks	Works (yes/no)?	Comments validator [EC]	Comments author [LP]
Model structure/parameters				
1.	Is the model structure appropriate? Are all appropriate transitions included? Is 'dead' an absorbing state? Is it possible to die from all appropriate states?	Yes		
2.	Check that the cohort numbers & the sum of all the health state transition probabilities add to 1 in all cycles / across the decision tree.	Yes		
3.	Were half cycle corrections applied consistently for all transitions?	Yes		
4.	Have subgroup specific parameters e.g., mortality based on age or disease severity etc been applied?	Yes	You will need to include more age groups for the sensitivity analysis based on a 20-year time horizon – you may find it easier to use Tables (rather than a list) within TreeAge Pro Healthcare 2021 for this	More subgroups included [start age: 80; 5- and 10-year time horizon
5.	Check for unrealistic results (negative costs, more events than possible …)	Yes		
Extreme/alternative values				
6.	Set all utilities to 0. The total QALY gain should be 0.	Yes		
7.	Set all costs to 0. The total costs should be 0.	Yes		
8.	Change the time horizon of the model and check that the outputs/results change accordingly.	Yes		
Discounting				
9.	Confirm that a discount rate of 3.5% has been applied for costs and QALYs		See comments in results section – need to check discounting [discounting produced lower estimates than discounted]	Checked discounting using the global function inbuilt into TreeAge Pro Healthcare 2021 and the manual function. Both

				produce the same results. Without discounting outcomes are smaller
10.	Set the discount rates to 0. The undiscounted costs and QALYs should equal the discounted costs and QALYs.			
External validity				
11.	Are the utilities in the model reasonable compared to patients in the general population of the same age? Check (table below, otherwise here):	Yes		
Probabilistic Sensitivity Analysis				
12.	Are all parameters assigned an appropriate distribution for probabilistic sensitivity analyses?	Yes		
13.	Check PSA output mean costs, QALYS and ICER compared to the deterministic results. If there is a large discrepancy, is this due to the nature of the model, or errors in the PSA?		See comment above – please check that you have sent correct PSA model file	Sent updated files; Deterministic and probabilistic analyses produce very similar results with a 10-year time horizon. The results differ in the lifetime horizon analysis. This can be explained by the lognormal distribution applied for the increased risk ratios for the adverse events. The mean of the skewed lognormal distribution is higher than the mean used in the deterministic model. The difference between the two is more evident over time.