Application of propensity scores and marginal structural models evaluating the effect of allopurinol in gout using primary care medical records

Trishna Rathod-Mistry

A thesis submitted for the degree of Doctor of Philosophy

June 2021

Keele University

Abstract

Background

Primary care electronic health records (EHR) capture real life patterns of healthcare utilisation over time. This provides the opportunity to estimate the effect of allopurinol on long term outcomes in people with gout. However, use of such data gives rise to confounding by indication which may change over time, a major impediment in treatment effect estimation.

Methods

A cohort of patients consulting for gout between 1997-2002 and not previously prescribed urate-lowering drugs were identified from the Clinical Practice Research Datalink GOLD and were followed up until the end of 2014. Effect of allopurinol vs. non-use was evaluated on reaching target serum urate (SU) level ≤360μmol/L, mortality, healthcare utilisation, vascular and renal diseases.

Three statistical approaches with differing complexities and assumptions imposed were considered: (1) baseline measurement of allopurinol and covariates with confounding controlled for using propensity score (PS) subclassification; (2) extending the methods in (1) to repeated measures where allopurinol and covariates were measured yearly; (3) using marginal structural models (MSM) within the repeated measures set-up. Survival models estimated hazard ratios with 95% confidence intervals. Robustness of estimated treatment effects to unmeasured confounding was evaluated.

Results

16,876 patients were eligible for analysis (mean age (standard deviation) 62 (14.1) years, 77% male). Baseline analysis found allopurinol was associated with higher chance of reaching target SU level (2.32 (1.97, 2.74)) and fewer gout consultations (0.70 (0.65, 0.75)), and with

increased risk of mortality (1.10 (1.03, 1.17)), gout hospitalisation (1.82 (1.64, 2.02)), coronary heart disease (1.11 (1.02, 1.21)), and renal disease (1.19 (1.10, 1.28)).

In the repeated measures setting, issues with poor performance of PS estimation were identified in both time-varying PS subclassification and MSM. These were resolved by allowing associations between covariates and initiation and continuation of allopurinol to differ in MSM; larger treatment effect estimates were obtained for most outcomes compared with baseline analysis and statistical significance was lost for mortality. The treatment effect estimates for target SU level and gout hospitalisation were likely to be robust to unmeasured confounding however, unmeasured confounding may explain away the treatment effects for coronary heart disease and renal disease.

Conclusion

Fitting complex models to EHR is challenging and consideration needs to be given to both clinical and statistical assumptions made during data preparation and analysis. Associations of allopurinol with adverse outcomes persisted, regardless of statistical approach used. This may be due to remaining residual confounding and/or because allopurinol dosage and adherence is suboptimal in primary care. Nevertheless, the treatment effect estimates obtain are relevant to UK primary care and provide evidence that managing gout in the long term needs to be improved.

Dissemination

Planned publications:

Long-term effect of allopurinol use on serum urate levels, mortality and comorbidities in gout using marginal structural models: a CPRD study

Comparison of marginal structural models and propensity score stratification in estimating time-varying effect of treatment

Conferences and seminars:

British Society for Rheumatology 2019 (best oral abstracts session): *Long-term effect of allopurinol use on serum uric acid levels, mortality and comorbidities in gout patients: an electronic health record cohort study*

Invited talk to Oxford University, MSK pharmaco- and device epidemiology group 2018: *Experience of applying marginal structural models to electronic primary care medical records*

International Society for Clinical Biostatistics 2016 (big data in healthcare session): *Use of time varying propensity scores to mimic properties of randomised controlled trials: evaluating the effect of allopurinol in gout using the Clinical Practice Research Datalink*

NIHR Infrastructure Doctoral Training Camp 2016 (highly commended MPHrp poster award): *Effect of allopurinol on achieving serum uric acid level and on healthcare utilisation for the management of acute gout*

Keele METHSS seminar (2018): *Application of marginal structural models: modelling the effect of allopurinol on all-cause mortality in gout*

Keele Postgraduate Symposium (2014 – 2019)

2019: *Long-term effect of allopurinol use on serum uric acid levels, mortality and comorbidities in gout patients: an electronic health record cohort study*

2018: *Use of marginal structural models to estimate the long-term casual effect of allopurinol on mortality and comorbidity in gout*

2017: *Challenges in evaluating the treatment effect of allopurinol in gout using electronic medical records from the Clinical Practice Research Datalink*

2016: *Effect of allopurinol on achieving serum urate target and healthcare utilisation for the management of acute gout*

2015: *Statistical methods used to control for confounding in observational studies evaluating the treatment effect of allopurinol in gout: a literature review*

2014: *Long term effect of allopurinol on gout outcomes: use of propensity scores*

Table of contents

List of tables

List of figures

Abbreviations

List of abbreviations used throughout the thesis:

ACR: American College of Rheumatology

ATE: Average treatment effect

ATET: Average treatment effect for the treated

BMI: Body mass index

BSR: British Society of Rheumatology

CI: Confidence interval

CKD: Chronic kidney disease

Coxibs: Cyclooxygenase II inhibitors

CPRD: Clinical Practice Research Datalink

DF: Degrees of freedom

EHR: Electronic health records

EULAR: European League Against Rheumatism

FP: Fractional polynomial

FP1: Fractional polynomial terms of dimension 1 (or first degree)

FP2: Fractional polynomial terms of dimension 2 (or second degree)

GEE: Generalised estimating equations

GLM: Generalised linear model

GP: General practitioner

GPRD: General Practice Research Datalink

HES: Hospital Episode Statistics

HR: Hazard ratio

IPTW: Inverse probability treatment weighting

IMD: Index of multiple deprivation

IQR: Interquartile range

IV: Instrumental variable

LOCF: Last observation carried forward

MH: Mantel-Haenszel

MI: Multiple imputation

MIM: Missing indicator method

MSE: Mean squared error

MSM: Marginal structural models

MSU: Monosodium urate

NHANES: National Health and Nutrition Examination Survey

NSAIDS: Non-steroidal anti-inflammatory drugs

ONS: Office of National Statistics

OR: Odds ratio

PDC: Proportion of days covered

PH: Proportional hazards

PS: Propensity score

QOF: Quality and Outcomes Framework

RCT: Randomised controlled trial

RR: Risk ratio

SD: Standard deviation

SE: Standard error

SMD: Standardised mean difference

SNM: Structural nested models

SU: Serum urate

SUTVA: Stable unit treatment value assumption

THIN: The Health Improvement Network

ULT: Urate-lowering therapy

Acknowledgments

Firstly, I would like to thank my lead supervisor Dr Milica Bucknall-Blagojevic and co-supervisor Dr Edward Roddy. I am deeply grateful for their on-going support, tireless feedback, expertise, patience and kindness over the course of this PhD. I have gained a huge amount of invaluable knowledge for which I am thankful for. I would also like to acknowledge and thank my advisors Prof Kelvin Jordan and Prof Christian Mallen for their support and insight.

I am grateful to the School of Primary, Community and Social Care for the opportunity to complete this research. Furthermore, I would like to thank friends and colleagues for their encouragement.

Last but not least, I thank my family for their support in everything that I decide to do.

This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone.

1 Introduction and aims

1.1 Estimation of treatment effect: RCTs versus observational studies

Randomised controlled trials (RCTs) are considered to be the gold standard to infer causal effectiveness of treatment on outcome. Randomisation ensures observed and unobserved patient characteristics are balanced across treatment groups thus, any differences observed in outcome may be attributed to treatment. RCTs however often face a range of limitations and restrictions. They cannot address clinical questions where randomisation is unfeasible or interventions are potentially harmful; they typically specify strict inclusion and exclusion criteria, meaning that clinically important subgroups (such as those with comorbidity or the oldest age) may be ignored; they may be inappropriate for assessment of rare events or long-term outcomes, for example death; furthermore they may be subject to financial constraints and other practical and ethical issues (Sanson-Fisher et al., 2007, Black, 1996). Therefore, RCTs cannot support all treatment decisions and consequently many decisions are based on sub-optimal evidence or clinical based practice, knowledge and consensus (Frieden, 2017).

Another drawback of RCTs is their reliance on the intention-to-treat principle which assumes that once patients are randomly assigned to treatment, they actually received that treatment as intended and patient characteristics remained balanced across the treatment groups (Hernán and Hernández-Díaz, 2012). However, in practice, observing treatment on a single occasion (or time-invariant treatment) does not match real life clinical management where treatment varies over time. For example, the treatment given at presentation of a symptom or morbidity may be varied later as the condition progresses or improves, or side effects arise which may lead to non-compliance with the treatment.

Observational data collected on large populations over a long period of time have the potential to overcome such shortfalls of RCTs and provide estimates of benefits and harms of treatment in real life clinical settings. The increasing availability of data gathered and coded electronically in the course of routine health care contacts, provides the opportunity to follow a patient's course of illness from first presentation to long-term outcome and to assess the effect of treatment on outcome. For example, the Clinical Practice Research Datalink (CPRD), a database of primary care electronic health records (EHR), provides clinical records for up to 30 years for approximately 8.5 million patients; it has been used to estimate treatment outcomes for a range of health conditions such as assessing the effect of metformin use on risk of developing cancer (Farmer et al., 2019) and kidney morbidity on adverse cardiovascular events (Currie et al., 2019). Therefore, EHR potentially allow us to study the outcomes of real-life patterns of healthcare use and prescribing that varies over time; they further allow us to study both rare and long-term outcomes, thus increasing generalisability particularly as diverse patient populations and wider spectrum of disease severity that may otherwise be excluded are available. Furthermore, using EHR in research is a relatively cheap way to study treatment effect without investing considerable time collecting data (Patorno et al., 2013).

1.2 Confounding by indication

Confounding, defined as a spurious association between treatment and outcome due to a third variable that is associated with both, is prevented through randomisation in RCTs however, it is a major area of concern for researchers using observational data to estimate outcomes of healthcare. Treatment decisions may be influenced by many pre-treatment characteristics (covariates), such as severity of the disease or comorbidity, which may also be associated with subsequent outcomes, in which case such covariates are termed confounders.

RCTs remove this “confounding by indication” by ensuring treatment allocation cannot be influenced by such prognostic covariates beyond chance hence enabling outcomes to be directly comparable between treatment groups, with any differences attributable to effects of treatment. In observational studies, in contrast, any systematic differences in covariates between treatment groups mean that differences in outcome could be the result of either treatment effect or differences in covariates, or both.

Conducting observational studies to the same level of academic rigor as RCTs by minimising confounding effects can lead to comparable treatments effects with RCTs. Anglemyer et al. (2014) conducted a Cochrane review of 14 systematic reviews and methodological reviews of reviews published between 1990 and 2013; each review compared the effect of treatment between RCTs and observational studies (cohort and case-control studies). On average, no significant differences in treatment effect were found, even when separately comparing cohort and case-control studies with RCTs, and when stratifying by non-pharmacological and pharmacological interventions. Discrepancies in treatment effect were found between RCTs and observational studies included in three reviews, two of which may be due to insufficient control of confounding. Similar findings was observed in an earlier review that found little evidence that treatment effect estimates differed between observational and RCTs (Benson and Hartz, 2000).

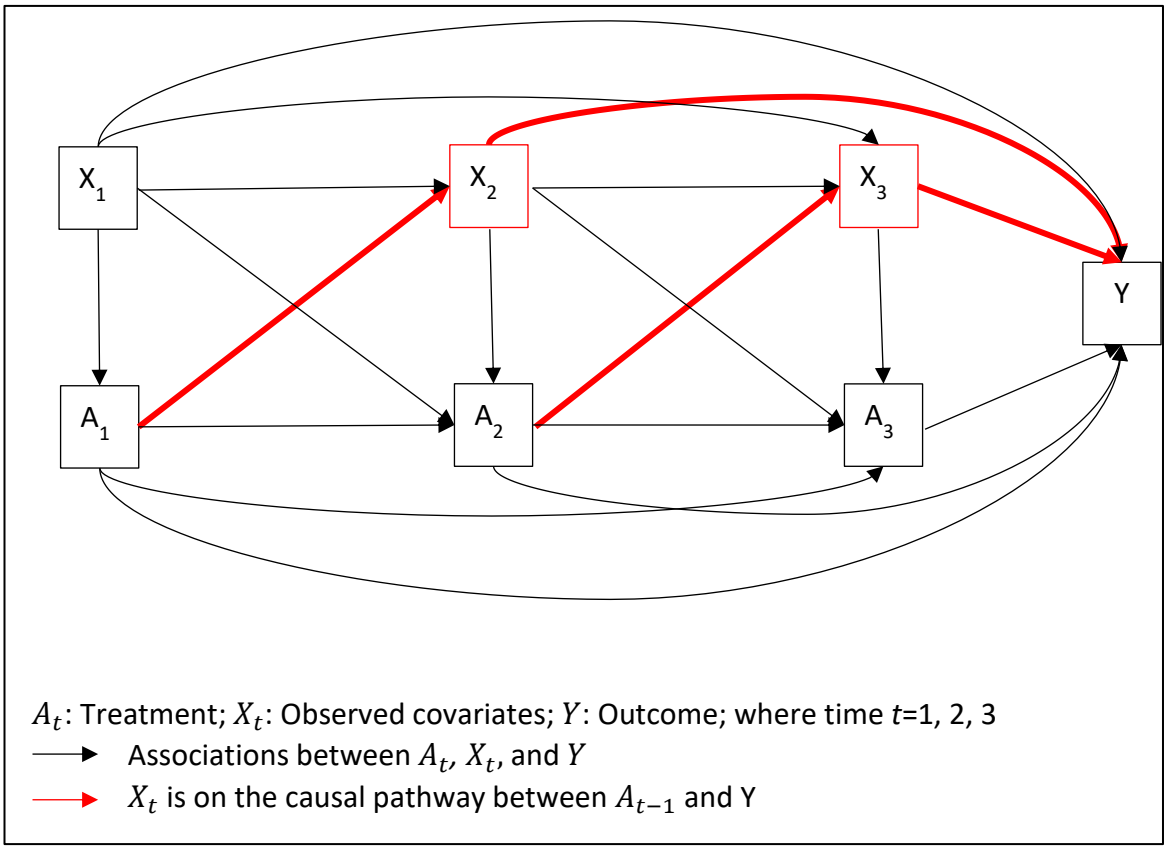
Multivariable regression modelling, which involves fitting a statistical model to estimate the association between a dependent variable (outcome) and one or more independent variables (treatment and covariates), is a popular approach used to account for observed confounders, but this approach has various drawbacks. For example, different regression models are subject to specific model assumptions which often go untested in practice. Furthermore, the estimated treatment effect would be biased if there is major imbalance in covariates between treatment groups and/or treatment effect varies across values of covariates (D'Agostino and

Kwan, 1995). Often, no attempt is made to ensure treatment groups are comparable by assessing whether the distribution of covariates is similar (adequate overlap) between treatment groups. Regression models will extrapolate data to cover areas where there is no overlap in incomparable groups. Matching may somewhat alleviate this problem by pairing treated patients with untreated patients on one or more covariates. This ensures adequate overlap between treatment groups as unmatched patients are excluded from analysis. However, matching may be limited as the number of covariates it is possible to match for is restricted by sample size.

As was stated earlier, in real life treatment is likely to vary over time. Furthermore, the covariates that influence choice of treatment and outcome may also change over time, giving rise to time-varying confounding as time-varying covariates are associated with both outcome and time-varying treatment. Estimating the overall treatment effect whilst accounting for time-varying covariates is complex. The key issue is that time-varying covariates also behave as mediators as they are affected by past treatment and are therefore on the causal pathway between treatment and outcome. This is illustrated in Figure 1.1 with treatment A and a single covariate X measured at three time points, assuming no unmeasured (or unobserved) covariates. The red arrows indicate when covariates act as mediators. At time point 1, covariates are adjusted for in regression modelling ensuring treatment groups are comparable. However, at time point 2, covariates and treatment are updated modifying the risk of outcome across treatment groups, meaning that treatment groups are now incomparable thus introducing residual confounding. A dilemma arises whether covariates at time point 2 need to be adjusted for, as this adjustment would remove the effect of treatment at time point 1 on outcome. The same issue applies at time point 3 when adjusting for these covariates would remove the effect of treatment at time point 2. Use of standard regression

models, for example the time-varying Cox model, in such instances may yield biased treatment effects thus their use is not recommended (Hernan et al., 2000, Robins et al., 2000).

Figure 1.1: Time-varying covariates affected by past treatment use



1.3 Estimation of treatment effect: Propensity score

An approach that avoids issues encountered in multivariable regression, and has gained widely in popularity over the last two decades, is propensity score (PS) methodology, originally formulated by Rosenbaum and Rubin (1983). The approach has been recommended for use in observational studies using EHR estimating effectiveness of treatment (Stuart et al., 2013a, Brookhart et al., 2010b). The attractiveness of this approach lies in the PS acting as a balancing score so that treatment groups homogeneous in PS have similar distribution of covariates, thus removing some of the bias due to confounding by indication before considering outcome and estimating treatment effect. Information from several covariates is collapsed into a single

score (the PS) which reflects the probability of a patient receiving treatment given these covariates. Once PS has been adjusted for in the analysis of treatment effect via matching, subclassification or weighting, any differences in outcome can be attributed to treatment assuming there is no unmeasured confounding.

The PS approach has widely been investigated and applied, however predominantly in scenarios where treatment is observed only once. In this instance, two systematic reviews have found that there was no difference in findings relating to effect of type of surgery on a range of outcomes between RCTs and PS-based analyses using observational data (Kuss et al., 2011, Lonjon et al., 2014).

In repeated measures design, a common approach to estimating effect of time-varying treatment is via one of Robins' G-methods (Robins, 1986, Robins et al., 2000, Robins et al., 1992), for example G-computation formula, however this approach assumes specification of the entire covariate history. Alternatively, one may use inverse probability of treatment (i.e., inverse PS) weighting (IPTW) of marginal structural models (MSM) (Robins 2000). Similar to PS, the probability of receiving treatment is estimated over time given treatment and covariate histories which are then converted to time-varying inverse probability of treatment weights. The weights reflect by how much observations are under-represented or over-represented in the study sample compared to the pseudo-population where there are no confounding effects. The treatment effect is then estimated in the pseudo- (or weighted) population. MSM are being increasingly used in EHR in various clinical areas such as diabetes (Gamble et al., 2017, Farmer et al., 2019), chronic kidney disease (Anderson et al., 2015) and gout (Desai et al., 2018).

In terms of more direct extension of the PS methodology as originally specified, Leon (2011b) estimated PS over time accounting for changing treatment status and covariates. Observations are then stratified into subclasses based on the PS, with treatment groups

comparable within each subclass. Treatment effects are estimated within each subclass and are then pooled. The authors mostly applied this method to the evaluation of ordinal doses of antidepressants in mental illness using a small number of time points (Leon et al., 2001, Leon et al., 2003); there are very few applications in other settings.

However, most of these approaches have been restricted in complexity to the case where all patients are observed at common discrete time points. This of course may be unrealistic in many clinical set-ups and is suboptimal for exploring treatment behaviour of individuals as represented by continually collected data such as that in CPRD where covariate and treatment observation time points differ between patients. Therefore, on one hand, EHR provide real life patterns of covariate, treatment and outcome data, but on the other hand lead to potentially complex analyses.

Performance and comparison of the different PS methods could be assessed under various scenarios via a simulation study. Some examples of existing simulation studies in this field include identifying the optimal PS matching approach (Austin, 2009b, Austin, 2011b, Austin, 2014), assessment of performance of time-varying PS subclassification in the estimation of treatment effect on different types of outcome data (Leon and Hedeker, 2005, Leon and Hedeker, 2007a, Leon, 2011b), and assessment of performance of different types of weights in reducing variability in the estimated treatment effect in MSM (Xiao et al., 2010). Ideally, the data generated in a simulation study should reflect the type of data observed in real life to assess how statistical models may perform in practice. Most existing relevant simulation studies have explored model performance under simple study designs with typically a small number of repeated measures or considered treatment at one time point. It is currently unclear what sort of data management and computational challenges may be encountered when applying PS based methods in large and complex observational studies, such as those based on EHR data. In order to design an informative simulation study, it would be necessary

to know what these challenges may be. This could be investigated by first applying different PS based approaches in analysis of real data, to give an indication of the model implementation challenges, types of observed treatment patterns, strength of plausible associations between covariates and treatment/outcome etc, all of which could subsequently inform a thorough simulation study. A simulation study was therefore considered to be outside the scope of this PhD project.

This PhD project will use gout as an exemplar to explore these issues.

1.4 Effectiveness of allopurinol in gout

Gout is the most common type of inflammatory arthritis. In the UK population, prevalence and incidence is reported to be 2.49% and 1.77 cases per 1,000 person-years respectively in 2012. Both prevalence and incidence increase with age and is higher in males than in females (Kuo et al., 2015b).

The key risk factor for gout is persistent hyperuricaemia causing monosodium urate (MSU) crystals to form and deposit in and around the joints; this leads to various clinical manifestations. The most common manifestation is the self-limiting painful flares of joint inflammation and swelling, typically seen in the first metatarsophalangeal joint, that lasts up to two weeks. If hyperuricaemia is not treated, subsequent flares often become more frequent, last longer, and may affect different joints. Over a long period of time, chronic gouty arthritis develops where joint inflammation becomes persistent due to recurrent flares; subcutaneous tophaceous deposits, crystals compressed with debris from the inflammatory response, harden and cause further joint damage. Consequently, all this leads to irreversible joint damage, pain, and disability (Roddy et al., 2013, Chandratre et al., 2018, NICE, 2018a).

Gout is often treated in primary care. The British Society for Rheumatology (BSR) (Jordan et al., 2007b, Hui et al., 2017) and the European League Against Rheumatism (EULAR) (Zhang et al., 2006a, Richette et al., 2017) have published guidelines for the management of gout. Long-term management involves urate-lowering therapy (ULT) with a treat-to-target strategy to lower serum urate (SU) levels to below the threshold of urate saturation to prevent the formation of new crystals and allow deposited crystals to dissolve and tophi to shrink; the aim of ULT is to prevent flares and 'cure' the patient of gout. The BSR guidelines recommend an initial target SU level ≤300μmol/L that can be relaxed to ≤360μmol/L once the patient no longer has tophi and flares; the same target levels are recommended by the EULAR guidelines but those with severe gout are recommended to have a lower target of ≤300μmol/L. ULT involves addressing modifiable risk factors for hyperuricaemia, for example diet and medication use, and taking urate-lowering drugs. Allopurinol, a xanthine oxidase inhibitor, works by decreasing SU production and is the first line drug treatment. The BSR guidelines state patients with comorbidities associated with gout (renal impairment and hypertension) and those with more severe gout (recurrent flares, tophi, chronic gouty arthritis, and joint damage) should be offered allopurinol. Allopurinol treatment is recommended to be life-long.

Allopurinol was developed for the treatment of gout over 50 years ago however, only a few RCTs have investigated its effectiveness. Seth et al. (2014) conducted a Cochrane review in 2014 evaluating the effect and safety of allopurinol in chronic gout. Eleven RCTs were identified with eight studies comparing allopurinol with other urate-lowering or uricosuric drugs. The review highlighted a lack of high quality RCTs due to small samples and attrition bias. Studies were limited up to 52 weeks follow-up and employed strict inclusion and exclusion criteria, for example only including patients with severe hyperuricaemia or excluding patients with renal impairment, which may limit their generalisability to the majority of patients with gout (Becker et al., 2005, Schumacher et al., 2008, Becker et al., 2010). More

recently, larger RCTs with more than 6,000 participants have been undertaken, however, strict inclusion and exclusion criteria were still used (White et al., 2018, MacDonald et al., 2014), and length of follow-up remained limited to around three years. One RCT generalisable to the UK gout population compared the efficacy of nurse-led care with usual general practitioner (GP)-led care on reaching primary outcome, target SU level ≤360µmol/L, with ULT; within nurse-led care, all participants were treated with allopurinol at baseline and had reductions in number and size of tophi, number of flares, reduction in SU level with increased number of patients reaching target SU level (≤300µmol/L and ≤360µmol/L), and improved physical health over two years (Doherty et al., 2018).

Despite the availability of guidelines, management of gout is often suboptimal and that uptake, adherence, and persistence to ULT are often poor. Various studies have shown only 30-40% of patients with gout were ever prescribed ULT, of which the majority were prescribed allopurinol dose ≤300mg/day (Cottrell et al., 2013, Roddy et al., 2007b, Clarson et al., 2017). The majority of patients with indications for ULT were not prescribed ULT; at diagnosis 44% of patients had indications for ULT of whom <1% were prescribed ULT (Kuo et al., 2014). Once patients had started allopurinol, 39% stopped treatment after a year; 64% of these restarted treatment within five years. Median patient time of allopurinol adherence (time covered with a prescription) was 67% (Scheepers et al., 2018). Consequently, only up to 40% of patients reach target SU level (Cottrell et al., 2013) hence approximately <1 in 10 patients are `cured'. Gout is managed ineffectively for a number of reasons including lack of GP training, focus on treating flares only, underestimation of the long-term effects of gout, lack of patient and practitioner understanding of gout pathogenesis, and the benefits and role of ULT (Doherty et al., 2012).

Estimating the effect of allopurinol using EHR is challenging. Confounding by indication is likely to be present as guidelines suggest patients with poorer health should be prescribed

allopurinol however, in practice this is often not realised. Allopurinol treatment is time-varying with patients often stopping and restarting treatment over time, with each treatment choice being influenced by time-varying and time-invariant covariates measured up to that time point.

Around the time when this PhD project was conceptualised in 2013, there were a few existing EHR based studies evaluating the effect of allopurinol in gout using PS methodology; compared with those not using allopurinol, allopurinol users were found to have an increased risk of severe cutaneous adverse reactions although the study was not limited to patients with gout (Kim et al., 2013a); Wei et al. (2011) reported no difference in risk of cardiovascular events between allopurinol users and non-users of ULT however, this study too was not restricted to gout. There was a general lack of consideration of a wider range of outcomes, in particular renal and cardiovascular diseases.

Since then, EHR observational studies evaluating the effect of allopurinol on different outcomes have been on the increase, for example mortality (Dubreuil et al., 2015, Kuo et al., 2015a), chronic kidney disease (Roughley et al., 2018, Vargas-Santos et al., 2018), and vascular diseases (Sultan et al., 2019); three of these studies used PS matching (Dubreuil et al., 2015, Kuo et al., 2015a, Vargas-Santos et al., 2018). However, these studies evaluated the effect of *initiating* allopurinol treatment thus ignoring time-varying confounding.

One issue frequently encountered in EHR studies using time-to-outcome data is immortal time bias, defined as, a period of follow-up time during which outcome cannot occur. This can arise during the period from study entry to prescription of treatment. The landmark method overcomes this issue by designating a period of time from study entry (the landmark period) to determine treatment status; follow-up commences after the landmark period in patients who did not have outcome up until then. This method has frequently been used in EHR studies in the evaluation of effectiveness of allopurinol (Sultan et al., 2018, Kuo et al., 2015a, Roughley

et al., 2018). However, the disadvantage of the landmark method is that it can potentially misclassify treated patients as untreated if the landmark period is too short, or exclude too many patients who had outcome occurring early on if the landmark period is too long (Dafni, 2011). Alternatively, to reduce misclassification bias from patients starting allopurinol many years from the start of follow-up, follow-up may instead be assumed to start when allopurinol was first prescribed (Sultan et al., 2019, Vargas-Santos et al., 2018, Dubreuil et al., 2015). However, these studies assumed patients did not change treatment status during follow-up.

Some studies may alternatively perform per-protocol analysis where analysis only includes patients who adhered to treatment. For example, Vargas-Santos et al. (2018) censored patient follow-up time when allopurinol treatment had stopped or changed. However, this may introduce selection bias if the reasons for stopping or changing allopurinol treatment, for example may have potentially experienced an adverse reaction to allopurinol or no longer had gout flares, is associated with outcome (Hernán and Hernández-Díaz, 2012).

None of these studies evaluated the effect of allopurinol on outcome when patients initiate, stop and restart treatment in the presence of time-varying confounding. Therefore, an observational study based on EHR is both needed and timely to investigate the realistic effect of allopurinol use versus non-allopurinol use over a long period of time on a range of outcomes. CPRD is one of the largest UK databases of primary care EHR and is thus an ideal data-source given that gout is generally diagnosed and managed in primary care.

As explained above, bias due to confounding by indication is a major impediment to valid assessment of treatment effect based on observational data. Use of PS and MSM methodologies will allow causal inferences to be made provided careful consideration is given to the study design, data manipulation and choice of covariates to adjust for. Naturally, not all important covariates will be measured or even observable, such as adherence of allopurinol uptake and genetic factors, thus some residual confounding will remain. Often, no attempts

are made to understand the extent to which residual confounding can impact treatment effect estimation.

To our knowledge, these methodologies have rarely been used to estimate effect of time-varying allopurinol use and never in EHR setting in the UK.

1.5 Thesis aims and overview

This PhD aims to approach estimation of allopurinol effect using observational EHR data in a comprehensive and thorough manner. From a clinical aspect, this will be achieved by considering a wide range of outcomes in patients with gout, relaxing exclusion criteria, and stratifying analyses on severity of SU levels and renal disease.

From a statistical aspect, allowance will be made for time-varying allopurinol use and time-varying covariates thus giving rise to repeated measures data structure within which both PS and inverse probability treatment weights and subsequently treatment effect will be estimated. Robustness of allopurinol treatment effect estimates to missing data and omission of important covariates will be tested. Comparisons of methodology and results will be made between PS and MSM.

The clinical objectives are:

1) To examine the effect of allopurinol versus not taking allopurinol on reaching target SU level ≤360μmol/L, the primary outcome. Secondary outcomes are all-cause mortality; gout hospital admission; joint replacement; and gout comorbidities (cerebrovascular, coronary heart, peripheral vascular diseases and renal disease).
2) Repeat objective 1 stratified on baseline levels of SU level (above and below 480μmol/L) and on presence of renal disease to assess whether effect of allopurinol varies by severity of SU level or renal disease.

The statistical objectives which will be addressed within objectives 1 and 2 above are to compare and contrast:

3) The effect of including time-varying covariates in addition to baseline covariates in PS subclassification on treatment effect.
4) Compare and contrast estimation of treatment effect from time-varying PS subclassification from objective 3 to MSM.
5) To examine the sensitivity of treatment effect estimates obtained in objective 1 to missing data and omission of covariates (with varying degree of strength of association with outcome and/or treatment) from modelling of treatment assignment to address the impact of unobserved covariates.

A brief description of each chapter is given below.

Chapter 2*: Gout*

A description of the clinical characteristics, risk factors, prognosis, and diagnosis of gout is given. Short- and long-term management of gout is discussed and the extent to which patients are prescribed allopurinol and remain on treatment.

Chapter 3*: A narrative review of observational studies evaluating the effect of allopurinol in gout*

A systematic search to identify all published observational studies evaluating the effect of allopurinol in gout. A narrative review of the eligible studies summarised the study design, adjustment for confounding, sources of data, definition of allopurinol exposure, and outcomes used. The findings from the review will identify the extent of use of EHR in evaluation of allopurinol effect as well as highlight methodological limitations of existing studies and where possible methodological improvements can be made.

Chapter 4*: Data source and study sample*

How the population-based cohort study is set up, how the study sample is defined, and the definitions for allopurinol use, covariates, and outcomes from electronic primary care medical records in CPRD are provided.

Chapter 5*: Propensity scores and marginal structural models*

The concept of the casual inference framework and advantages and disadvantages of common methods used to control for confounding are discussed. Reasons why PS and MSM methodologies are chosen to control for confounding is justified. Further details on how to fit such models are described.

Chapter 6*: Statistical analysis plan*

The four methods applied to CPRD data in estimating treatment effect of allopurinol are described. The first two methods focussed on PS subclassification performed at a single point in time and then repeatedly over time; description of covariate selection for PS estimation, how the number of subclasses are determined, and assessment for balance and treatment effect estimation are provided. The third and fourth methods focussed on MSM and the assumption regarding associations between treatment and covariates are stated; initially it is assumed that the reasons for prescription of treatment are the same amongst patients initiating and continuing with treatment, and subsequently these reasons are then allowed to differ.

Chapters 7*: Effect of allopurinol: time-invariant PS subclassification*

Treatment effect estimates are presented after using PS subclassification at a single time point to control for confounding. This chapter addresses all the clinical objectives.

Chapter 8*: Effect of allopurinol: time-varying PS subclassification*

Similar to chapter 7, treatment effect estimates are presented after using time-varying PS subclassification.

***Chapter 9**: Modelling simple mechanisms of allopurinol via MSM*

Reasons for treatment are assumed not to differ between patients initiating and continuing treatment. This assumption led to unreliable treatment effect estimates. Results from various methods employed to improve weight estimation are presented. Outcome considered is all-cause mortality.

***Chapter 10**: Modelling complex mechanisms of allopurinol via MSM*

Treatment effect estimates are presented after using MSM allowing reasons for prescription of treatment to differ amongst patients initiating and continuing with treatment.

***Chapter 11:** Discussion*

The final chapter describes the overall conclusions, strengths and limitations of the methodologies, and further advancements required.

2 Gout

The aim of this chapter is to provide background information about gout that is of relevance to this thesis. An overview of clinical manifestations, risk factors, and poor outcomes is provided. Diagnosis of gout, management, and implementation of management for gout is described.

2.1 Hyperuricaemia, crystal formation and clinical presentation

The development and progression of gout is viewed as three overarching disease states (Bursill et al., 2019). These include the pre-clinical state, gout flare and advanced gout manifestations, which are now briefly described.

Pre-clinical state

Asymptomatic hyperuricaemia with monosodium urate (MSU) crystal deposition precedes gout (Choi et al., 2005). Uric acid is the end product of the degradation of purines, nucleotide bases which are a key component of DNA. Uric acid largely exists in its ionised form, urate, at physiological pH and temperature. Urate level is dependent upon endogenous purine metabolism, dietary intake of purine rich food, and rate of excretion via the kidneys and gut. In 90% of cases hyperuricaemia results from renal under-excretion of urate, and in the 10% of cases a combination of urate underexcretion and overproduction (Choi et al., 2005). MSU crystal formation occurs when the urate level persistently exceeds 380µmol/L; the synovial fluid becomes supersaturated with urate causing MSU crystals to form and deposit in the surrounding cartilage, bone, and tissues (Seegmiller, 1965, Loeb, 1972).

Gout flare

The most common clinical manifestation of gout is occurrence of episodic flares. Gout flares occur from shedding of crystals from the deposits into the joint space eliciting an inflammatory response at the joint site. The gout flare has distinct characteristics of reaching peak inflammation within 24 hours of onset accompanied by swelling, tenderness, and excruciating pain in the affected joint that resolves itself within one to two weeks (Roddy et al., 2013). Gout is frequently monoarticular and typically affects the lower limb with the majority of flares occurring in the first metatarsophalangeal joint. Other joints frequently involved include the mid-foot, ankle and knee, and less frequently the upper limbs and fingers (Roddy, 2011).

After the gout flare has resolved, the patient enters the inter-critical period and generally remains symptom free until the next flare. Despite being symptom free, if hyperuricaemia is not treated, crystals continue to form and cause low persistent inflammation with the majority of patients having a second flare within two years (Yu and Gutman, 1961). Subsequent flares become more frequent, last longer, and more joints become involved (oligoarticular (affecting 2-4 joints) or polyarticular gout (≥5 joints)), and may affect also the upper limbs (Dalbeth et al., 2016).

Advanced gout manifestations

Untreated hyperuricaemia over a number of years may lead to chronic gouty arthritis, tophi, and bone erosion in some patients. Joint inflammation becomes persistent due to crystals causing inflammation from recurrent flares and low-level inflammation during the inter-critical period.

Crystals harden and compress forming tophaceous deposits mainly in subcutaneous and periarticular areas such as fingers, toes, knees, olecranon processes, Achilles' tendons, and

helix of the ears. Subcutaneous tophi are typically pain free. They may have a white to yellow appearance and are asymmetrically shaped (Roddy, 2011, Dalbeth et al., 2016).

A combination of synovial hypertrophy from chronic inflammation, gout flares, and tophi presence and size leads to irreversible structural joint damage (Wu et al., 2019) and poor health-related quality of life (Khanna et al., 2012c). Bone erosion is dominantly due to tophi infiltrating into the bone (Dalbeth et al., 2016).

2.2 Diagnosis

The European League Against Rheumatism (EULAR) have published guidelines on how gout should be diagnosed (Zhang et al., 2006b, Richette et al., 2020). The definitive diagnosis of gout involves joint aspiration of synovial fluid or tophi for microscopic examination of MSU crystals. The procedure can be performed during a gout flare or during the inter-critical period between flares (Roddy et al., 2013). Joint aspiration is rarely performed in primary care which may be due to lack of facilities or expertise (Underwood, 2006, Kienhorst et al., 2014).

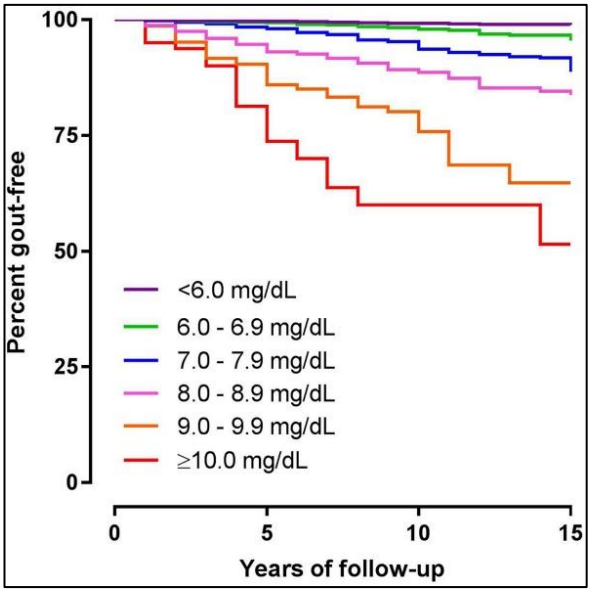
In practice, diagnosis is often based on clinical signs and symptoms. Clinical features of gout and hyperuricaemia are highly suggestive but not specific for gout. Consequently, gout can be misdiagnosed. Gout can be mistaken for a flare of osteoarthritis in the 1st MTP joint in the presence of hyperuricaemia or pseudo-gout, a type of arthritis resulting from deposits of calcium pyrophosphate crystals (Sturrock, 2000). Joint aspiration should be undertaken if diagnosis of the inflamed joints is unclear.

2.3 Risk factors for gout

2.3.1 Hyperuricaemia

Longitudinal population-based studies have long established the association between hyperuricaemia (elevated serum urate (SU) level) and gout. These studies have shown higher levels of SU are associated with greater risk of incident gout (Campion et al., 1987), with incidence being higher in men compared to women (Bhole et al., 2010). Similarly, hyperuricaemia was found to increase the risk of recurrent gout flares (Trifiro et al., 2013). More recently, Dalbeth et al. (2018) pooled participants from four large cohort studies, and showed the cumulative incidence of gout over 15 years was 1.1% in patients with SU <6mg/dL but 49% in patients with SU level ≥10mg/dL. Figure 2.1 illustrates the percentage of patients remaining gout free over time based on different SU categories (Dalbeth et al., 2018).

Figure 2.1: Kaplan-Meier plot showing the percentage of participants who were gout-free over the follow-up period, based on baseline SU categories in mg/dL



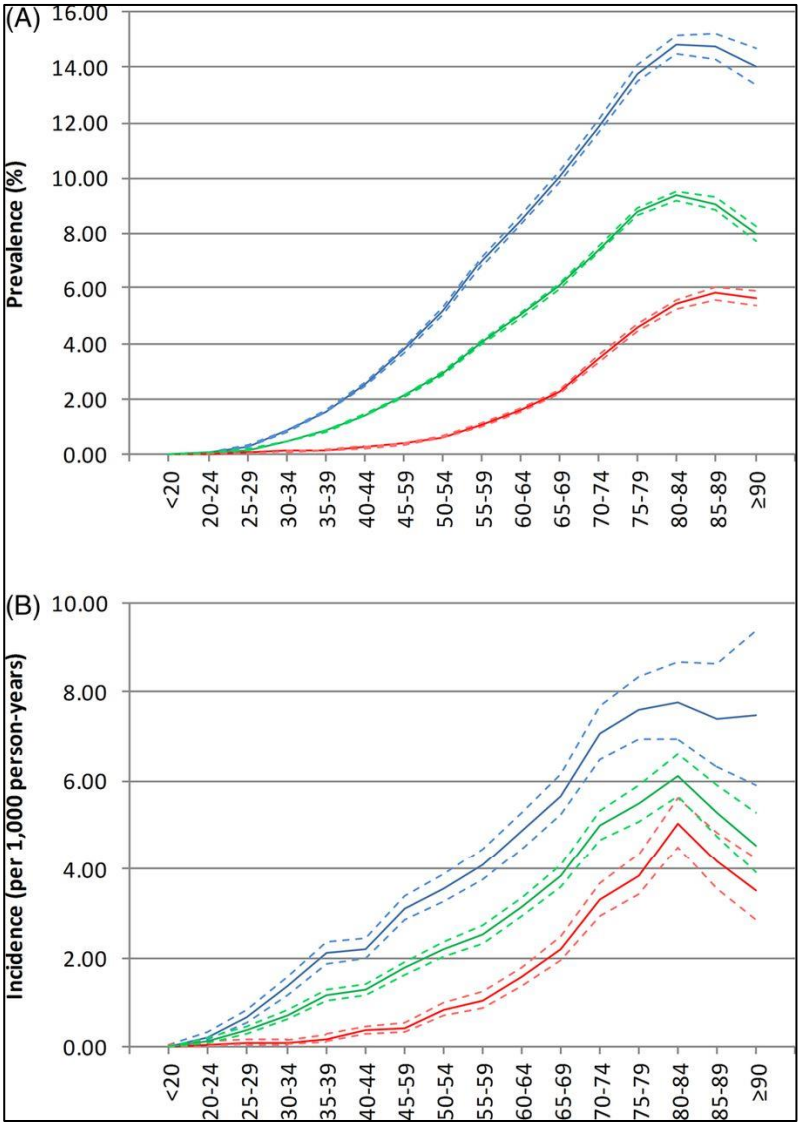
Reproduced from Relationship between serum urate concentration and clinically evident incident gout: an individual participant data analysis, Dalbeth et al, vol. 77, pg. 1048-1052, 2018, with permission from BMJ Publishing Group Ltd. The American system measured SU level as mg/dL; 6.0mg/dL is equivalent to 360µmol/L; 7.0mg/dL is equivalent to 420µmol/L; 8.0mg/dL is equivalent to 480µmol/L; 9.0mg/dL is equivalent to 540µmol/L; 10.0mg/dL is equivalent to 600µmol/L.

Not all individuals with hyperuricaemia go on to develop gout with prevalence of hyperuricaemia (20%) being higher than the prevalence of gout (3.9%) (Chen-Xu et al., 2019).

2.3.2 Demographics

In the UK population in 2012, both prevalence and incidence of gout increased with increasing age, and were higher in males than in females (Figure 2.2) (Kuo et al., 2015b).

Figure 2.2: Age-specific prevalence (A) and incidence (B) of gout in 2012



Reproduced from Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study, Kuo et al, vol. 74, pg. 661-667, 2015, with permission from BMJ Publishing Group Ltd. (Blue: men; red: women; green: total; dotted lines show 95% confidence bounds)

Gout has historically been associated with male sex however prevalence and incidence of gout has increased over time in both males and females. In females, prevalence and incidence rises after menopause. Hak et al. (2010) found amongst menopausal women, that those taking hormone replacement therapy had lower risk of gout than non-users. Oestrogen aids urate

renal excretion (Nicholls et al., 1973) and once women stop producing oestrogen, SU levels begin to rise and consequently the risk of gout increases.

Historically, gout was perceived to be associated with socioeconomic privilege and frequently referred to as the 'disease of kings'. Contemporary studies have otherwise shown the opposite to be true; gout is both more common and more severe with greater deprivation. Prevalence of gout was higher in a more deprived English town (4.8%) compared with an English town that was considered less deprived (3.9%) (Gardner et al., 1982). Further education (odds ratio (OR) 0.54 (95% CI: 0.36, 0.81)) and low area-level deprivation determined by the Index of Multiple Deprivation (OR 0.71 (95% CI: 0.51, 0.98)) were associated with lower risk of having ≥2 flares in the preceding 12 months. Although deprivation does not have a direct impact on SU levels, it is linked to poorer health and greater number of visits to the general practitioner (GP), and may contribute to delayed consultation and reluctance to acknowledge having gout (Bowen-Davies et al., 2018).

2.3.3 Diet

Most urate is produced due to the process of aging and the associated increased cell death in the human body however, urate is also produced by the metabolism of purines found in the diet.

Li et al. (2018) performed a systematic review and meta-analysis of dietary factors and risk of incident gout (19 prospective cohort and cross-sectional studies). Red meat, seafood, alcohol consumption, and fructose sweetened soft drinks increased risk of incident gout. On the other hand, dairy products, soy foods, high-purine vegetables, and coffee consumption were protective against gout. Another study had shown higher vitamin C consumption yielded lower risk of gout (Choi et al., 2009).

Studies have also investigated whether dietary factors trigger recurrent gout flares. Two case-cross over studies had found greater alcohol consumption was associated with greater odds of flares (Neogi et al., 2014), whereas greater cherry consumption was associated with lower odds (Zhang et al., 2012).

2.3.4 Genetics

Gout tends to cluster within families, implying that it may be related to lifestyle and genetic factors. Genome-wide association studies had found a large number of genes involved in the development of hyperuricaemia and gout. The majority of these genes are involved with renal function where urate is under excreted. The heritability, defined as the percentage variance of phenotype that is explained by inherited genetic variants, can be estimated by studying the phenotypic correlation between related individuals (Major et al., 2018). The heritability of SU levels and gout in Europeans was estimated to be between 27% and 41% (Köttgen et al., 2013) and approximately 30% (Cadzow et al., 2017), respectively.

2.3.5 Metabolic syndrome

The association between gout and the metabolic syndrome is well known with several large studies evaluating these associations (Roddy and Choi, 2014). There are various definitions for the metabolic syndrome but they all share the same comorbidities of obesity, hypertension, dyslipidaemia, and either or both hyperinsulinaemia or hyperglycaemia (Eckel et al., 2005).

Prevalence of the metabolic syndrome and its individual components was higher in patients with gout compared with those without (Choi et al., 2007).

A systematic review and meta-analysis of 11 cohort studies found obesity (risk ratio (RR) 2.24 (95% CI: 1.76, 2.86)) and hypertension (RR 2.11 (95% CI: 1.64, 2.72)) doubled the risk of incident gout (Evans et al., 2018).

Risk of gout differs in patients pre- and post-diagnosis of diabetes. Studies have shown SU level is higher in prediabetes vs. non-diabetics (Herman and Goldbourt, 1982) and increases with HbA1c levels (Choi and Ford, 2008); this is explained by high insulin levels impairing renal urate excretion leading to hyperuricaemia (Ter Maaten et al., 1997). Conversely, analysis of primary care medical records has shown uncomplicated and complicated diabetes lowers the risk of incident gout (OR 0.90 (95% CI: 0.85, 0.96) and (OR 0.87 (95% CI: 0.76, 1.00), respectively) compared with not having diabetes (Kuo et al., 2016b), and lowers the risk of recurrent flare (hazard ratio (HR) 0.92 (95% CI: 0.85, 0.99)) (Rothenbacher et al., 2011). The negative association between diabetes and gout may be explained by high glucose levels in the urine (glycosuria) leading to increased urine volume (polyuria) and enhanced urate excretion, lowering SU levels (Cook et al., 1986).

2.3.6 Chronic kidney disease

The association between chronic kidney disease (CKD) and gout is well established. A Clinical Practice Research Database (CPRD) study found renal disease was associated with incident gout (OR 6.63 (95% CI: 5.18, 8.48) (Kuo et al., 2016b). This is due to impaired kidney function leading to reduced urate excretion and increased risk of hyperuricaemia and gout.

Conversely, various studies have shown that gout is associated with incident CKD. A meta-analysis study of three cross-sectional studies found that compared to patients without gout, those with gout had increased odds (OR 2.41 (95% CI: 1.86, 3.11)) of CKD stage ≥3 (Roughley et al., 2015); two studies in CPRD found gout was associated with increased risk of advanced

CKD (HR 1.29 (95% CI: 1.23, 1.35)) and its components including end stage kidney disease, estimated glomerular filtration rate <10 mL/min/1.73m^2, and doubling of serum creatinine from baseline (Stack et al., 2019), and incident renal disease (Kuo et al., 2016b). The risk of renal disease is higher in patients with gout than those without due to comorbid hypertension and diabetes, hyperuricaemia, chronic inflammation, and non-steroidal anti-inflammatory drug (NSAID) use (Roughley et al., 2018).

Although high SU levels have been shown to be associated with kidney disease (Li et al., 2014), recent studies have suggested SU level may not have a causal relationship. Randomised controlled trials (RCTs) have demonstrated urate-lowering therapies (ULTs) do not slow the rate of kidney function decline (Badve et al., 2011, Badve et al., 2020, Kimura et al., 2018, Doria et al., 2020) whilst Mendelian randomisation studies have shown no effect of SU level on CKD (Jordan et al., 2019).

2.3.7 Osteoarthritis

There is limited evidence osteoarthritis is a possible risk factor for gout, under the mechanism that MSU crystals deposit more easily in osteoarthritic joints (Ma and Leung, 2017). A cross-sectional study found flares at a particular joint were associated with presence of osteoarthritis within that joint (OR 7.94 (95% CI: 6.27, 10.05)), with statistically significant associations found at the 1st metatarsophalangeal joint, mid-foot, knee and distal interphalangeal joints in those with gout (Roddy et al., 2007a). An analysis of CPRD data found osteoarthritis was associated with incident gout (OR 1.27 (95% CI: 1.20, 1.34)) (Kuo et al., 2016b).

As stated in Section 2.1, joint damage may occur in gout. A small cross-sectional study had found presence of tophi over 5mm at a particular joint was associated with bone erosion

within that joint, and there is a positive strong correlation between tophi size and the degree of erosion (McQueen et al., 2014). A cohort study using data from CPRD had shown prevalence of total joint replacement was higher in people with gout than those without (2.61% vs. 1.76%) (Kuo et al., 2018). Given osteoarthritis is the largest risk factor for joint replacement and predisposes gout, gout is therefore a risk factor for joint replacement.

2.3.8 Medications

Antihypertensives are widely used to treat hypertension and thus have protective effects against myocardial infarction and strokes. However, diuretics (thiazides and loop diuretics), used to treatment hypertension and heart failure, commonly cause hyperuricaemia and gout (Pascual and Perdiguero, 2006). A systematic review and meta-analysis of three cohort studies found diuretic use increased the risk of developing gout (RR 2.39 (95% CI: 1.57, 3.65)) (Evans et al., 2018). Choi et al. (2012) evaluated the risk of various antihypertensives on incident gout in a large case-control study using primary care electronic health records (EHR) from The Health Improvement Network (THIN) database. Current prescriptions of calcium channel blockers and losartan had lower odds of gout yielding OR 0.87 (95% CI: 0.82, 0.93) and 0.81 (95% CI: 0.70 to 0.94) respectively. On the other hand, diuretics (OR 2.36 (95% CI: 2.21, 2.52)), beta-blockers (OR 1.48 (95% CI: 1.40, 1.57)), angiotensin converting enzyme inhibitors (OR 1.24 (95% CI: 1.17, 1.32)) and other non-losartan angiotensin II receptor blockers (OR 1.29 (95% CI: 1.16 to 1.43)) had greater odds of gout compared with non-use.

Aspirin, a blood thinning drug, lowers urate excretion when used in lower cardioprotective doses whereas high anti-inflammatory doses have the opposite effect of increasing urate excretion. A case-crossover study found low dose aspirin increased odds of recurrent gout flares compared with no use in the last two days (OR 1.81 (95% CI: 1.20, 2.51)). The odds of

recurrent gout flares were higher with lower aspirin doses compared with non-aspirin use (Zhang et al., 2014). However, since gout is associated with cardiovascular disease, the cardioprotective effects of low dose aspirin are thought to outweigh its slight effect to increase urate levels.

2.4 Long-term outcomes from gout

Gout increases the risk of poor outcomes over a long period of time, particularly cardiovascular and renal diseases. This may partly be attributed to poor management of hyperuricaemia, as hyperuricaemia has been verified as an independent risk factor of cardiovascular and renal diseases (Gaffo et al., 2009). Persistent inflammation may also play a role.

Several studies analysing primary care medical records from CPRD have shown gout increased the risk of various vascular diseases, genitourinary diseases, and comorbidities (Table 2.1).

Gout is associated with many poor outcomes and unsurprisingly, is also associated with a greater risk of premature mortality. Using Swedish medical records, compared with non-gout patients, those with gout were found to be at increased risk of death due to cardiovascular disease (HR 1.27 (95% CI: 1.22, 1.33)), renal disease (HR 1.78 (95% CI: 1.34, 2.35)), and diseases of the digestive system (HR: 1.56 (95% CI: 1.34, 1.83)) (Vargas-Santos et al., 2019).

Table 2.1: Summary of the findings from previous studies of gout outcomes undertaken in CPRD

Outcomes	Overall HR (95% CI)	Men HR (95% CI)	Women HR (95% CI)
Vascular diseases			
Any vascular disease (Clarson et al., 2015)	-	1.06 (1.01, 1.12)	1.25 (1.15, 1.35)
Any coronary heart disease	-	1.08 (1.01, 1.15)	1.25 (1.12, 1.39)
Angina	-	1.02 (0.92, 1.13)	1.28 (1.09, 1.51)
Myocardial infarction	-	1.12 (1.00, 1.27)	0.97 (0.77, 1.22)
Any cerebrovascular disease	-	0.95 (0.83, 1.09)	1.17 (0.99, 1.38)
Transient ischaemic attack	-	1.02 (0.88, 1.18)	1.26 (1.05, 1.53)
Cerebrovascular attack	-	0.93 (0.81, 1.06)	1.34 (1.15, 1.57)
Peripheral vascular disease	-	1.18 (1.01, 1.38)	1.89 (1.50, 2.38)
Atrial fibrillation (Kuo et al., 2016a)	1.09 (1.03, 1.16)	1.09 (1.01, 1.16)	1.12 (1.02, 1.24)
Venous thromboembolism (Sultan et al., 2019)	1.25 (1.15, 1.35)	1.20 (1.09, 1.33)	1.32 (1.14, 1.52)
Cardiac arrhythmias (Kuo et al., 2016b)	1.59 (1.48, 1.70)	-	-
Congestive heart failure (Kuo et al., 2016b)	1.81 (1.65, 1.98)	-	-
Valvular heart disease (Kuo et al., 2016b)	1.80 (1.60, 2.04)	-	-
Chronic pulmonary disease	1.10 (1.02, 1.18)	-	-
Genitourinary diseases			
Chronic kidney disease (Roughley et al., 2018)	1.78 (1.70, 1.85)	1.78 (1.69, 1.87)	1.79 (1.66, 1.93)
Renal disease (Kuo et al., 2016b)	3.18 (2.88, 3.50)	-	-
Urolithiasis (Kuo et al., 2016b)	1.26 (1.02, 1.55)	-	-
Comorbidities			
Hypertension (Kuo et al., 2016b)	1.51 (1.43, 1.58)	-	-
Hyperlipidaemia (Kuo et al., 2016b)	1.40 (1.31, 1.50)	-	-
Hypothyroidism (Kuo et al., 2016b)	1.46 (1.32, 1.61)	-	-
Osteoarthritis (Kuo et al., 2016b)	1.45 (1.35, 1.54)	-	-
Depression (Kuo et al., 2016b)	1.19 (1.12, 1.26)	-	-
Other			
Erectile dysfunction (Abdul Sultan et al., 2017)	-	1.31 (1.24, 1.40)	-
Joint replacement (Kuo et al., 2018)	1.14 (1.05, 1.22)	-	-
Fractures (Sultan et al., 2018)	0.97 (0.92, 1.02)	-	-

HR: Hazard ratio; CI: Confidence interval

2.5 Management of gout

Various organisations had published guidelines on the management of gout flares and chronic gouty arthritis. The British Society of Rheumatology (BSR) published guidelines in 2007 (Jordan et al., 2007b) and later updated the guidelines in 2017 (Hui et al., 2017). The European League Against Rheumatism (EULAR) published guidelines in 2006 (Zhang et al., 2006a) and updated them in 2016 (Richette et al., 2017). The American College of Rheumatology (ACR) published guidelines in 2012 (Khanna et al., 2012a, Khanna et al., 2012b) and updated in 2020 (FitzGerald et al., 2020).

The more recent guidelines (BSR 2017, EULAR 2016 and ACR 2020) emphasised the need to educate patients to understand the causes and consequences of gout and hyperuricaemia, the importance of ULT, associated comorbidities, and required lifestyle changes.

2.5.1 Treatment of gout flares

Treatment of flares aims to rapidly reduce pain and swelling. Commonly prescribed pharmacological treatments are NSAIDS, colchicine, and corticosteroids. Choice of treatment is dependent upon comorbidity, risk of side effects, and patient preference.

NSAIDS are the most frequently prescribed medication for flares (Roddy et al., 2010). All guidelines recommend quick-acting NSAIDS at full dose, co-prescribed with gastro-protective drugs. Several RCTs have demonstrated the efficacy of NSAIDS in the treatment of flares (Khanna et al., 2014). There is little difference in anti-inflammatory effect between NSAIDS and there is no evidence any one NSAID is superior to another (Roddy et al., 2013). NSAIDS are prescribed with caution in patients with vascular diseases, impaired renal or liver function, and gastro-intestinal problems (for example, ulceration and haemorrhage) (NICE, 2019).

Low dose colchicine, the second most used drug (Roddy et al., 2010), is an effective treatment that rapidly reduces pain and inflammation. Despite its effectiveness, adverse events of diarrhoea, nausea and vomiting are common (Terkeltaub et al., 2010, Ahern et al., 1987).

A RCT compared the effect of naproxen versus low-dose colchicine on treating flares. No difference in change of pain scores over seven days were found between the two treatments although, naproxen use had fewer side effects, less analgesic use, and lower overall cost (aggregate of drug, GP, nurse, emergency GP, A&E and intervention costs) than colchicine use (Roddy et al., 2019).

Corticosteroids are not often used in primary care (Roddy et al., 2010) but they are an alternative in patients who have contraindications to NSAIDS and colchicine. In RCTs, prednisolone, an oral corticosteroid, was equally effective as NSAIDS in reducing pain scores (Janssens et al., 2008, Man et al., 2007). More recent guidelines (ACR 2012, BSR 2017, and EULAR 2016) recommended the use of combination therapy of NSAIDS, colchicine with corticosteroids when monotherapy was not sufficient to resolve a severe flare or flares affected were polyarticular.

Similarly, analgesics such as paracetamol and codeine are not often prescribed to treat flares (Roddy et al., 2010) but are recommended to be used as clinical adjuncts if pain is not adequately controlled when taking NSAIDS or colchicine (Jordan et al., 2007b).

2.5.2 Urate-lowering drugs

Gout is a consequence of crystal deposition due to hyperuricaemia. The plausible treatment of gout is to lower SU level below the urate saturation threshold of 380µmol/L (Roddy et al., 2013, Seegmiller, 1965, Loeb, 1972). This allows deposited MSU crystals and tophi to dissolve away and prevent new crystals from forming, thus preventing flares. A patient is considered 'cured' of gout if they are crystal and tophi free and had no flares, typically after two years of treatment (Roddy et al., 2013). The management guidelines for ULT are summarised in Table 2.2.

The treat-to-target concept is the cornerstone of ULT and has been adopted across most guidelines. EULAR 2006 and 2016 guidelines recommend SU level should be lowered to a target of ≤360µmol/L, with the 2016 guidelines stating a lower target of <300µmol/L may be needed in those with severe gout. BSR 2007 opted for a stricter target of <300µmol/L which was upheld in the 2017 guidelines, but with a further recommendation that once the patient

is in clinical remission, the target can be relaxed and SU maintained below ≤360µmol/L. This target allows SU level to be well below the saturation point allowing for fluctuations in SU level without increasing the risk of flares. Similarly, ACR 2012 opted for target SU level ≤360µmol/L but also recommended a lower target of <300µmol/L that may be needed to rapidly improve signs and symptoms in those with severe gout.

Xanthine oxidase inhibitors target the metabolism of purine by preventing the degradation of hypoxanthine to uric acid, thus lowering SU level. Allopurinol is the most commonly used xanthine oxidase inhibitor in ULT. Another xanthine oxidase inhibitor febuxostat, was approved by NICE in 2009 for use in patients who are intolerant of allopurinol.

Uricosuric drugs raise excretion of uric acid in the urine thus lowering SU level. Guidelines recommend using either benzbromarone, sulfinpyrazone and probenecid. Off-label drugs with mild uricosuric properties for gout are also considered in patients with comorbidities. These drugs include fenofibrate and losartan for the treatment of hyperlipidaemia and hypertension, respectively. Recently, a new uricosuric drug, lesinurad, co-prescribed with a xanthine oxidase inhibitor has marketing authorisation for treating hyperuricaemia in adults with gout provided the standalone xanthine oxidase inhibitor failed to lower SU level to target, however the drug does not have approval for use in the UK from the National Institute for Health and Care Excellence (NICE, 2018b).

The initial phases of ULT and titration increases the risk of flares. Although such flares are unwanted, they are an indication that SU level is decreasing and crystal dissolution occurring, and are markers of successful treatment (Roddy et al., 2013). To prevent such flares, anti-inflammatory prophylaxis can be co-prescribed with ULT. Guidelines recommend to either prescribe low dose colchicine or NSAIDS if not contradicted.

Uricosurics are uncommonly used in the UK and account for <1% of patients prescribed gout treatment in primary care practices (Annemans et al., 2008). Scarce use of uricosurics may be because allopurinol is an effective drug, allopurinol intolerance is uncommon, and febuxostat is available as an alternative when allopurinol is not tolerated. Probenecid and benzbromarone are unlicensed drugs and are only prescribed by rheumatologists within secondary care. Uricosurics should not be prescribed for over-producers of urate or in CKD, and benzbromarone carries a severe risk of hepatotoxicity (Hui et al., 2017).

Checking whether patients reach and maintain SU target, particularly during ULT initiation and titration, requires frequent SU monitoring. BSR 2007 guidelines state SU level should be monitored monthly during allopurinol titration and then yearly after SU target has been obtained. However in practice, measuring SU level is variable; an audit of a UK primary care practice found only 22% of patients with gout had SU measured in the last year, with that figure rising to 34% amongst allopurinol users (Cottrell et al., 2013).

Table 2.2: Comparison of guidelines in the long-term management of gout

BSR 2007 (Jordan et al., 2007b)	BSR 2017 (Hui et al., 2017)	EULAR 2006 (Zhang et al., 2006a)	EULAR 2016 (Richette et al., 2017)	ACR 2012 (Khanna et al., 2012a)	ACR 2020 (FitzGerald et al., 2020)
Who to start treatment in					
No comorbidities and had first flare, consider ULT after recurrence of flares within 1 year In patients with comorbidity, consider ULT: • Visible tophi • Renal insufficiency • Uric acid stones • Prescribed diuretics	ULT offered to all patients & particularly advised in those with: • ≥2 flares in a year • Tophi • Chronic gouty arthritis • Joint damage • Renal impairment • History of urolithiasis • Diuretic use • Young age	Severe established gout: • Recurrent flares • Gouty arthropathy • Tophi • Radiographic changes of gout • Multiple joint involvement • Uric acid nephrolithiasis	Considered and discussed in all patients, particularly if: • ≥2 flares per year • Tophi • Urate arthropathy and/or renal stones • Young age <40 years • SU level >480µmol/L • Renal impairment • Hypertension • Ischaemic heart disease • Heart failure	ULT indicated in: • ≥2 flares annually • Tophi • CKD stage 2 or worse • Past urolithiasis	ULT indicated in: • ≥2 flares annually • ≥1 subcutaneous tophi • Radiographic damage • Previously experienced >1 flare but had infrequent flares (<2 flares per year) • First flare and CKD stage ≥3, SU >9mg/dL, or urolithiasis
SU level					
• Target and maintain at <300µmol/L • Target SU level should be reached within 4 weeks • Measure SU every 3 months for the first year, then annually including creatinine level	• Target <300µmol/L • Maintain <360µmol/L once SU level is stable	• Target and maintain at ≤360µmol/L	• Target and maintain at <360µmol/L • <300µmol/L in severe gout (tophi, chronic arthropathy, frequent flares) • <3mg/dL not recommended in the long term	• Minimum target ≤360µmol/L • Lower target of <5mg/dL may be needed to improve signs and symptoms of gout • Treatment is indefinite	• Target and maintain at <6mg/dL
When to start treatment					
1-2 weeks after flare has ended	• After flare has ended • Can start during a flare if appropriate timing cannot be found	Not stated	Not stated	Can start during a flare provided anti-inflammatory medication has already been initiated	Can start during a flare

BSR 2007 (Jordan et al., 2007b)	BSR 2017 (Hui et al., 2017)	EULAR 2006 (Zhang et al., 2006a)	EULAR 2016 (Richette et al., 2017)	ACR 2012 (Khanna et al., 2012a)	ACR 2020 (FitzGerald et al., 2020)
Allopurinol					
First line drug • Start at 50-100mg/day • Increase by 50-100mg every few weeks until target SU level reached • Maximum dose 900mg • Lower dose according to impaired renal function • Rare life-threatening side effects • Co-prescribe with benzbromarone to reduce SU level further	First line drug • Start at 50-100mg/day • Increase by 100mg every 4 weeks until target SU level reached • Maximum dose 900mg • Increase by 50mg in patients with renal impairment • Associated with rare side effects • Should not be given to patients with a positive screening for HLA-B*5801 • Co-prescribe with uricosuric if target SU level not reached	Appropriate long term ULD • Starting dose of 100mg/day • Increase by 100mg every 2-4 weeks if needed • Adjust dose in renal impaired patients • May cause life-threatening AHS	First line ULT • Start at low dose 100mg/day • Titrate upwards by 100mg every 2-4 weeks until target SU level reached • Adjust dose according to renal function • Co-prescribe with uricosuric if allopurinol alone is unsuccessful in reaching SU target	First line drug • Start at ≤100mg OD • Start at 50mg in CKD stage 4 or worse • Titrate every 2-5 weeks • Maximum dose 800ng/day • Monitored every 2-5 weeks during titration • Once SU level achieved, monitored every 6 months • Risk of AHS • Prescribe in patients intolerant to febuxostat • If target SU level not met, co-prescribe with one uricosuric as second line approach	First line drug • Starting dose ≤100mg/day, and at lower doses for those with CKD stage ≥3 • Subsequent titration • In patients with prior allergic response to allopurinol and cannot be treated with other ULT, allopurinol desensitisation is an option • Screen for HLA-B*5801 in patients of Southeast Asian descent and African American patients
Febuxostat					
	Second line drug in patients • Contraindicated to allopurinol • Renal impairment that prevents increased allopurinol dose to reach SU level • Starting dose 80mg/day, increase after 4 weeks to 120mg/day if necessary to reach target SU level		Prescribe febuxostat if patients: • Cannot achieve SU target • Intolerant to allopurinol, especially in patients with impaired renal function	First line drug • Upward titration • Maximum dose 80mg/day • Prescribe in patients intolerant to allopurinol • If target SU level not met, co-prescribe with one uricosuric as second line approach	Second line drug • Cannot achieve target using allopurinol • Starting dose ≤40mg/day and subsequent titration • Not to be prescribed in patients with a new or history of cardiovascular event

BSR 2007 (Jordan et al., 2007b)	BSR 2017 (Hui et al., 2017)	EULAR 2006 (Zhang et al., 2006a)	EULAR 2016 (Richette et al., 2017)	ACR 2012 (Khanna et al., 2012a, Khanna et al., 2012b)	ACR 2020 (FitzGerald et al., 2020)
Uricosurics					
Prescribed in patients: • Intolerant to allopurinol • Under-excretes urate • Not over producers of urate • With normal renal function, prescribe sulphinpyrazone or probenecid • With mild-to-moderate renal insufficiency, prescribe benzbromarone	Prescribed in patients: • Intolerant to allopurinol and febuxostat • With normal-mild impaired renal function, prescribe sulfinpyrazone or probenecid • With mild-moderate impaired renal function, prescribe benzbromarone • Contraindicated in those with urolithiasis or severe renal impairment	Uricosuric drugs are an alternative: • With normal renal function, prescribe sulfinpyrazone or probenecid • with mild-to-moderate renal insufficiency , prescribe benzbromarone • Contraindicated in patients with urolithiasis	Uricosuric drugs are an alternative: • If target SU level not reached on allopurinol alone • If intolerant to allopurinol • Benzbromarone is more potent than probenecid • Benzbromarone prescribed in patients with renal impairment	Prescribed in patients: • Intolerant to allopurinol or febuxostat • With normal renal function, prescribe probenecid • Contraindications in patients who overproduce urate or have a history of urolithiasis • Consider fenofibrate and losartan	Alterative in patients who cannot have allopurinol or febuxostat: • Prescribe low dose probenecid with subsequent titration
Prophylaxis					
Following initiation of allopurinol or uricosuric drugs, consider: • Colchicine 0.5mg BD for maximum 6 months • NSAID or coxibs in those intolerant of colchicine, provided no contraindications, for a maximum of 6 weeks	Following ULT initiation or up-titration, consider: • Colchicine 500µg BD up to 6 months • Low dose NSAID or coxibs with gastro-protection if there are no contraindications • Lose-dose colchicine adjusted for renal-disease is safer than low dose NSAID	During the first months of ULT, consider: • Colchicine 0.5-1mg/day • NSAID with gastro-protection if required • Benefits and harms need to be considered for both drugs	In the first 6 months of ULT, consider: • Colchicine 0.5-1mg/day • Reduce dose in patients with renal impairment • Potential neurotoxicity and/or muscular toxicity in renal impairment or statin treatment • Alternatively use low dose NSAID	After initiating ULT, consider: • Colchicine 0.5-0.6mg/day • Low dose NSAID with gastro-protection • Prescribe prednisolone If NSAID and colchicine not tolerated/indicated • Continue for >6 months, or 3/6 months once SU target reached and absence/presence of tophi	Strongly recommends prophylaxis for 3-6 months • Colchicine • NSAIDS • Prednisolone

CKD: Chronic kidney disease; Coxibs: Cyclooxygenase II inhibitors; NSAIDS: Non-steroidal anti-inflammatory drugs; SU: Serum urate; ULT: Urate-lowering therapy

Modification of lifestyle and diet, and the management of comorbidities are adjunct measures in ULT. Patient education is required to improve diet and lifestyle. Guidelines recommend that patients should reduce consumption of alcohol and foods rich in purine and fructose, where excessive, although the evidence that these are effective management strategies is sparse.

2.5.3 Indications for urate-lowering therapy

Indication for ULT differs across guidelines (Table 2.2). Early guidelines (BSR 2007, EULAR 2006) recommended that allopurinol should be used in patients with established gout. Despite these guidelines, uptake of ULT is poor. Analysis of primary care medical records from CPRD of patients with incident gout in 1997-2010 found 44% of patients at diagnosis had indications for ULT of whom, <1% were prescribed ULT. At 5 years from diagnosis, 86% of patients were indicated for treatment but only 30% were prescribed ULT (Kuo et al., 2014). This builds upon previous estimates that between 25% and 56% of patients were on ULT (Kuo et al., 2015b, Roddy et al., 2007b, Cottrell et al., 2013, Annemans et al., 2008). The majority of patients on ULT would be prescribed allopurinol (Cottrell et al., 2013).

ULT prescription was more common in people with indications for ULT (recurrent flares, tophi, CKD, diuretic use etc.) apart from urolithiasis. Other factors associated with receiving ULT were male sex, higher deprivation, and higher Charlson comorbidity score (a summary measure of 17 diagnostic categories representing a person's severity of health) (Charlson et al., 1994, Deyo et al., 1992, Kuo et al., 2014).

Clarson et al. (2017) investigated factors influencing initiation of allopurinol. Only 40% of patients were prescribed allopurinol. Median time to first allopurinol prescription was 8 months. Similarly to Kuo et al. (2014), CKD, diuretic use, and tophi were associated with increased likelihood of allopurinol prescription. In addition, patients with urolithiasis, two or

more gout consultations in the preceding 12 months, and who were overweight were more likely to be prescribed allopurinol. Conversely, males and those with increased Charlson comorbidity score were less likely to be prescribed allopurinol. Charlson score at diagnosis was higher in those who received allopurinol suggesting that people with comorbidity were more likely to receive allopurinol.

Allopurinol is recommended across all guidelines to be taken daily starting at the low dose of up to 100mg and to be titrated upwards by 100mg every 2-4 weeks until target SU level is reached. This approach is not applied in most allopurinol users. Cottrell et al. (2013) found only 62% of allopurinol users were correctly prescribed a starting dose of 100mg, with 32% prescribed 300mg, and titration was not performed in 57% of allopurinol users. Annemans et al. (2008) found the most common average daily dose was 200-300mg in 63% of patients followed by 50-100mg in 21% of patients; only a minority of patients (2%) were prescribed >300mg.

2.5.4 Adherence to allopurinol

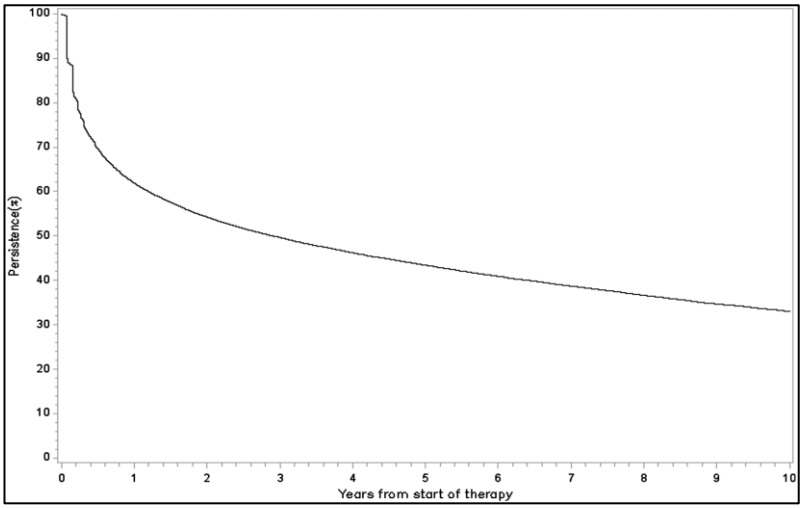
Allopurinol treatment is intended to be life-long however it is well known prescribing of allopurinol is suboptimal.

Kuo et al. (2015b) evaluated the proportion of days (covered) (PDC) patients were prescribed ULT during each year of follow-up from diagnosis within CPRD. Adherence to treatment (PDC ≥80%) improved over time from 28% in 1997 to 39% in 2012, although overall adherence was still poor.

Also within CPRD, Scheepers et al. (2018) published a more comprehensive study evaluating persistence and adherence to allopurinol among patients diagnosed between 1987-2014 with gout. Median time to non-persistence (defined as no allopurinol prescription for at least 90

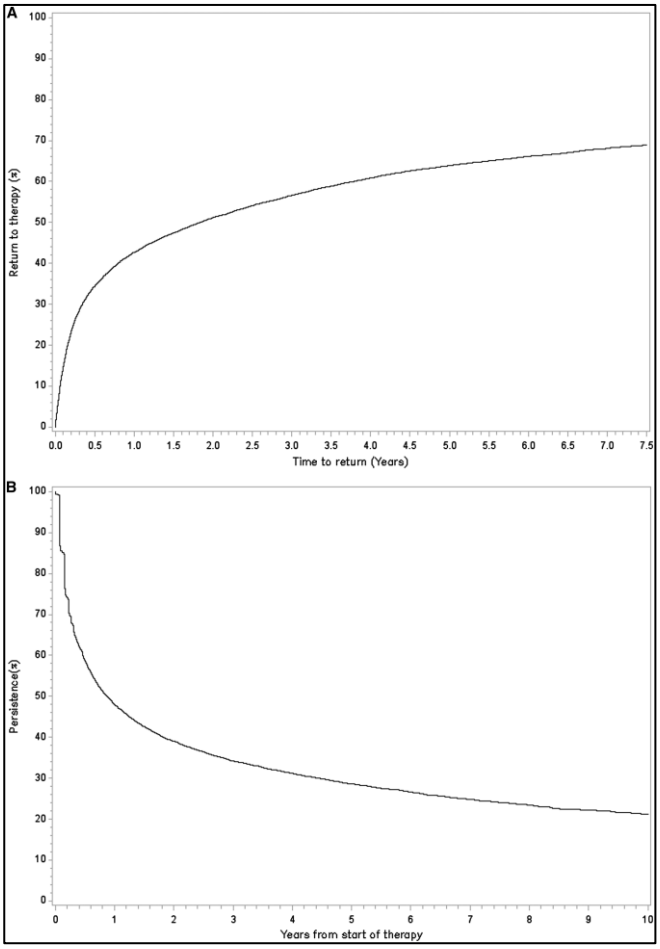
days) was 1,029 days. Non-persistence increased over time (Figure 2.3) with 39% stopping allopurinol at 1 year and 57% at 5 years. PDC was moderate with median 67% of patient observation time had a prescription for allopurinol. Of patients who stopped allopurinol, 57% restarted treatment and median time to restarting treatment was 643 days (Figure 2.4). Non-persistence (52%) and PDC (mean 49%) was poor in this group of patients (Scheepers et al., 2018).

Figure 2.3: Kaplan-Meier curve for persistence (90-day gap) to treatment with allopurinol medication in the total study sample



Scheepers et al, Medication adherence among gout patients initiated allopurinol: a retrospective cohort study in the Clinical Practice Research Datalink (CPRD), Rheumatology, 2018, vol. 57, issue no. 9, pgs. 1641-1650, by permission of Oxford University Press

Figure 2.4: Probability of restarting treatment over time



Scheepers et al, Medication adherence among gout patients initiated allopurinol: a retrospective cohort study in the Clinical Practice Research Datalink (CPRD), Rheumatology, 2018, vol. 57, issue no. 9, pgs. 1641-1650, by permission of Oxford University Press

A systematic review conducted in 2013 evaluated adherence to allopurinol (De Vera et al., 2014). The systematic review found non-adherence was high in other countries (US, New Zealand, Israel, Netherlands and Spain); PDC ranged from 17-36% in three studies which were lower than 39-57% in the UK studies.

Scheepers et al. (2018) found factors associated with non-persistence and non-adherence (PDC <80%) were female sex, smoking, and greater number of primary care consultations. Whereas factors associated with persistence and adherence of allopurinol use were older age, higher BMI, ex-smoking, alcohol consumption, more recent initiation of allopurinol, initiation of treatment within 90 days of diagnosis, prescribed anti-hypertensive treatment, diagnosis

of dementia, diabetes, and dyslipidaemia. This is comparable to the systematic review which found among the six included studies that older patients and those with hypertension were consistently reported to have higher adherence. Furthermore, occurrence of flares, absence of tophi and incident gout were associated with lower adherence (De Vera et al., 2014). An analysis of Swedish health records further found reduced kidney function was associated with non-adherence (Dehlin et al., 2017).

2.6 Summary

This chapter introduced gout, the condition of interest for this thesis. Epidemiological research has identified risk factors for gout and long-term consequences of gout. Organisations had published how gout should be diagnosed and managed.

Since the main aim of this thesis is to investigate the effectiveness of allopurinol using observational EHR data, the next chapter will describe a narrative review which will examine how published observational studies of the effectiveness of allopurinol have accounted for confounding in their analysis.

3 A narrative review of observational studies evaluating the effect of allopurinol in gout

3.1 Objectives

The objective of this chapter was to perform a narrative review of observational studies assessing the effects of allopurinol in the treatment of gout, and to establish the range of study designs and statistical methods used to control for confounding variables in the published literature.

3.2 Literature search strategy and data extraction

A literature review protocol (Appendix A) was developed specifying details relating to search strategy and study selection process in order to optimise identification of relevant studies. Throughout the literature review process, assistance was sought from the designated departmental systematic literature review support team when needed, and attendance of the departmental systematic literature review workshop.

Search terms were compiled from guidance from the supervisory team and the search was conducted in five databases (AgeLine, CINAHL, EMBASE, MEDLINE, and Web of Science) from the date of inception to 4th October 2014, for full-text articles published in English. The search terms 'Gout' and 'Allopurinol' or equivalent e.g., 'podagra' and brand names for allopurinol, were searched for in the title and abstract; the full search strategy for MEDLINE is given in Appendix B. The search strategy was broad to include any outcome and any comparator group including different drugs, dosage of allopurinol, or a non-pharmacological intervention.

The inclusion criteria were:

- articles evaluating the effectiveness of allopurinol (versus any comparator group) on any outcome in patients with gout
- observational studies of any design including: cross-sectional, cohort, case-control, and case-crossover studies

The exclusion criteria were:

- studies without a comparator group
- randomised controlled trials
- studies undertaken exclusively in children aged <18 years
- non-published material/grey literature

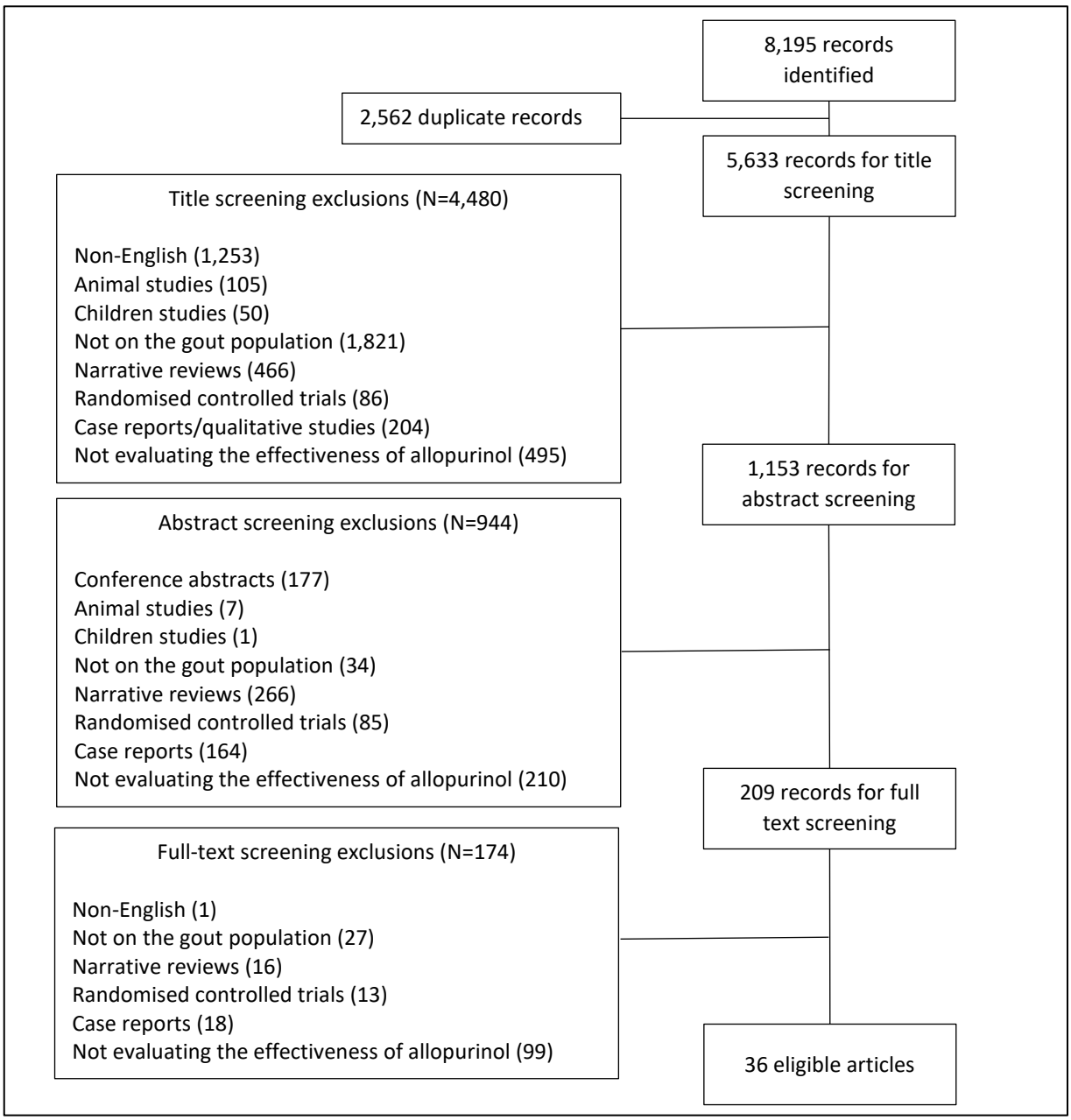
Screening of search results (titles, abstracts, and full texts) using pre-defined inclusion and exclusion criteria, was performed independently by one reviewer (T Rathod-Mistry (TRM)) with a subsample screened by two independent reviewers (M Blagojevic-Bucknall (MBB)/E Roddy (ER)). Originally, articles not published in English were included however approximately half of eligible titles with no abstracts (n≅200) came from non-English journals. Due to time constraints, articles not published in English were excluded. Further articles were excluded if the entire study sample under analysis did not have gout, for example studies with mixed populations where people with hyperuricaemia but not gout were included alongside people with gout, or if there was contamination of treatment groups such that the exposure group contained participants taking any anti-gout medication, for example allopurinol or probenecid. To ensure eligible articles were not missed from the search strategy, the references of the eligible articles were manually screened. Reviewers ER and MBB screened titles that had no abstract; any disagreements were resolved by retrieving full text and discussion.

Data extracted from eligible articles were: study design that may have been used to control for confounding; setting and source of data; sample size used in analysis and length of follow-up period; gout outcomes; how exposure to allopurinol, dosage, and treatment duration were recorded and analysed; statistical models used to adjust for confounding variables; how unmeasured confounding and missing data were handled. Treatment effect estimates were not extracted as meta-analysis and meta-regression were not performed as the aim was to conduct a narrative review of the methods that were used to control for confounding. Extracted data was entered into a pre-tested Excel spreadsheet devised with supervisors' guidance. Data extraction was performed independently by TRM. Uncertainties concerning data extraction were discussed with ER and MBB. The following characteristics of the included studies were described narratively: study design, methods used to adjust for confounding, exposure to allopurinol, data sources, and gout outcomes.

3.3 Description of included articles

The search of databases yielded 8,195 records, 2,562 of which were duplicates. The exclusion criteria were applied to the remaining 5,633 titles, and subsequently to 1,153 abstracts, and then 209 full-text articles. Reasons for exclusion can be found in Figure 3.1. Thirty-five studies met the inclusion criteria. Manual reference checks of eligible articles yielded no further articles. One known eligible article was not identified by the database search as it was published ahead of print and was included. A total of 36 articles were therefore reviewed.

Figure 3.1 Systematic review flow chart



Appendix C summarises the data extracted of the 36 articles. Study designs used to estimate the effects of allopurinol were:

- Cohort studies (n=20)
- Cross-sectional studies (n=11)
- Case-control studies (n=3)
- Case-crossover studies (n=2)

3.4 Accounting for confounding in study design

Cohort study

Cohort design, which has the potential to infer causality as the temporal ordering of treatment and outcome is known, was the most popular study design. Two studies controlled for covariates via the study design. Kok et al. (2014) created a matched cohort between allopurinol users and non-users matching on seven baseline variables to make the two treatment groups similar. Despite initially finding 12,563 allopurinol users, only 20% were matched to a non-user. The number of valid matches was limited by its sample size and the number of covariates to match on. The incident user cohort study design of Dubreuil et al. (2015) improved upon the previous study by using propensity score (PS) matching that was conditional on a greater number of covariates (24), and more allopurinol users were matched to non-users (85%). One cohort study was nested within a case-control study design although no further details were given on how matching was performed (Alvarez-Nemegyei et al., 2005). The other 16 studies did not use matching and simply assembled a cohort of patients with gout to follow-up.

Cross sectional study

Cross-sectional was the second most commonly used study design. As exposure to allopurinol and outcome is collected at a single time-point or over a short period of time, causality cannot be inferred as the temporal order between exposure and outcome is unknown, therefore only an association can be estimated.

Case-control study

A common misconception within case-control studies is that matching removes confounding of the matched covariates. Matching ensures outcome groups, rather than treatment groups, are comparable on the matching covariates however this may underestimate the treatment

effect as cases and controls may become too similar with regards to the distribution of treatment. The primary reason to match on covariates is to improve precision in the effect estimates upon adjustment for the matched covariates (Pearce, 2016).Three studies used a case-control study design however none of these studies adjusted for the matching covariates in treatment effect estimation. Thanassoulis et al. (2010) case-control study was nested within a cohort study, with cases and controls matched on calendar day of admission to the cohort allowing equal follow-up periods between matched sets of cases and controls. Both Hutton et al. (2009) and Stamp et al. (2012) had matched on demographics, with Stamp et al. (2012) also matching on diuretics and renal function.

Case-crossover study

Two studies, Zhang et al. (2012) and Neogi et al. (2014), used a case-crossover study design. Case-crossover studies are typically used to investigate intermittent exposures with short-term effects on the risk of an acute outcome. The key advantage of case-crossover studies is that they allow cases to serve as their own control with individuals crossing between periods of exposure and non-exposure, thus eliminating confounding of individual characteristics that remain constant over time e.g., sex. However, within-individual confounding can remain when individual characteristics change over time e.g., medication use, thus requiring further analysis. Case-crossover studies are ideally used for a short follow-up period to minimise time-varying confounding.

3.5 Sources of data

Data was collected from three sources: national administrative databases of routinely collected electronic health records (EHR) (n=9); medical record data collected from individual

or few rheumatology clinics and hospitals (n=22); recruiting patients from the general population (n=5).

Administrative databases

National administrative health care databases contain EHR for claims for services, procedures, and drugs for health insurance programs. Studies based on such data were most frequently based in the USA (Halpern et al., 2009, Hatoum et al., 2014, Kim et al., 2013b, Pandya et al., 2011), with three other studies conducted in Taiwan (Kok et al., 2014), Canada (Thanassoulis et al., 2010) and New Zealand (Stamp et al., 2012). Use of such data allow large longitudinal studies to be conducted; six studies had sample sizes ranged from 1,768 to 35,577 patients and median follow-up ranged from 0.2 to 5.25 years. Furthermore, such data allowed one study to evaluate the effect of allopurinol on a rare outcome, allopurinol hypersensitivity syndrome (Stamp et al., 2012). Although administrative databases can be largely representative of the general population, there is less detailed clinical information in the form of free text, and generalisability may be limited if health care systems only cover those with health insurance; such patients may be healthier, more affluent, and more likely to be employed (Schneeweiss and Avorn, 2005).

Two studies used data from The Health Improvement Network database (THIN) which hold records of 11.1 million patients from 562 primary care practices in the UK, covering 6.2% of the population. Dubreuil et al. (2015) and Rothenbacher et al. (2011) were able to analyse a large sample of 9,590 and 6,795 patients respectively however, the associated length of follow up was short (2.9 years and 6 months respectively).

Rheumatology clinics and hospitals

Twenty-two studies either reviewed medical records or recruited patients who underwent clinical assessments to collect data from rheumatology clinics or hospitals. These studies

tended to be small with sample sizes ranging from 31 to 1,288. Patients attending rheumatology clinics or hospitals may be more likely to have severe gout and more comorbidities, therefore results obtained from these studies may not be generalisable to the general population; on the other hand, diagnosis of gout is more likely to have been definite and may allow more covariates to be measured that may not be available in administrative databases.

General population

These studies recruited gout patients from the general population and collected data from questionnaires, medical record review and/or clinical assessment. Alternative study designs were used such as the case-crossover design used in two studies (Neogi et al., 2006, Zhang et al., 2012). These studies tended to be small with sample sizes ranged from 57 to 290 participants although two studies were quite large with up to 724 participants. A disadvantage of using questionnaire data is that responses may be inaccurate if participants do not respond with introspective ability and honesty.

3.6 Gout outcomes

The 36 eligible studies reported association of allopurinol with various outcomes. Shown in Table 3.1, outcomes considered could be grouped into seven categories with the most number of studies frequently reporting SU level (n=14), gout flare (n=6), and cardiovascular outcome (n=4). Although many of the key outcomes relevant to gout were considered, actual time to occurrence of outcome was only taken into account in four studies (SU level <360µmol/L; various cardiovascular diseases; gout flare; all-cause mortality).

Table 3.1: Outcomes measured

Outcome category	Number of studies evaluating outcome	Outcome (number of times analysed)
SU level	14	SU level (11); time to or attained SU level <360µmol/L (9); clearance or urinary excretion of urate/uric acid level (7)
Gout flare	6	Frequency of gout flares (3); occurrence of gout flare (3); time to gout flare (1)
Vascular outcome	4	Cardiovascular event or sudden death due to cardiac causes (1); heart failure re-admission or all-cause mortality (1); myocardial infarction (1). Time to cardiovascular outcome requiring hospitalisation (1); coronary heart disease (1); stroke (1), hypertensive heart disease (1), heart failure (1), other cardiovascular disorders (1)
Reactions to allopurinol use	5	Adverse event (1); allopurinol hypersensitivity syndrome (1); oxypurinol level (3);
Renal disease	5	Chronic kidney disease (1); clearance of creatinine (2); renal failure (1); Serum cystatin C concentration (1)
Tophi	3	Diameter of tophi (1); presence of tophi (2); time until tophi resolution (1); velocity of reduction in tophi (1)
Urolithiasis	2	Stone composition (11); stone formation (1);
Other	4	Hospital admission (1); MSK physical disability (1); time to all-cause mortality (1); urinary pH (1); urinary volume (1)

MSK: Musculoskeletal; SU: Serum urate.

3.7 Exposure to allopurinol

All 36 studies recorded allopurinol use at baseline therefore estimating the effect of initiating treatment. Five studies had taken into account the duration spent on allopurinol in different ways. Dubreuil et al. (2015) performed sensitivity analyses truncating follow-up at 1, 2 and 3 years of follow-up to address the potential of patients discontinuing treatment although that study did not report the number of patients who actually discontinued treatment. Kim et al. (2013b) censored follow-up time when patients discontinued allopurinol use which was accounted for in modelling effect of allopurinol, and also stratified analysis by the number of days (<30, 31-90, 91-120 days) spent on allopurinol. Mak et al. (2009) adjusted for number of

years allopurinol was taken for in a regression model. Thanassoulis et al. (2010) compared outcome amongst non-users to those on allopurinol for less or more than 30 days in regression modelling.

Eighteen studies modelled the effect of allopurinol dose on outcome in different ways. Five studies evaluated the effect of giving a higher or lower than recommended dose of allopurinol on outcome based on creatinine level (Dalbeth et al., 2006, Stamp et al., 2000, Stamp et al., 2011b, Stamp et al., 2012, Vazquez-Mellado et al., 2001). One study stratified analysis by dose (Perez-Ruiz et al., 1998). The remaining studies adjusted for dose in regression models or compared the distribution of dose between those who had or did not have outcome.

Zhang et al. (2012) and Neogi et al. (2014) were the only two studies to have collected information on allopurinol use every three months however, allopurinol use was not modelled as time-varying as conditional logistic regression was performed. Hatoum et al. (2014) reported the number of patients who changed treatment from allopurinol to febuxostat and vice versa however treatment was not modelled as time-varying. Six studies stated not accounting for adherence or compliance to allopurinol was a study limitation.

3.8 Adjustment for confounding variables in analysis

Apart from controlling for covariates via the study design (described in Section 3.4), covariates can also be adjusted for in subsequent analysis via the outcome model. Twenty studies did not adjust for confounding variables in their analysis whereas 13 studies did; three studies performed both unadjusted and adjusted analyses.

Unadjusted analyses

Statistical tests used to determine if an association between allopurinol exposure and outcome exists without adjusting for covariates were:

- Analysis of variance, Mann Whitney U test, Wilcoxon independent groups test, and T-test (n=17 studies)
- Chi-square test/Fisher's test (n=14 studies)
- Regression models, more specifically linear regression, Cox regression and mixed effects linear model (n=3 studies)
- Likelihood ratio test (n=1 study)
- Mantel-Haenszel odds ratio (n=1 study)

The chi-square test/Fisher's test, Mann Whitney-u test, analysis of variance, and the likelihood ratio test only determine if a treatment effect between allopurinol exposure and outcome exists at a pre-defined significance level, typically at 5%. These statistical tests do not inform the magnitude and direction of the treatment effect. On the other hand, the t-test, regression models, and the Mantel-Haenszel odds ratio estimates the magnitude and direction of the actual treatment effect and the associated standard error. Treatment effect estimates were expected to be biased due to no adjustment for covariates.

Adjusted analyses

Types of regression models used to determine if an association between allopurinol exposure and outcome existed whilst adjusting for covariates were:

- Logistic regression (n=5 studies)
- Conditional logistic regression (n=4 studies)
- Cox regression (n=3 studies)
- Linear regression (n=3 studies)

- Poisson regression (n=1 study)

The disadvantages of using regression models are that the number of covariates that can be adjusted for may be limited by the study sample size, and model regression assumptions are imposed.

A wide range of covariates that were controlled for are listed in Table 3.2. The covariates encountered can be grouped into five general categories: demographics, medication use, comorbidities, health care utilisation, and lifestyle factors. The most commonly controlled confounding variables were age (n=13), sex (n=12), and SU level (n=9). The median number of confounding variables adjusted or matched on was seven, ranging between one and 29.

The majority of studies had at least controlled for demographics and comorbidities, with lifestyle factors rarely being adjusted for. Cardiovascular and renal diseases, either measured in the form of biomarkers, use of medications, or diagnosis, were often adjusted for in analysis. It is possible that adjusting for information on gout and healthcare utilisation may have indirectly adjusted for severity of gout and general health.

All studies but two adjusted for confounding variables at baseline with the assumption that covariates were time-invariant. This assumption may not hold especially if the follow-up period is long, as patients may have new diagnoses, severity of gout may worsen or lessen, or there may be changes in lifestyle and other covariates. Although the two case-crossover studies by Zhang et al. (2012) and Neogi et al. (2014) collected information on confounding variables every three months for a year, subsequent analysis did not take this into account.

All studies performed complete case analysis i.e., patients with missing covariate data were excluded from analysis, as their primary analysis; no studies looked at the sensitivity of missing data. Fifteen studies acknowledged treatment effects may be biased due to unmeasured confounding.

Table 3.2: Confounding variables controlled for via the study design or statistical analysis

Category (no. of studies)	Sub-category (no. of studies)	Confounding variables (no. of studies)
Demographics (n=14)	-	Age (13); sex (12); ethnicity (5); index date (3)
Medication use (n=8)	Gout medications (n=6)	NSAIDS (6); colchicine (2); benzbromarone (1); corticosteroids (1); coxibs (1); intra-articular steroids (1); intravenous steroids (1); opioids (1); oral steroids (1)
	Cardiovascular medications (n=8)	ACE inhibitors (4); ARBs (3); aspirin (2); beta-blockers (4); calcium channel blockers (2); diuretics (8); fibrates (2); hydrochlorothiazide (3); losartan (2); statins (3); antiplatelet agents (1); anticoagulants (1); aldosterone antagonists (1);
	Other (n=1)	Insulin (1)
Comorbidities (n=13)	Heart disease (n=9)	Atrial fibrillation (1); cardiovascular disease (3); diseases of the heart (1); heart failure (1); hypertension (7); ischaemic heart disease (2); myocardial infarction (1); stroke (1) Biomarkers: C-reactive protein (1); erythrocyte sediment rate (1)
	Lipodystrophic disorders (n=8)	Disorders of lipid metabolism (1); dyslipidaemia (1); hyperlipidaemia (3) Biomarkers: Cholesterol (2); high density lipoprotein cholesterol (2); lipoprotein cholesterol (1); triglyceride (1)
	Renal disease (n=11)	Chronic kidney disease (2); chronic renal failure (2); diseases of the urinary system (1); renal function (1); renal stones (1); stage of chronic kidney disease (1); stage of renal function (1); uremia (1); confidence to keep SU under control (1) Biochemical tests/biomarkers of renal function: Creatinine clearance (2); glomerular filtration rate (2); serum creatinine level (1)
	Other (n=8)	Body mass index (4); COPD (1); Charlson comorbidity index (4); comorbidity index (1), diabetes (6); gastric ulcer (1); non-traumatic joint disorders (1); obesity (1) Biomarkers: Albumin level (2)
Gout characteristics	Gout (n=9)	Duration of gout (1); tophi (2); year of gout diagnosis (1) Biomarkers: Serum urate level (9)
Healthcare utilisation (n=6)	-	Cardiac procedures (1); emergency room visits (1); hospitalisations (1); prescription drugs (1); office visits (1); primary care practice visits (4); gout consultations (1) rheumatology visits (1); speciality of prescribing physician (1)
Lifestyle factors (n=3)	-	Alcohol consumption (3); cherry intake (1); purine intake (2); smoking status (1)

ACE: Angiotensin-converting enzyme; ARBs: Angiotensin II receptor blockers; COPD: Chronic obstructive pulmonary disorder; NSAIDS: Non-steroidal anti-inflammatory drugs.

3.9 Implications for the PhD project

This narrative review systematically identified and assessed 36 observational studies that had evaluated the effect of allopurinol on gout outcomes.

3.9.1 Strengths and limitations

The key strengths of this literature review were that a systematic search was performed and in addition the references of eligible articles were manually checked to identify all observational studies evaluating the treatment effect of allopurinol in gout. Additional reviewers had screened a subsample of the search results and reviewed data extraction separately.

It is possible that eligible articles could have been missed if they were published ahead of print but not identified by the database search; regression analysis adjusted for allopurinol but was not reported in the abstract due to lack of association with outcome; treatment effectiveness analysis was performed in a different population, say hyperuricaemia, and as a sensitivity analysis restricted analysis to patients with gout but this not reported in the abstract. Furthermore, some articles may have been eligible from the 1,253 articles published in languages other than English that were excluded. Cross validation was not performed, potentially increasing the chance of missing eligible articles and data extraction errors.

This literature review highlighted possible improvements that can be made in the approach taken to estimate effect of allopurinol, and these are now briefly summarised.

3.9.2 Time-varying confounding

Control for confounding could be improved in the majority of studies. Under half of the 36 eligible studies controlled for baseline confounding; the remaining studies ignored confounding issues thus potentially yielding biased treatment effect estimates. Confounding was mostly controlled for using regression models. Only one study used PS matching to create comparable treatment groups. No studies attempted to model for time-varying treatment or covariates although few studies acknowledged that ignoring time-varying allopurinol use, adherence and compliance to treatment were limitations of the study. Fifteen studies acknowledged that residual confounding was likely owing to unmeasured confounding variables and lack of randomisation however, no attempt was made to revise the estimated treatment estimates to take this into account. Across studies, a wide range of covariates were considered however within studies, the median number of confounding variables adjusted or matched on was seven and ranged between one and 29.

As explained in Section 2.5.4, adherence to allopurinol is poor and changes over time. In addition to time-varying treatment, time-varying confounding will also exist. Indications for allopurinol are likely to vary over time as flares become more frequent and may last longer and affect more joints, chronic gouty arthritis, tophi and bone erosion may develop in untreated patients.

None of these studies have assessed such time-varying nature of indication for allopurinol nor its effect on outcome.

3.9.3 Electronic health records

Large EHR databases are increasingly being used for health research (Schneeweiss and Avorn, 2005). Several studies analysed data from administrative databases for health insurance

programs mostly based in the USA. Only two studies had used primary care data from THIN that represents 6.2% of the UK population. As stated in Section 1.1, the advantages of using such data are that their large size is generalisable to the population of interest and represents routine primary care.

No study had evaluated the effect of allopurinol on gout outcomes using primary care medical records from the Clinical Practice Research Datalink (CPRD). CPRD is comparable with THIN as both databases recruit practices that use the management software Vision (that records medical record data) thus there is considerable overlap in their patient pools. CPRD is representative of 8% of the UK population and allows linkage to more secondary databases such as Hospital Episode Statistics data. Since gout is primarily managed in primary care it was surprising to not see more studies using this data source.

3.9.4 Allopurinol exposure

Exposure to allopurinol was measured in different ways. Studies had either compared the effect of allopurinol use vs. non-use, amount of time taking allopurinol, or differing allopurinol doses.

In practice, the recommended starting dose of allopurinol is 100mg daily which is gradually titrated upwards by 100mg every 2-4 weeks in order to lower SU levels below 360µmol/L (Zhang et al., 2006a); dosages greater than 300mg are needed to meet this target in over half of patients (Khanna et al., 2012a). Furthermore, allopurinol dosage needs to be adjusted for in patients with renal impairment (Khanna et al., 2012a). In practice, allopurinol dose is often not titrated and patients may remain on their starting dose (Cottrell et al., 2013). As stated above in Section 3.9.2, in addition to dose potentially changing over time, indications for treatment, and adherence to treatment may also change over time.

Definition of allopurinol will be dependent upon how in practice one could model time-varying treatment in the presence of time-varying confounding.

3.9.5 Outcomes

Gout is widely known to be associated with comorbidity particularly with vascular and renal diseases. Unsurprisingly, studies evaluating the effectiveness of allopurinol have considered a wide range of outcomes however only mortality and gout flares have been considered so far in studies undertaken using data from primary care medical records. Primary care EHR provide an opportunity to observe a wide range of outcomes that can take many years after diagnosis to develop. Therefore, a high quality observational study using data from a large primary care EHR database, such as CPRD, to evaluate the effect of allopurinol on such outcomes is needed. A retrospective cohort study is the ideal observational study design to evaluate effectiveness of allopurinol using EHR data as it would allow calculation of incidence rates or the relative risk of multiple outcomes between the treatment groups. The advantage of using EHR data is that the outcomes of interest are routinely collected and available for analysis; prospectively collecting data is not feasible due to very long follow-up required for some outcomes, such as death, to occur.

Outcomes of interest in this thesis are target SU level, mortality, vascular and renal diseases, health care contacts for gout in primary and secondary care, and joint replacement. Reasons why these outcomes were chosen, and limitations of existing studies identified from this review are described below.

Target SU level

As explained in Section 2.3.1, hyperuricaemia is a primary risk factor for gout and reducing SU level is the most important treatment objective (Section 2.5.2). Thus, reaching target SU level

≤360µmol/L, as recommended by EULAR (Zhang et al., 2006a), was considered the primary outcome.

This review identified SU level as the most frequently evaluated outcome. Eight studies had evaluated the effect of allopurinol on attaining target SU ≤360µmol/L. Studies had found the median time to achieve target was 37 weeks (Lim et al., 2012), target SU level was reached by 29-41% of allopurinol users by 6 months and 48-72% by 12-24 months (Lim et al., 2012, Hatoum et al., 2014), and attaining target was more commonly achieved in those taking allopurinol than non-users (Dalbeth et al., 2006, Roddy et al., 2007b) and in those with higher doses (Pandya et al., 2011).

No high quality study exists comparing the effects of allopurinol use with non-use on the time to achieve target SU level using primary care EHR databases. Only cross-sectional studies had compared allopurinol use vs. non-use but cannot infer causal effects (Dalbeth et al., 2006, Dalbeth et al., 2012, Roddy et al., 2007b), whilst cohort studies had used febuxostat or uricosuric drugs as the comparator group (Hatoum et al., 2014, Perez-Ruiz et al., 1998, Stamp et al., 2011b). Two studies that had used EHR from large administrative databases were based in the USA (Hatoum et al., 2014, Pandya et al., 2011), with the majority of studies conducted in non-UK rheumatology/hospitals; only one small cross-sectional study had recruited people from the general UK population (Roddy et al., 2007b). Few studies had adjusted for baseline covariates (Dalbeth et al., 2012, Hatoum et al., 2014, Pandya et al., 2011) however no studies had modelled time-varying treatment or covariates.

All-cause mortality

Gout is known to be associated with premature mortality (Lottmann et al., 2012) with common causes of death being vascular disease (particularly from coronary heart disease, stroke and heart failure), endocrine and metabolic diseases (mostly diabetes), and kidney

disease (Kuo et al., 2011). Interest is in whether allopurinol use is protective against premature morality in people with gout.

This review identified one high quality cohort study using data from THIN (a UK primary care EHR database) that had evaluated the effect of allopurinol on time to all-cause mortality and had shown allopurinol use was protective. Although this study had used PS matching to create comparable treatment groups on several covariates, information was lacking on how PS was performed and the assessment of comparability of treatment groups. To address the potential for allopurinol users to discontinue treatment, that study had used a suboptimal approach censoring patient follow-up at 1, 2, and 3 years after treatment initiation (Dubreuil et al., 2015), when a better approach would have been to model allopurinol use as time-varying.

Further investigation is needed whether modelling allopurinol as time-varying yields similar conclusions, and to perform statistical analysis in a more rigorous manner ensuring PS analysis performed well.

Vascular diseases

Gout has been shown to be associated with coronary heart disease and peripheral vascular disease among men and women, and associated with cerebrovascular disease among women (Clarson et al., 2015). A systematic review conducted in 2013 had shown there is limited evidence allopurinol reduces risk of cardiovascular disease (Fleeman et al., 2014) thus a need for higher quality studies is needed.

The second most commonly evaluated outcome as identified by the review in this chapter was vascular disease which covered a range of conditions such as myocardial infarction and stroke.

Two studies had used administrative databases of EHR to evaluate effectiveness of allopurinol. One matched cohort study found allopurinol users had an overall increased risk of any

cardiovascular event, and had separate increased risks for coronary heart disease, hypertensive heart disease and heart failure than non-users but no association was observed for stroke (Kok et al., 2014). Conversely, a case-control study nested within a cohort found allopurinol use had reduced odds of heart failure readmission/death than non-users (Thanassoulis et al., 2010). Despite using EHR data, median follow-up was short ranging from 2 to 5 years in the two studies. Two other studies had considered myocardial infarction and cardiovascular event or death however they were of cross-sectional design and did not adjust for any covariates.

No studies had evaluated effect of treatment on vascular disease including peripheral vascular disease using UK primary care EHR data especially modelling allopurinol and covariates as time-varying.

Renal disease

As seen in Section 2.3.6, the association of renal disease and gout is widely known to be bi-directional. Few studies had evaluated the effectiveness of allopurinol on renal disease using various outcome definitions, with one systematic review showing there is limited evidence allopurinol reduces risk of renal disease (Fleeman et al., 2014).

Five small studies based in rheumatology clinics or hospitals, had evaluated the effect of allopurinol on renal disease were identified in this review. Various definitions for renal disease were used including chronic kidney disease and creatinine clearance. Four of these studies had found no association between allopurinol use and renal disease (Alvarez-Nemegyei et al., 2005, Cheyoe et al., 2012, Perez-Ruiz et al., 1998, Perez-Ruiz et al., 2010). One study had found allopurinol users had worse renal function than non-users, and was the only study to use non-allopurinol use as the comparator group (Choe et al., 2010). Follow-up in the two cohort

studies were short of approximately one year and did not adjust for any covariates (Perez-Ruiz et al., 1998, Perez-Ruiz et al., 2010).

Overall effectiveness of allopurinol on renal disease is limited with mixed results on presence of association being reported, suboptimal methods, did not model allopurinol and covariates as time-varying, and not had used UK primary care EHR data.

Health care contacts for gout

Few studies had evaluated frequency of gout flares within EHR data and hospital admissions.

A large cohort study of an administrative data containing health insurance claims found allopurinol users had a higher (adjusted) rate of claims of gout flares than colchicine users over a short follow-up period of less than a year; follow-up was short as it was censored when patients discontinued treatment (Kim et al., 2013b). A small hospital based cohort study used clinical interviews to ascertain the frequency of flares treated with either colchicine or corticosteroids over a 15 month period; that study found longer periods of allopurinol use was not associated with the number of gout flares (Mak et al., 2009). One small case-control study found allopurinol use compared with non-use, and lower doses were associated with recurrent hospital admission during a 12-month period over five years (Hutton et al., 2009).

Evidence is lacking on whether allopurinol use leads to reduced number of primary care consultations for gout-related hospitalisations.

Joint replacement

As described in Section 2.3.7, osteoarthritis has been shown to be associated with gout flares within a joint. People with gout may be at an increased risk of joint replacement as osteoarthritis is the main reason for patients to undergo this procedure. This review had found no studies had evaluated whether allopurinol use was associated with joint replacement.

3.9.6 Effectiveness of allopurinol in subgroups

One of the clinical objectives of this PhD project is to evaluate effectiveness of allopurinol stratified by presence of renal disease and severity of hyperuricaemia. No studies have evaluated effectiveness of allopurinol within these clinically important subgroups.

3.10 Conclusions

To conclude, this chapter presents a narrative review of observational studies of the effectiveness of allopurinol. Common limitations of these studies were that none modelled allopurinol use as time-varying, and often studies did not even control for baseline covariates. The majority of studies were non-UK based and frequently set in rheumatology clinics and hospitals. Only a few studies were based in the UK and used primary care data. The next chapter will describe how a UK EHR database (CPRD) will be used to address some of the limitations identified from this review.

4 Data source and study sample

Preparing electronic primary care medical records for research is a complex task. This chapter describes the data source used in this PhD project as well as data preparation and management procedures. More specifically, the objectives for this chapter are to:

1) Describe the data source, namely the Clinical Practice Research Datalink (CPRD).
2) Define and describe how the study sample was derived.
3) Define the start and end dates of follow-up.
4) Define outcomes, treatment, and covariates.

4.1 Primary care electronic health record databases

In the UK, gout is largely managed in primary care with annual consultation prevalence of 4.7/1,000 in 2007 (Elliot et al., 2009). As identified from Chapter 3, there are a lack of studies using large EHR databases evaluating effectiveness of allopurinol in gout. For this PhD project, data from CPRD was utilised, as CPRD is the largest, well established UK primary care database, has been validated the most, and has been used in over 2,500 publications.

CPRD (GOLD) contains anonymised, routinely collected research quality medical records spanning from 1987 to the present time, jointly supported by the Medicines and Healthcare products Regulatory Agency and National Institute for Health Research. The database holds medical records on over 11.3 million residents registered with over 674 UK primary care practices. Over 95% of the population is registered with a primary care practice (Lawrenson et al., 1999). Active patients who are alive and currently registered, represent 6.9% (4.4 million patients) of the UK general population and are similar in terms of age (although there may be some under-representation in younger age groups), sex, and ethnicity. However, CPRD may not be representative of all practices as CPRD practices tend to be larger than the national

average (Herrett et al., 2015, Campbell et al., 2013). CPRD GOLD collects data from practices that use the management software Vision which records medical record data however, Vision is only used by 9% of English practices compared to more popular systems such as EMIS (56%) and SystmOne (34%). There are geographical differences; Vision practices are predominantly based in London, South of England, greater Manchester, and Birmingham and are under-represented in the North and East of England (Kontopantelis et al., 2018). Furthermore, many Vision practices have migrated to EMIS from 2014 resulting in patient follow-up being censored.

Data available in CPRD include demographic details, lifestyle factors, consultations, prescriptions, referrals, death, and other data available via linked datasets, such as the Index of Multiple Deprivation (IMD) data, Office of National Statistics (ONS) mortality data, and Hospital Episode Statistics (HES) records. Data entries such as diagnoses, symptoms and processes of care are recorded and stored as Read codes and prescribed drugs stored according to their British National Formulary equivalent product code. The Quality and Outcomes Framework (QOF) was introduced in 2004 and aimed to improve quality of care by financially incentivising general practitioners (GPs) to record and monitor patient health, thus patient data had become more complete over time, for example, smoking status (Doran et al., 2011). The accuracy and completeness of the CPRD data has been validated extensively (Herrett et al., 2010). In the UK, the GP is responsible for the majority of patients' medical care, including referral to specialist care, illness prevention and co-ordination of healthcare following hospitalization, and other medical events, making CPRD an ideal data source for assessing the impact of allopurinol on long term outcomes in patients with gout.

Approval for access to anonymised medical records was obtained by the Independent Scientific Advisory Committee (ISAC). The ISAC approval number is 14_163 and the ISAC application form is in Appendix D. Primary care medical records were obtained in January 2015

with linkage obtained in July 2015. Details of the relevant data required for this thesis are given in Table 4.1.

All patients were assigned a unique patient identification number thus linking each patient's medical history across different aspects of care. Dates were provided when patients first registered with the practice, at each point of contact with primary care, and if applicable when the patient left the practice.

Table 4.1: Relevant information obtained from CPRD

Category	Data collected
Patient details	Demographics: Year of birth, sex, primary care practice Registration details: date of first registration, date and reason of transfer out of primary care practice, date of death
Primary care practice details	Region where the primary care practice is located, date of last data collection, last known date data is checked to be of research quality
Clinical details	Recorded medical codes and date of symptoms, signs, and diagnoses and the date they occurred
Referrals	Recorded medical codes and date of patient referrals to secondary care such as hospitals and specialist care centres
Additional information	Various information supporting medical codes on height, weight, body mass index, alcohol consumption and smoking status
Test results	Results of pathology tests ordered. Serum urate levels were of interest only.
Prescriptions	Recorded product codes and date of prescription including schedules and issues for repeat prescriptions, dosage, and instructions on how to take the therapy
Index of multiple deprivation	Categorisation of the Index of Multiple Deprivation 2004 into quintiles, deciles and twentiles where the patient resides
ONS mortality	Date and cause of death
Hospital episode statistics	Recorded ICD-10 codes of reason for inpatient and day case admissions to hospital and date of discharge

ONS: Office of National Statistics

Details of the patient consultation were recorded in various formats. Clinical symptoms, signs, and diagnoses were entered via Read codes. The Read code system is a standard clinical coding system used by GPs where a Read code describes the clinical (or Read) term of the main reason for the patient's consultation. Read codes have a hierarchical structure where the first character indicates the chapter, for example a class of conditions, and subsequent characters are subchapters, grouping similar conditions, until the disease itself has been specified. For

example, the Read code for Gout is C34 where ‘C’ denotes all endocrine, nutritional, metabolic and immunity disorders; within this, ‘3’ denotes other metabolic and immunity disorders; within this, ‘4’ denotes gout. Subchapters of C34 identify different types and causes of gout; for example, C345 identifies gout due to impairment of renal function. For certain Read codes, additional information may be entered into the structured data area in to provide complete information; for example, the Read term 'O/E Blood Pressure Reading', the actual blood pressure reading needs to be recorded. Each distinct Read code was assigned a unique numeric medical code derived by CPRD in order to retrieve necessary details.

Gemscript is a coding system based on the NHS dictionary of medicine and devices used by GPs to manage prescription of therapies. Each distinct Gemscript therapy was assigned a unique numerical product code within CPRD, which were used to retrieve therapy events.

ONS mortality and HES data were coded using the International Classification of Diseases (ICD) system version 10. Similar to the Read code system, the ICD-10 codes have a hierarchical structure. For example, the ICD-10 code for gout is M10; the first character ‘M’ refers to conditions of the musculoskeletal system and connective tissue and the number ‘10’ refers to gout.

Primary care medical records are considered to be accurate in terms of the information they contain. A systematic review of 212 publications validated 183 different diagnoses in CPRD (formally known as the General Practice Research Datalink (GPRD)) and had shown the median 89% of cases were confirmed using additional internal or external information (Herrett et al., 2010). With regards to a diagnosis of gout, a GPRD study has shown that among patients recorded as having gout and were prescribed gout medication (allopurinol, colchicine, probenecid, indomethacin, or other NSAIDS), 86% were confirmed as having gout; in addition if a patient had a high serum urate (SU) level, 100% were confirmed as having gout (Meier and Jick, 1997). Data are also considered complete with 87% of diagnoses recorded on the

database from 58 UK primary care practices (Jick et al., 1991). Disadvantages of using primary care medical records are that patients only present the illness they want to treat to their GP thus mild cases may not present, data quality across GPs and practices may vary although practices are required to have good quality data in order to contribute to CPRD, variable recording of lifestyle factors such as body mass index (BMI), alcohol consumption, diet, and smoking status, and loss of follow-up if patients moved practice.

4.2 Study design and population

A retrospective cohort study design was used. The study sample consisted of patients consulting for gout between the 1st January 1997 and the 31st December 2002. Patients who were under 18 years of age at their first gout consultation within this period were excluded. This would have resulted in minimal exclusions as gout is rare in patients younger than 20 years with 5.11 patients consulting for gout per 100,000 patients in this age group (estimate was derived from CPRD) (Kuo et al., 2015b). This five-year period was chosen as CPRD linkage to HES was available from 1997 onwards and allowed patients to have a sufficiently long follow-up period of a maximum of 18 years. This five-year period was chosen as there was a sufficient number of consulters for gout needed to adequately power analyses to detect treatment effect (see Section 4.5).

For each patient, the date of start of the follow-up (known as the index date), was defined as the first consultation for gout in the period 1st January 1997 to 31st December 2002 and had no prescription for allopurinol or uricosuric drugs in the two years prior to consultation. If the patient was prescribed allopurinol or uricosuric drugs (sulfinpyrazone, probenecid or benzbromarone, definitions can be found in Section 4.4.4) in the two years prior to their gout consultation, a cycle occurred where the subsequent gout consultation, up until the end of

2002, was identified and checked for any prescription (regardless of dosage and duration) for allopurinol and uricosuric drugs in the two years beforehand. This was repeated until a gout consultation was identified as the index date or the patient was excluded if no index date was found. The study sample contained a mixture of prevalent and incident consulters for gout. This was to ensure previous effects of allopurinol and uricosuric prescription did not have an impact on estimating treatment effect. Febuxostat was licensed for use in 2009 hence it was not required to consider this drug when selecting the index date. Figure 4.1 illustrates this process.

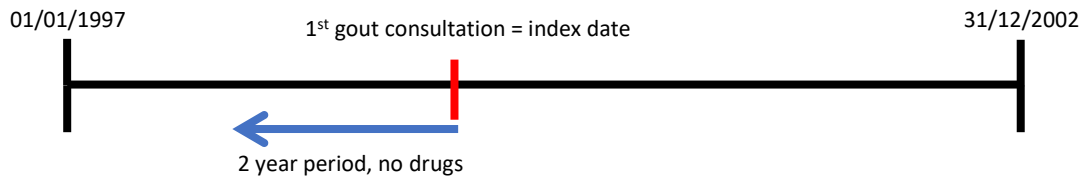
Practices had to have consented to linkage to the HES, ONS mortality, and IMD databases to be included in this study. Patient-level IMD measured in 2004 was available for England-based practices but was not available for Wales, Scotland, and Northern Ireland. Therefore, patients registered to non-England based practices were excluded and findings may not be generalisable to the rest of the UK. Patients had to have been registered with their practice for at least two years prior to their gout consultation and had to have at least one year of follow-up time.

The latest date of follow-up was defined as the minimum of the date of transfer out of the practice, practice last data collection (December 2014), CPRD derived date of death and ONS date of death (see Section 4.4.1 for further details on recording of death).

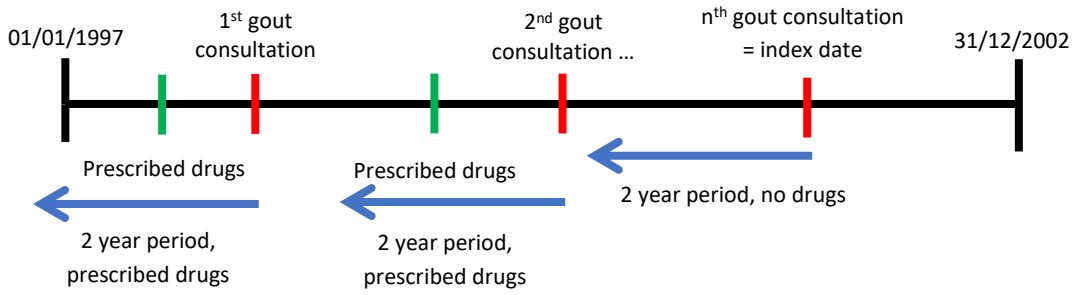
Patients were followed up from their index date and follow-up ended if one of the five scenarios occurred first: (1) the outcome of interest, (2) prescribed sulfinpyrazone, probenecid, benzbromarone or febuxostat; this allowed one to evaluate the effect of allopurinol without interference from the effect of these drugs, (3) transferred out of practice, (4) last data collection (31st December 2014), or (5) death.

Figure 4.1: Selection of the index date

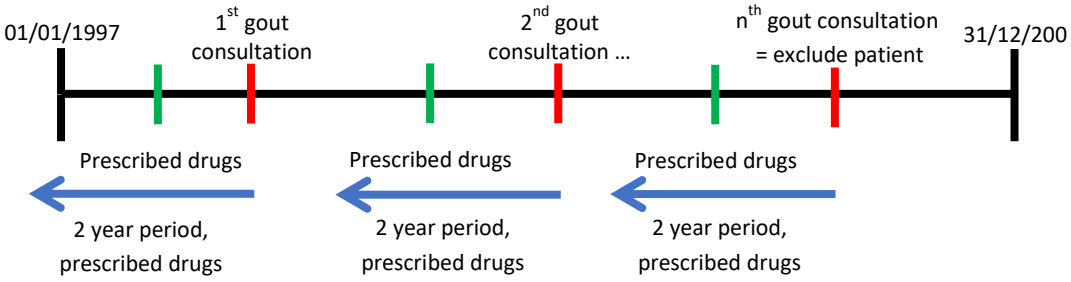
Scenario A: No allopurinol and uricosuric drugs were prescribed in the two years prior to the 1st gout consultation. The 1st gout consultation date was taken as the index date.



Scenario B: Allopurinol and uricosuric drugs was prescribed in the two years prior to the 1st gout consultation. A cycle occurred where the next gout consultation was identified and checked if allopurinol and uricosuric drugs were prescribed in the two years beforehand. The first gout consultation date with no allopurinol and uricosuric drugs prescribed in the last two years was taken as the index date.



Scenario C: Allopurinol and uricosuric drugs were prescribed in the two years prior to all gout consultations between 1997 and 2002. This patient was excluded.



The code lists identifying gout is listed in Section 4.4.4, allopurinol is listed in Section 4.4.2, and uricosuric drugs and febuxostat are listed in Section 4.4.4.

4.3 The landmark method

4.3.1 Time-invariant treatment and covariates

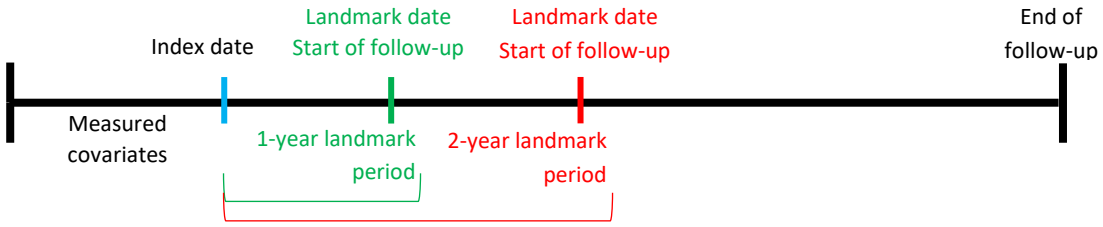
In practice, allopurinol can be prescribed many years after a gout diagnosis (Kuo et al., 2014). The time period prior to prescription of allopurinol needs to be adequately handled in analysis otherwise the treatment effect may suffer from immortal time bias. Immortal time bias is defined as a period of follow-up where outcome cannot be considered to occur. For example, using death as a outcome, for allopurinol users survival time would commence upon prescription for allopurinol thus death cannot be considered to occur prior to date of prescription and patients are considered 'immortal'; in non-users such a requirement that does not apply as survival time starts from study entry. Allopurinol users would accrue survival time waiting for treatment and consequently appear to live longer than non-users biasing treatment effect (Lévesque et al., 2010, Dafni, 2011).

To overcome this issue, one approach is the landmark method that was implemented for this study. A time point after the index date (i.e., the landmark date) was chosen allowing allopurinol treatment to be determined using prescriptions from the index date up until and including, the landmark date (i.e., the landmark period). Follow-up commenced from the landmark date for both allopurinol users and non-users ensuring that outcome was dependent on treatment status at the landmark date. Patients who had the outcome during the landmark period were excluded and non-users who were prescribed allopurinol after the landmark date, a change in treatment status, was ignored in the analysis.

The landmark method is best used in situations where outcome is unlikely to occur early on during follow-up but the likelihood of being treated early on was high. Given the range of outcomes with some likely to occur early on, for example gout consultation, and some likely to occur later, for example cerebrovascular disease, the primary analysis was conducted using

a one-year landmark period. Sensitivity analysis using a two-year landmark period was then performed to evaluate robustness of treatment effects against misclassification of treatment. There is a balance between lessening misclassification of treatment status and excluding too many patients who had the outcome prior to the landmark date leading to a loss of power. Furthermore, Kuo et al. (2014) had shown of the patients who received urate-lowering therapy (ULT), most had received treatment within 5 years. The landmark method is illustrated in Figure 4.2. Patients prescribed three or more months of allopurinol were deemed allopurinol users (treatment definition can be found in Section 4.4.2). To ensure temporal ordering, covariates were measured in the interval prior to the index date (the baseline period).

Figure 4.2: The landmark method at 1 and 2 years



Within each landmark period, if a patient was prescribed ≥3 months of allopurinol, they would be classed as an allopurinol user otherwise classed as a non-user. Follow-up would commence from the landmark date. As the length of the landmark period increases, more patients would be classified as allopurinol users but less patients would be analysed due to increased number of patients having outcome in the landmark period.

Several studies had used the landmark method for treatment effect estimation. For example, Kuo et al. (2015a) evaluated the effect of allopurinol on all-cause mortality in the gout population, and Hsiang et al. (2015) evaluated the effect of statins on death in patients with hepatitis.

4.3.2 Time-varying treatment and covariates

To capture changes in treatment and covariates over time, follow-up time was divided into discrete intervals. As medical record collection is not set-up for research, data will not be measured repeatedly at set time periods that one may be able to control in a prospective cohort study. Data is only collected when patients need to consult their GP, and data of interest will be measured at different times. For example, a patient presenting with gout for the first time may be recorded as gout, but SU level measurements and screening for comorbidities may occur later. It is anticipated there will be missing data over time.

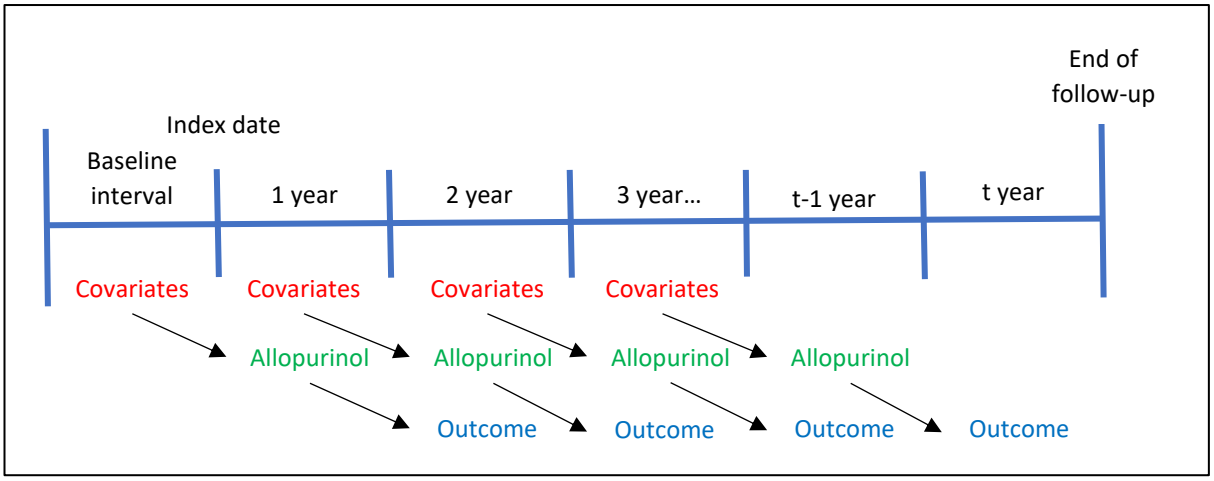
The length of intervals needs to capture how frequently data is recorded but not to have too many intervals such that analyses are not cumbersome and computationally intensive. SU level is the key confounding variable as it is a strong indication for allopurinol use. Guidelines suggest this should be measured every 6 months from treatment and thereafter yearly (Jordan et al., 2007b). Initial interval length considered was 6 months however when proceeding with time-varying PS subclassification analysis (Section 6.3), fitting of multi-level logistic models did not converge or had taken a long time to converge. Given SU level was mostly missing and had to be treated as time-invariant, and the majority of covariates whose status could only change once thus repeated measures were mostly static (see Section 4.4.3), one-year intervals were instead chosen.

Within each interval during follow-up, covariates, treatment, and outcome were repeatedly measured. Patients had a baseline interval (the period prior to the index date) with follow-up represented by at least two intervals. The number of intervals did not succeed the maximum follow-up time.

To establish temporal ordering between covariates, treatment, and outcome, covariates measured in the previous interval were used to predict treatment in the current interval. Then

treatment was used to predict outcome in the subsequent interval. This is illustrated in Figure 4.3; covariates measured in the last two intervals and treatment measured in the last interval of follow-up were not required as outcome was not recorded in subsequent intervals. This is an extension of the landmark method with the interval for covariates and treatment repeatedly created over time.

Figure 4.3: Temporal ordering of covariates, allopurinol treatment and outcome



Longer intervals were considered as a sensitivity analysis, for example two years. However, due to the way temporal ordering was imposed, covariates used to predict treatment may have occurred at the most four years ago; covariates measured in the same interval as treatment were more likely to have stronger associations. Therefore, longer intervals as a sensitivity analysis was not considered.

4.4 Definitions of outcomes, treatment, covariates and other

Definitions for most outcomes, covariates, and treatment were based on Read codes or Gemscript codes. To identify all relevant codes a process was undertaken:

1) Code lists were identified from previous published work using medical record data undertaken in Primary Care Centre Versus Arthritis, Keele University.

2) An online clinical codes repository (clinicalcodes.org (Springate et al., 2014)) was used to identify further codes.
3) Read codes under the highest subchapter from codes identified in (1) and (2) were searched for using the CPRD medical browser; the drug substance names were searched for using CPRD product browser.
4) The lists of consolidated codes were then reviewed and finalised by two GPs and a rheumatologist where applicable.

In some instances, vascular and renal diseases were both covariates and outcome; the same Read codes were used to define both in this instance. Table 4.2 lists the source of studies whose Read code lists were available and were potentially used in this PhD project.

Table 4.2: Source of Read codes and Gemscript codes

Comorbidities	**Source for Read codes**
Anxiety	Burton et al. (2013), Prior et al. (2015), Walters et al. (2012)
Depression	Kontopantelis et al. (2012), Burton et al. (2013), Prior et al. (2015), Rait et al. (2009)
Cerebrovascular disease	PCC (2012), Reilly et al. (2015), Kontopantelis et al. (2014), Khan et al. (2010), Clarson et al. (2015)
Coronary heart disease	Bhattarai et al. (2012), Doran et al. (2011), Hawkins et al. (2013), Horsfield (2004), PCC (2012), Khan et al. (2010), Reeves et al. (2014), Reilly et al. (2015), Parisi et al. (2015), Kontopantelis et al. (2015b), Kontopantelis et al. (2015a), Kontopantelis et al. (2014), Clarson et al. (2015), Roughley et al. (2018)
Type I and II diabetes	Khan et al. (2010), Kontopantelis et al. (2014), Kontopantelis et al. (2015b), Kontopantelis et al. (2015a), PCC (2012), Reilly et al. (2015), Reeves et al. (2014), Clarson et al. (2015), Roughley et al. (2018), Horsfield (2004)
Gout consultation	Clarson et al. (2015), Chandratre et al. (2018)
Hyperlipidaemia	Clarson et al. (2015), Roughley et al. (2018)
Hypertension	Horsfield (2004), PCC (2012), Doran et al. (2011), Kontopantelis et al. (2015b), Kontopantelis et al. (2015a), Reeves et al. (2014), Clarson et al. (2015), Roughley et al. (2018)
Hip or knee joint replacement	Culliford et al. (2015)
Osteoarthritis	Kontopantelis et al. (2015a), Reilly et al. (2015)
Peripheral vascular disease	Doran et al. (2011), Khan et al. (2010), PCC (2012), Clarson et al. (2015), Roughley et al. (2018)
Renal disease	Doran et al. (2011), Reilly et al. (2015), PCC (2012), Kontopantelis et al. (2015b), Khan et al. (2010), Clarson et al. (2015), Roughley et al. (2018)
Lifestyle factors	
Body mass index	(Clarson et al., 2015, Doran et al., 2011, Reeves et al., 2014, Stocks et al., 2015, Fairhurst et al., 2014);
Alcohol consumption	Fairhurst et al. (2014).
Smoking status	(Clarson et al., 2015, Doran et al., 2011, Fairhurst et al., 2014, Kontopantelis et al., 2014, Kontopantelis et al., 2015b, Reeves et al., 2014, Springate et al., 2015, Stocks et al., 2015).
Medications	**Source for Gemscript codes**
Allopurinol	Clarson et al. (2017)
Analgesics	Bedson et al. (2013)
Colchicine	Clarson et al. (2015)
Diuretics	Clarson et al. (2015), Kontopantelis et al. (2015b), Springate et al. (2015), Stocks et al. (2015)
NSAIDS	Bedson et al. (2013), Clarson et al. (2015)

NSAIDS: Non-steroidal anti-inflammatory drugs

4.4.1 Outcomes

The choice of which outcomes to analyse was previously described in Section 3.9.5. The primary outcome was time taken to reach target SU level ≤360µmol/L for the first time. Patients were eligible for this analysis if they had a baseline measurement (as defined in Section 4.4.3.3) that was above target (>360µmol/L), and SU level was measured during follow-up. If patients did not reach target, follow-up was censored at the last date of measured SU level.

The range of time to first secondary outcomes measured during follow-up that were considered is listed in Table 4.3. With the exception of repeated gout consultations, patients were excluded from analysis if they had outcome in the baseline period or if follow-up ended during the landmark period. For repeated gout consultations, consultations that occurred during the landmark period were ignored as it was expected a large proportion of patients would consult during this period.

Table 4.3: Primary and secondary outcomes

Primary outcome
Target SU level ≤360μmol/L
Secondary outcomes
All-cause mortality
Repeated gout consultations
Gout hospitalisation
Hip or knee joint replacement
Cerebrovascular disease
Coronary heart disease
Peripheral vascular disease
Renal disease

SU: Serum urate

The definition of outcomes is described below.

Target SU level

SU level and its units were not consistently recorded as implausible values, duplicate entries, and a range of units used were observed in the data. Based on the raw data, the following process was undertaken to remove implausible data and to standardise it to one set of units, μmol/L, that is commonly used in the UK.

1) Only one SU level per date per patient was allowed:
    i. Duplicates in terms of patient identifier, SU level, unit of measurement, and date of measurement were removed.
    ii. If there were two or more different measurement recorded on the same date, implausible values were removed. If the different measurements appeared to be plausible, one measurement was retained at random.

2) To standardise SU measurements, for each recorded unit of measurement:
    i. The distribution of the SU levels was checked.
    ii. SU levels were then categorised based on the location of the gaps in the distribution.
    iii. If it was believed the wrong units was recorded within a category, a more appropriate unit was chosen based on the distribution of SU levels recorded in other units.
    iv. If there were less than 20 observations within a category, the patient's history of SU level was checked to determine plausibility of SU levels and its units.

The results of standardising SU levels can be found in Appendix E.

All-cause mortality

Date of death was recorded in two ways within CPRD, (1) date the patient transferred out of the practice due to death and (2) CPRD derived date of death based on an algorithm. When the practice is notified that a patient had died, the relevant Read code for death is entered in the patient's medical history alongside with information on date and cause of death. The patient's registration status would be changed to 'transferred out' with death being the reason; the date of transfer is often after the actual date of death. Due to this delay, CPRD developed an algorithm to identify records of death to estimate the date the patient had died. Date of death was defined to be the earliest of (1) the transferred out of practice date due to death, (2) date of death or date of recording information on death, and (3) date when Read code for death was recorded.

A more accurate date of death was obtained from ONS mortality data. It is a legal requirement to register all deaths in England to the General Register Office with a medical certificate completed by the medical practitioner establishing the cause of death. Deaths should be

registered within five days from the date of death however late registration can occur for up to a number of years if the cause of death is not known, for example, deaths referred to coroners. Data is validated to ensure data is entered correctly. Death registration data was collected from 01/01/1998 up until 30/04/2014.

Most deaths recorded in the ONS are also recorded in CPRD however, there may be small differences in the recording of the exact date of death in CPRD; this difference had become smaller over time (Gallagher et al., 2019).

Date of death was defined as the earliest date recorded from either CPRD based on its algorithm or from ONS.

Gout hospitalisation

A hospitalisation due to a gout was defined using HES via ICD-10 codes. The list of codes is shown in Table 4.4. As not all patients would have a record for any hospitalisation, it was assumed they were not hospitalised due to gout.

Table 4.4: ICD-10 codes for gout hospitalisation

ICD-10 code	Description
M10	Gout
M10.0	Idiopathic gout
M10.1	Lead-induced gout
M10.2	Drug-induced gout
M10.3	Gout due to impairment of renal function
M10.4	Other secondary gout
M10.9	Gout, unspecified

Repeated gout consultations

This was defined as any consultation for gout. The Read codes used to define gout in Section 4.2 was also used to define gout consultations; Read codes can be found in Table 4.10 Section 4.4.4.

Vascular diseases

Coronary heart disease, also known as ischaemic heart disease or coronary artery disease, is caused by build-up of fatty deposits in the arteries around the heart and restricting blood flow to the heart. Common symptoms are angina (chest pain) and myocardial infarction (heart attack). Read codes identified patients with angina and myocardial infarction, as well as surgery for treatment such as angioplasty and coronary artery bypass graft.

Cerebrovascular disease refers to disorders of the brain where the blood flow is affected. Read codes identified the different types of cerebrovascular disease including stroke (such as haemorrhage and transient ischaemic attack), aneurysm, embolism, thrombosis, and carotid artery stenosis.

Peripheral vascular (or arterial) disease is caused by fatty deposits in the arteries restricting blood supply to the leg muscles. Read codes identified symptoms include intermittent claudication and ischaemic legs, foot and toes, and surgical treatment including peripheral bypass surgery.

Renal disease

The ability of kidneys to filter and excrete waste (for example creatinine) and excess fluids from the blood becomes impaired over time. Read codes identified those with chronic kidney disease (stages 1-5), or acute or chronic renal failure.

Joint replacement

Read codes identified total hip and knee joint replacement (arthroplasty), a major surgery that replaces the damaged joint with a prosthetic.

4.4.2 Allopurinol

All prescriptions for allopurinol during follow-up were identified along with date of prescription, drug quantity, numeric daily dose, and number of days of prescription. Allopurinol was analysed as a binary variable (allopurinol use vs. non-use). Various definitions of allopurinol use had been used in medical record studies. Kuo et al. (2015a) had used a six-month prescription of allopurinol during the one-year landmark period; Dubreuil et al. (2015) had used any prescription for allopurinol during a six-month period; Rothenbacher et al. (2011) had used any prescription of allopurinol within 30 days of gout diagnosis.

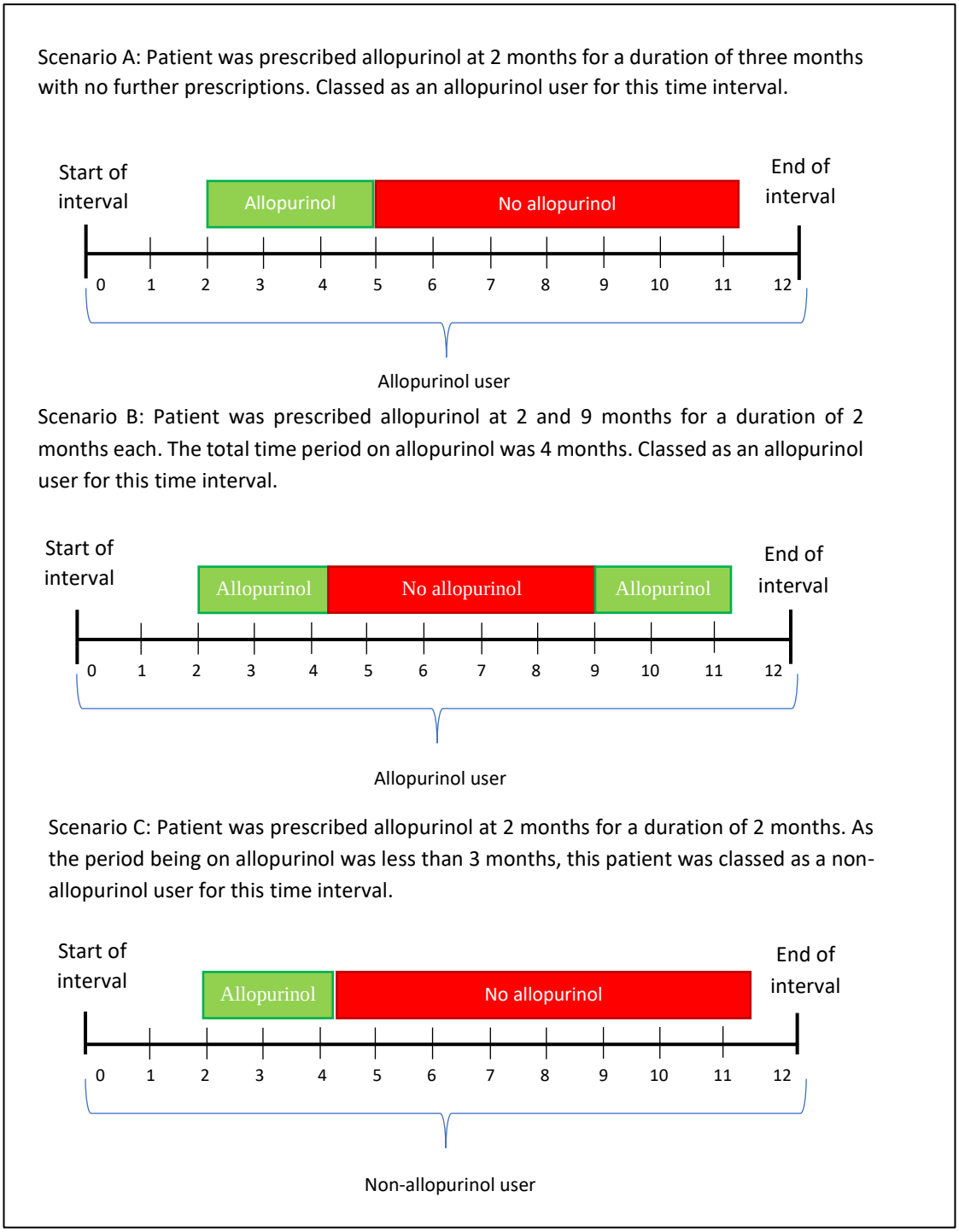
Choice of length of allopurinol prescription is dependent on how long patients adhered with treatment. Kuo et al. (2015b) defined adherence as the proportion of days covered (PDC) with a prescription of ULT (majority of prescriptions is for allopurinol) over a one-year period using data from CPRD. In 2012, 39% adhered with treatment (PDC >80%), 42% partially adhered (PDC 20-79%), and 17% were not adherent (PDC <20%). In 1997, 28% of ULT patients adhered with treatment and approximately 50% were partially adherent.

For this PhD project, it was assumed non-adherent allopurinol users were unlikely to observe treatment effects on long-term outcomes, whereas partially and adherent patients are more likely to observe treatment effects. Given 20% PDC is equivalent to 73 days in a year, and prescription of allopurinol is either given for one, two, or three months, patients with a total of three or more non-consecutive months of allopurinol prescription were deemed allopurinol users, whereas those with no prescription or had less than three months of allopurinol prescription were deemed non-users.

When allopurinol use was time-invariant, prescriptions of allopurinol during the one-year landmark period was used to determine treatment status. For the two-year landmark period, patients were required to have a three-month prescription in the first year or second year of

follow-up to be deemed as an allopurinol user. When allopurinol use was time-varying, prescriptions within a one-year intervals was used to determine treatment status. Repeatedly measuring allopurinol status in this way captured when patients were prescribed and not prescribed allopurinol. Hypothetical scenarios where patients were classed as allopurinol or non-users within a time interval is illustrated in Figure 4.4.

Figure 4.4: Examples of how patients were classed as allopurinol users and non-users within an interval



The Gemscript code and its associated product codes for prescription of allopurinol were obtained from Clarson et al. (2015) study. To identify any further relevant codes, using the product browser supplied by CPRD, search terms of 'allopurinol' and brand names for allopurinol were searched for. The list of product codes used to identify prescription of allopurinol is given in Table 4.5.

Table 4.5: Products codes for prescription of allopurinol

Product code	Gemscript code	Product name
76	59735020	Allopurinol 300mg tablets
368	52350020	Zyloric 100mg tablets (Aspen Pharma Trading Ltd)
413	59734020	Allopurinol 100mg tablets
5182	74888020	Xanthomax 100 tablets (Ashbourne Pharmaceuticals Ltd)
7805	53508020	Zyloric 300mg tablets (Aspen Pharma Trading Ltd)
11975	88998020	Allopurinol 100mg/5ml sugar free oral suspension
13467	48562020	Caplenal 300mg tablets (Teva UK Ltd)
17255	58338020	Hamarin 300 Tablet (Roche Products Ltd)
19037	74889020	Xanthomax 300 tablets (Ashbourne Pharmaceuticals Ltd)
19201	48380020	Allopurinol 100mg tablets (IVAX Pharmaceuticals UK Ltd)
23368	58335020	Hamarin 100 Tablet (Nicholas Laboratories Ltd)
24215	48561020	Caplenal 100mg tablets (Teva UK Ltd)
30768	48366020	Allopurinol 100mg tablets (A A H Pharmaceuticals Ltd)
33484	56807020	Allopurinol 100mg tablets (Actavis UK Ltd)
34005	53683020	Allopurinol 300mg tablets (Teva UK Ltd)
34278	48367020	Allopurinol 300mg tablets (A A H Pharmaceuticals Ltd)
34566	60172020	Allopurinol 300mg tablets (Generics (UK) Ltd)
34573	48371020	Allopurinol 300mg tablets (Wockhardt UK Ltd)
34711	60171020	Allopurinol 100mg tablets (Generics (UK) Ltd)
34930	48370020	Allopurinol 100mg tablets (Wockhardt UK Ltd)
34947	55358020	Allopurinol 100mg Tablet (Lagap)
41520	48381020	Allopurinol 300mg tablets (IVAX Pharmaceuticals UK Ltd)
41541	53684020	Allopurinol 100mg tablets (Teva UK Ltd)
41612	56808020	Allopurinol 300mg tablets (Actavis UK Ltd)
41664	48362020	Allopurinol 100mg Tablet (Celltech Pharma Europe Ltd)
44239	60082020	Cosuric 100mg Tablet (DDSA Pharmaceuticals Ltd)
44240	60083020	Cosuric 300mg Tablet (DDSA Pharmaceuticals Ltd)
45352	48390020	Allopurinol 300mg tablets (Ranbaxy (UK) Ltd)
46941	55359020	Allopurinol 300mg Tablet (Lagap)
52409	18891020	Allopurinol 100mg/5ml oral suspension
54139	18907020	Allopurinol 300mg/5ml oral suspension

To calculate the duration of time the patient was on allopurinol, the start and end date of prescription needed to be determined. The start date of prescription was assumed to be the date when prescription was entered into the Vision system, however the end date of

prescription was not recorded. Although number of days related to each prescription can be recorded, this is not a required field thus was mostly missing (94% of all prescriptions).

Duration was therefore calculated by dividing drug quantity by daily dose which was more complete. For example, if the prescription quantity is 56 tablets and the daily dosage is 2, then duration of prescription would be 28 days. However, it was observed daily dose and quantity may also be missing as some patients may be instructed to take treatment 'as needed' or 'as directed' and values were found to be implausible.

Based on the observed data, rules were derived how to impute missing and implausible data. Implausible daily dose was defined as those with values <0.5 and >6, or equal to 0.75 and 2.25 after seeking guidance from a GP. Implausible drug quantity was defined as a prescription that would last less than a month (for example quantities less than multiples of 28 and 30) with the exception of values 7, 14 and 100 as they appeared to be frequently used.

Imputation of missing and implausible values across all patients

1) If daily dose was missing, the most common daily dose was taken for that particular drug quantity only if it was extremely unlikely another daily dose could have been used.
2) If daily dose was missing, daily dose from the previous prescription was used if drug quantity was the same.
3) If drug quantity was missing, drug quantity from the previous prescription was used if daily dose was the same.
4) If drug quantity and daily dose were both missing, the duration of the previous prescription was taken.
5) If duration was still missing or greater than 30 days (as the majority of durations was for a month), it was assumed duration was for 28 days instead.

The distribution of quantity of drugs and numeric daily dose before and after imputing data is shown in 0. The date the prescription stopped was thus calculated by adding the duration of prescription to the start date of the prescription. Using this information, patients were determined as either allopurinol user or non-user as described above. The distribution of numeric daily dose and quantity of drugs before and after imputing for missing data is shown in Appendix F.

4.4.3 Covariates

Covariates considered for time-invariant and time-varying analyses is listed in Table 4.6 and covered a range of demographics, comorbidities, medication usage and lifestyle factors.

Table 4.6: Covariates

	Time-invariant analysis	***Time-varying analysis***
Demographics		
Age	X	Time-varying: increased yearly
Sex	X	Time-invariant
Deprivation	X	Time-invariant
Comorbidities		
Anxiety	X	Time-varying: status changed once
Depression	X	Time-varying: status changed once
Cerebrovascular disease	X	Time-varying: status changed once
Coronary heart disease	X	Time-varying: status changed once
Type II and I diabetes	X	Time-varying: status changed once
Gout consultation	N/A	Time-varying: status may change multiple times
Hyperlipidaemia	X	Time-varying: status changed once
Hypertension	X	Time-varying: status changed once
Osteoarthritis	X	Time-varying: status changed once
Peripheral vascular disease	X	Time-varying: status changed once
Renal disease	X	Time-varying: status changed once
Lifestyle		
Alcohol consumption	X	Time-varying: status changed once
Body mass index	X	Time-varying: status may change multiple times
Smoking status	X	Time-varying: status changed once
Serum urate	X	Time-invariant
Medication use		
Analgesics	X	Time-varying: status may change multiple times
Colchicine	X	Time-varying: status may change multiple times
Diuretics	X	Time-varying: status may change multiple times
NSAIDS	X	Time-varying: status may change multiple times
Cumulative allopurinol use (years)	N/A	Time-varying: status can only increase in value
Year of follow-up	N/A	Time-varying: increased yearly

NSAIDS: Non-steroidal anti-inflammatory drugs

All time-invariant covariates, with the exception of lifestyle factors (see below), were measured in the two years prior to the index date.

With the exception of sex, deprivation and SU level, covariates were then considered time-varying measured in the two years prior to the index date (with the exception of lifestyle factors, see below) and then in one-year intervals after the index date until follow-up ended.

Selection of covariates were chosen on the basis that they are indications for treatment and risk factors for gout or poor outcomes due to gout. Genetics and diet is not recorded in CPRD hence is considered as unmeasured confounders.

Definition of covariates is described below.

4.4.3.1 Demographics

Only the patient's year of birth was obtained from CPRD. It was therefore assumed patients were born on the 1st January. Time-invariant age was calculated on the index date. Time-varying age was calculated at the start of each interval. Age was treated as a continuous covariate.

The IMD measures relative deprivation for small areas (known as lower layer super output areas) in England at the patient level. Each area is ranked from 1 being the most deprived area to 32,844 being the least deprived area with the ranking based on seven domains: income; employment; health deprivation and disability; education, skills and training; barriers, housing and services; crime; living environment. IMD is updated every few years and CPRD provided deprivation measure based on quintile, deciles and twentile scores based on the rank; deprivation measures were published in 2004 that had used data between 1997 and 2003. Deprivation measured in 2004 was used as this measure was the closest time point to the baseline period. Twentile scores were used for analysis rather than deciles and quintiles as it

retained the most information after categorisation of ranks. Deprivation was treated as a continuous covariate.

4.4.3.2 Comorbidities and medication use

For comorbidities defined by Read codes, absence of the code was assumed to imply absence of the relevant diagnosis. Thus, by default there were no missing data in these measures. Similarly, for drug use covariates, defined via Gemscript codes, absence of the code implied that a patient had not been prescribed the relevant medication. Thus, in this instance too, missing data were non-applicable.

For all time-varying comorbidities, except for gout consultation, once the patient was diagnosed with the comorbidity, they were assumed to have that comorbidity for the rest of follow-up.

In defining analgesics, topical analgesics were excluded as it was believed they have little pain-relieving effects for the gout flare. NSAIDS were also not included as they were considered as a separate covariate.

In defining NSAIDS, topical NSAIDS were excluded as it was believed they have little pain-relieving effects for the gout flare. Unlicensed NSAIDS were also considered s drugs can fall in and out of fashion for treating gout flares.

Time-invariant gout consultation was not considered as adjustment for this covariate may partially adjust out the treatment effect. Patient index date was based on consulting for gout between 1997 and 2002 and was not prescribed allopurinol in the two years prior to consultation. In prevalent gout cases, adjusting for previous gout consultation may introduce the effect of allopurinol as that previous gout consultation was not chosen due to being prescribed allopurinol in the last two years. In treatment effect estimation, part of the effect

would be adjusted for. Therefore, it was assumed patients did not consult for gout in the two years prior to index date.

4.4.3.3 Lifestyle factors

As part of management of gout, GPs give general advice regarding making healthier lifestyle choices (stop smoking and reduce alcohol intake) and to lose weight. This information may be missing as consultations tend to be short (approximately 10 minutes) thus not all relevant data may be entered with the most important aspects of the consultation recorded. As the study sample contains prevalent and incident patients with gout, it is possible in prevalent cases the two-year period prior to the index date may not be long enough to capture these details when they may have been recorded when the patient was first diagnosed with gout. Patients with more severe gout may be more likely to have lifestyle factors recorded than less severe cases. Missingness may arise as healthier patients are less likely to consult their GP if they do not have gout flares.

Body mass index

Measurements of weight, height and BMI were obtained. For time-invariant BMI, initially the most recent BMI was taken in the two years prior to the index date however there was substantial missing data with only 30% of patients with a record. On further investigation, a total of 10,660 patients (63%) had a measurement recorded prior to the index date however some measurements were recorded up to 13 years prior. To minimise missing data, the most recent measurement was taken during this 13-year period (Table 4.7); the median (IQR) number of years prior to the index date in which weight or BMI was recorded was 2.13 (0.65, 4.51) years; 50% of measurements were recorded within five years of the index date.

Table 4.7: Number of eligible patients (n=16,876) with recorded weight and BMI prior to the index date

Period prior to index date	N (%)
2 years	4,981 (30)
5 years	8,356 (50)
10 years	10,519 (62)
13 years	10,669 (63)

Height, weight and BMI underwent a process of cleaning outlined by Bhaskaran et al. (2013) to remove implausible values and to calculate BMI. Further details of this process can be found in Appendix G. BMI was then categorised as normal weight (BMI<25), overweight (BMI 25-30), and obese (BMI 30+).

To further minimise missing data Read codes for BMI were utilised. The Read codes had to differentiate between BMI of normal weight (<25kg/m^2), overweight (25-30kg/m^2), and obese (≥30kg/m^2).

Alcohol consumption

Alcohol consumption was recorded as non-, current- or ex- drinker; additional information on start and stop dates of drinking and units of alcohol consumed per week were recorded as well. Recording of alcohol consumption in CPRD is a blunt measure as it is often recorded as non-, current- or ex-drinker. The number of units consumed is not often recorded (Stewart et al., 2017) and thus it was not possible to accurately distinguish between light, moderate and heavy drinkers.

Similarly to BMI, missing data was substantial in the two years prior to the index date (Table 4.8), therefore the most recent record was taken over the 13 year period prior to the index date. The median (IQR) number of years in which alcohol consumption was recorded prior to the index date was 3.05 (1.17, 5.50).

Table 4.8: Number of eligible patients (n=16,876) with recorded alcohol consumption prior to the index date

Period prior to index date	N (%)
2 years	3,467 (21)
5 years	7,222 (43)
10 years	10,131 (60)
13 years	10,344 (61)

To further minimise missing data, Read codes for alcohol consumption were obtained. The relevant Read Codes contained information on the patient's drinking status if they are a current, ex-, or non-drinker.

When considering alcohol consumption as a time-varying covariate, a record of ex-drinker was only plausible after a record of current drinker however GPs interchangeably use ex-drinkers and non-drinkers. Therefore, alcohol consumption was categorised as never- and ever-drinkers. The never drinkers consisted of non-drinkers provided there was no earlier record of ex- or current drinker. The ever-drinkers consisted of current- and ex- drinkers. This classification can be used over time as patients can switch from never drinker to ever drinker but not from ever drinker to never drinker.

Smoking status

Similarly to alcohol consumption, smoking status was recorded as either non-, current- or ex-smoker with additional information on how many cigarettes smoked per day and the start and stop dates of smoking. Substantial missing data was observed in the two years prior to the index date (Table 4.9), therefore the most recent record was taken over the 13-year period prior to the index date. The median (IQR) number of years in which smoking status was recorded prior to the index date was 2.62 (0.94, 5.19).

Table 4.9: Number of eligible patients (n=16,876) with recorded smoking status prior to the index date

Period prior to index date	N (%)
2 years	4,590 (27)
5 years	8,336 (49)
10 years	11,068 (66)
13 years	11,283 (67)

To further minimise missing data, Read codes for smoking status was obtained. The relevant Read Codes contained information on the patient's smoking status if they are a current smoker, ex-smoker, or non-smoker.

Recording of ex-smokers is under-utilised by GPs (Booth et al., 2013) and ex-smokers and non-smokers may be used interchangeably. Therefore, smoking status was categorised as never smoker and ever smoker. Never smokers consisted of non-smokers who did not have a previous record of ex- or current smoker. Ever-smokers consisted of non- and ex- smokers. This classification can be used over time as patients can switch from never smoker to ever smoker but not from ever smoker to never smoker.

SU level

The cleaning of SU level was already described above in Section 4.4.1.

Preference was for SU to be measured as a time-varying covariate as SU level is a strong indication for treatment and is associated with various outcomes such as vascular and renal diseases. On inspection of data, missing data was substantial in the two years prior to the index date which remained consistent over time. Therefore, SU was only considered time-invariant.

SU recorded more than two years prior to the index date could not be used as patients may have been prescribed allopurinol or uricosuric drugs which may have influenced SU level.

Therefore, within allopurinol users (identified in the one-year landmark period), the most recent SU level was taken prior to prescription of allopurinol up to one year after the index date. In non-users, the most recent SU level was taken up to one year after the index date.

In the primary outcome analysis (time to achieve target SU level ≤360 µmol/L), SU level was considered as a continuous covariate as there was no missing data. In the analysis of secondary outcomes, SU level was dichotomised such that patients with SU ≤360µmol/L were considered to have had reached target SU level, and patients with SU >360µmol/L were considered not to have had reached target SU level.

Cumulative years of allopurinol use

Previous level of exposure to allopurinol was considered as a time-varying covariate. As this covariate was measured in each year of follow-up, it was defined as the cumulative previous number of one-year periods in which the patient was exposed to allopurinol; this covariate will simply be referred to as cumulative allopurinol use.

4.4.4 Other definitions

The list of Read codes for gout consultations is shown in Table 4.10. The consultation for gout can be for any reason.

Table 4.10: Medical codes for gout consultation

Medical code	Read term
709	Gout
2857	Gouty arthritis
3759	H/O: gout
4440	Gouty tophi of other sites
9162	Renal stone - uric acid
9874	Gouty tophi of hand
10080	Gouty arthropathy
11462	Idiopathic gout
12594	Gouty arthritis NOS
14996	Initial gout assessment
16475	Gout monitoring
17284	O/E - auricle of ear -,phi
21687	Gout due, impairment of renal function
24153	Gout NOS
27521	Other specified gouty manifestation NOS
28999	Other specified gouty manifestation
29561	Pre-treatment uric acid level
29658	Joints gout affected
34006	Date gout treatment started
34105	Gout treatment changed
35660	Follow-up gout assessment
35664	Gouty arthritis of the ankle and foot
36481	Gouty tophi of ear
43646	Date gout treatment stopped
43744	Uric acid nepolithiasis
44566	Drug-induced gout
45465	Gouty arthritis of the forearm
49775	Gouty arthritis of the lower leg
50067	Gouty iritis
52101	Gouty arthritis of the hand
52103	Gout drug side effects
52117	Gout monitoring NOS
52969	Gouty nepopathy
57334	Gouty tophi of heart
58064	Gouty arthritis of multiple sites
58746	Gout associated problems
59344	Gouty neuritis
60541	Gouty arthritis of other specified site
61145	Gouty nepopathy NOS
68209	Date of last gout attack
72471	Gouty arthritis of the shoulder region
93677	Gouty arthritis of toe
93689	Gouty tophi of heart
97539	Gouty arthritis of the upper arm

Benzbromarone and febuxostat

As there was no available code list for benzbromarone, Gemscript codes were searched for within the CPRD product browser using the search term 'benzbromarone'. Similarly for febuxostat, search terms of 'febuxostat' and 'Adenuric' identified Gemscript codes.

Probenecid

The product codes for probenecid were obtained from Clarson et al. (2015) and in addition, the terms 'probenecid' and 'Benemid' were searched for within the product browser under product name and drug substance name.

Sulfinpyrazone

The product codes for sulfinpyrazone were obtained from Clarson et al. (2015) and in addition, the terms 'sulfinpyrazone' and 'Anturan' were searched for within the product browser under product name and drug substance name.

4.5 Power calculation

A feasibility count was undertaken to ensure there was sufficient number of patients consulting for gout between 1997 and 2002, and there was sufficient sample size and power to detect reasonable treatment effect.

It is assumed that the total CPRD annual registered population is 5.5 million patients and that approximately 50% of practices contributing to CPRD have consented to linkage to HES, deprivation and ONS mortality. This yields relevant annual population of approximately 2.75 million. It is estimated that 70,000 of these will have gout, based on the latest figure for prevalence of gout as 2.5% (Kuo et al., 2015b).

Between 1st January 1997 and 31st December 2002, 33,538 patients aged 18 years and over had a Read code for gout and were registered with their practice for at least two years prior. To determine the number of patients prescribed allopurinol, the landmark method described in Section 4.3.1 was used. The number of patients prescribed allopurinol is listed in Table 4.11. Approximately half these patients would have linkage.

Table 4.11: Number of patients prescribed allopurinol

Landmark period	Prescribed allopurinol N=33,538 N (%)	Prescribed allopurinol in patients with linkage N=16,769 N (%)
1 year	10,266 (30.61)	5,133 (30.61)
2 years	11,648 (34.74)	5,824 (34.73)

The primary outcome was SU level ≤360µmol/L. Based on a single small study, 77% of those on allopurinol are expected to reach this threshold as opposed to 25% of those not taking allopurinol (Roddy et al., 2007b). Secondary outcome of recurrent gout consultations had not been compared between treatment groups however, a nationwide population based study found 22% of newly diagnosed patients with gout prescribed with allopurinol had a recurrent gout attack within a year compared to 14% of non-users, which patients would consult for the GP for (Trifiro et al., 2013). Sample size and power calculations could not be performed taking into account time to outcome due to lack of studies. Therefore, calculations were based on the proportions of outcome between treatment groups.

Assuming 90% power with 5% significance and allocation ratio of two (as approximately 33% of patients were prescribed allopurinol (Table 4.11)), Table 4.12 shows the total number of patients required to detect a difference in proportion (percentage) in outcome (target SU level and recurrent gout flare) between treatment groups. The proportion of patients having outcome in the treatment groups was based on the above figures, which then varied if a smaller or larger proportion of outcome was to be observed. A maximum of 1,128 patients

were needed assuming proportion of 45% and 36% reached target SU level in allopurinol users and non-users respectively; a maximum of 3,753 patients were needed assuming proportion of 30% and 35% of recurrent flare in allopurinol users and non-users respectively.

Table 4.12: Total number of patients required to detect a difference in proportions in outcome between treatment groups

Outcome				
Target SU level		Allopurinol non-users		
Allopurinol users		15%	25%	35%
	45%	105	261	1,128
	55%	63	120	288
	65%	42	69	126
	75%	27	42	23
Recurrent gout flare		Allopurinol non-users		
Allopurinol users		5%	15%	25%
	10%	1,278	2,088	306
	20%	216	2,703	3,312
	30%	99	357	3,753
	40%	60	144	453

Across the 1- and 2-year landmark periods, the power to detect a difference of proportion in outcome (using the same proportions in Table 4.12) between allopurinol users and non-users was almost 100%.

The five-year period was sufficient to yield sufficient number of consulters for gout and was adequate to power analyses potentially across a range of outcomes.

4.6 Summary

This chapter describes how the retrospective cohort study was set up using primary care medical records from CPRD. The study sample was defined, how time-invariant and time-varying covariates and treatment would be measured over time, definitions of outcomes, allopurinol, and covariates, and sample size and power calculations were provided.

The next chapter aims to describe the statistical models used that will estimate treatment effects.

5 Propensity scores and marginal structural models

Observational studies are often used to estimate the causal effect of treatment on outcome. In such studies, controlling for confounding can be challenging and is one of the most important aspects in estimating valid treatment effect.

This chapter introduces the conceptual framework and the assumptions required to estimate casual effect. An overview of possible approaches to control for time-invariant and time-varying confounding is discussed, and methods applied to estimate effect of allopurinol on outcomes in patients with gout in this PhD project described.

5.1 Causal inference in observational studies

Neyman (1923) developed a conceptual framework to investigate causality in the context of randomised controlled trials (RCTs) which was later extended to observational studies by Rubin (1974).

Consider a binary treatment A (1 if treated and 0 if untreated) and binary outcome Y (1 if outcome occurred and 0 if outcome did not occur). Let Y^a denote two potential (or counterfactual) outcomes, $Y^{a=1}$ if treated and $Y^{a=0}$ if untreated. The idea is to compare the same patient under two different treatments at baseline and effect of treatment on outcome measured during follow-up; the difference in outcome is attributable to treatment (Equation 5.1).

$$Y^{a=1} - Y^{a=0} \qquad 5.1$$

In reality, both outcomes cannot be observed as the patient is either treated or untreated therefore, patient specific treatment effect cannot be calculated. Instead, the average causal

effect is estimated at the population level. Let $E[\mathrm{Y}^{\mathrm{a}=1}]$ and $E[\mathrm{Y}^{\mathrm{a}=0}]$ be the expected value of outcome in treated and untreated patients respectively. The average treatment effect (ATE) is then defined in Equation 5.2. For a binary outcome, $E[\mathrm{Y}^{a}]$ represents the probability of outcome $pr[Y^{a} = 1]$ therefore, Equation 5.2 represents the risk difference; alternatively, the risk ratio or odds ratio could be derived. For continuous outcome, Equation 5.2 represents the average difference in outcome.

$$E[Y^{a=1}] - E[Y^{a=0}] \qquad 5.2$$

So far above, treatment at a single time point was considered. In a repeated measures setting treatment is measured repeatedly over time. Consider time-varying treatment A_t measured at time t; t can be measured in continuous or discrete time (for example every year). $A_t = 1$ if patient was treated at time t and $A_t = 0$ if patient was untreated at time t. Let treatment history be denoted as $\bar{A}_t = (A_{0,} A_1, \ldots., A_t)$. The number of treatment histories can become large as there can be as many as 2^tcombinations. Each treatment history has an associated outcome measured at the end of follow-up thus there will be 2^t outcomes for each patient of which only one will be observed. For simplicity, two treatment histories are often compared in practice. For example, never treated history $\bar{A} = (0, 0, \ldots, 0)$ is often compared with always treated history $\bar{A} = (1, 1, \ldots, 1)$. The ATE is defined as the difference in expected outcome value between any two treatment histories (Equation 5.3) (Robins et al., 2000).

$$E[Y^{\bar{a}}] - E[Y^{\overline{a'}}] \qquad 5.3$$

On the other hand, outcome can be measured repeatedly at each time point. The difference in expected value between treatment groups is measured at time t (Equation 5.4) (Robins et al., 1999).

$$E[Y^{a_t=1}] - E[Y^{a_t=0}] \quad 5.4$$

However, the above definitions (Equations 5.2, 5.3 and 5.4) relies on potential outcomes, i.e., only one outcome will be observed by the patient. Causal inference is underpinned by finding the best substitute patients who in all ways are similar to the observed patients apart from treatment, in an attempt to observe potential outcomes. Randomisation is the gold standard approach as each patient is randomly assigned to treatment hence it is due to chance whether potential outcomes are observed or not. A key consequence of randomisation is that the treatment groups are exchangeable, meaning it does not matter which particular group received treatment of interest, the results will be the same regardless.

Another definition of treatment effect is the average treatment effect for the treated (ATET) which restricts attention to the subgroup of the population who received treatment. The causal effect in this subpopulation is the average difference in outcome between treated and untreated patients (Equation 5.5).

$$E[Y^{a_t=1} - Y^{a_t=0}] \mid A = 1 \quad 5.5$$

In an RCT, the ATE and ATET estimates coincide as the treated population will on average be the same as the overall population due to randomisation. Whereas in an observational study in the absence of randomisation, the treated and overall populations on average are likely to differ thus the two estimates are unlikely to coincide (Austin, 2011a). In this PhD project the ATE is of interest.

5.2 Fundamental assumptions

In observational studies, assignment of treatment is influenced by patient and other (for example, hospital level) covariates. Causal inference is determined by viewing the observational study as a 'conditionally randomised study' provided the following assumptions are met (Hernán and Robins, 2020, Cole and Hernan, 2008):

Consistency: Treatment is unique and well defined thus has the same effect on the patient regardless of how the patient received treatment. Consistency is not met when treatment is not well defined, such as having multiple versions of treatment with differing effects on outcome e.g., various treatment doses unaccounted for in analysis.

Exchangeability: Treated and untreated patients are exchangeable if the treated patients were instead untreated, they would on average have the same outcome if they were treated, and vice versa, conditional on a set of observed covariates.

Positivity: Patients' probability of being assigned to each treatment group, conditional on observed covariates, is greater than zero.

Correct model specification: The regression model used to obtain the associations between covariates, treatment, and outcome needs to be specified correctly, for example, using the correct functional form for continuous covariates.

The estimated ATE (in Equations 5.2, 5.3 and 5.4) is either considered to be conditional or marginal. A conditional effect is the average effect of treatment at the patient level. A marginal effect is the average treatment effect at the population level. When observed and unobserved confounding effects are absent, the conditional and marginal effects are the same.

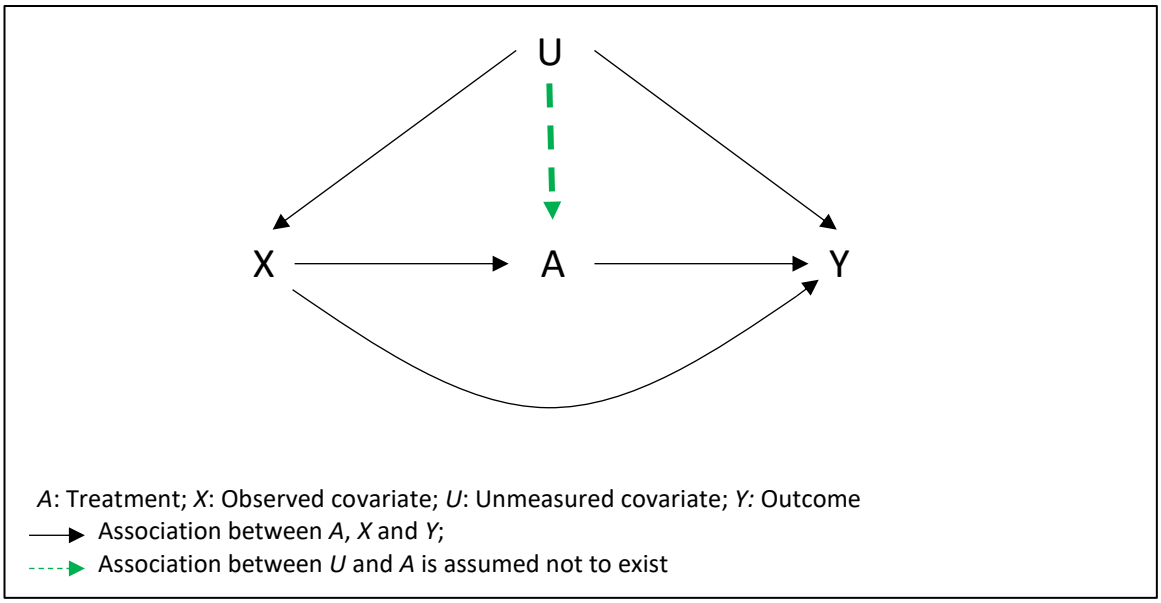
The challenge in estimating causal treatment effect is that treatment assignment is not random leading to poor exchangeability between treated and untreated patients. The

treatment groups may be dissimilar and some patients may be guaranteed to receive treatment, violating the exchangeability and positivity assumptions respectively.

The lack of exchangeability is otherwise known as confounding that biases treatment effect. Within a healthcare setting, confounding may arise from clinical decisions, known as confounding by indication. This is where assignment of treatment is based on covariates such as patient health status and prognosis, the clinician's past experience with the treatment, or the willingness of the patient to take the medication as prescribed. As a result, patients prescribed treatment will often differ (for example be more severely ill or have more comorbidities) from patients not prescribed treatment; if these covariates also predict outcome, confounding by indication is present.

The structure of confounding in the simplest case is represented in Figure 5.1 where a single covariate and treatment are measured at one time point. Figure 5.1 shows the casual effect of A on Y however, this association is confounded by observed covariate X; this induces additional association between A and Y from the confounding effect of X on A and Y (represented by the black arrows). Although covariates can be controlled for in treatment effect estimation, unmeasured covariates cannot be controlled for. In practice, assumption of no unmeasured confounding is often made, i.e., no association between U and A illustrated by the dashed green arrow in Figure 5.1.

Figure 5.1: Illustration of confounding: treatment and covariate measured at a single time point

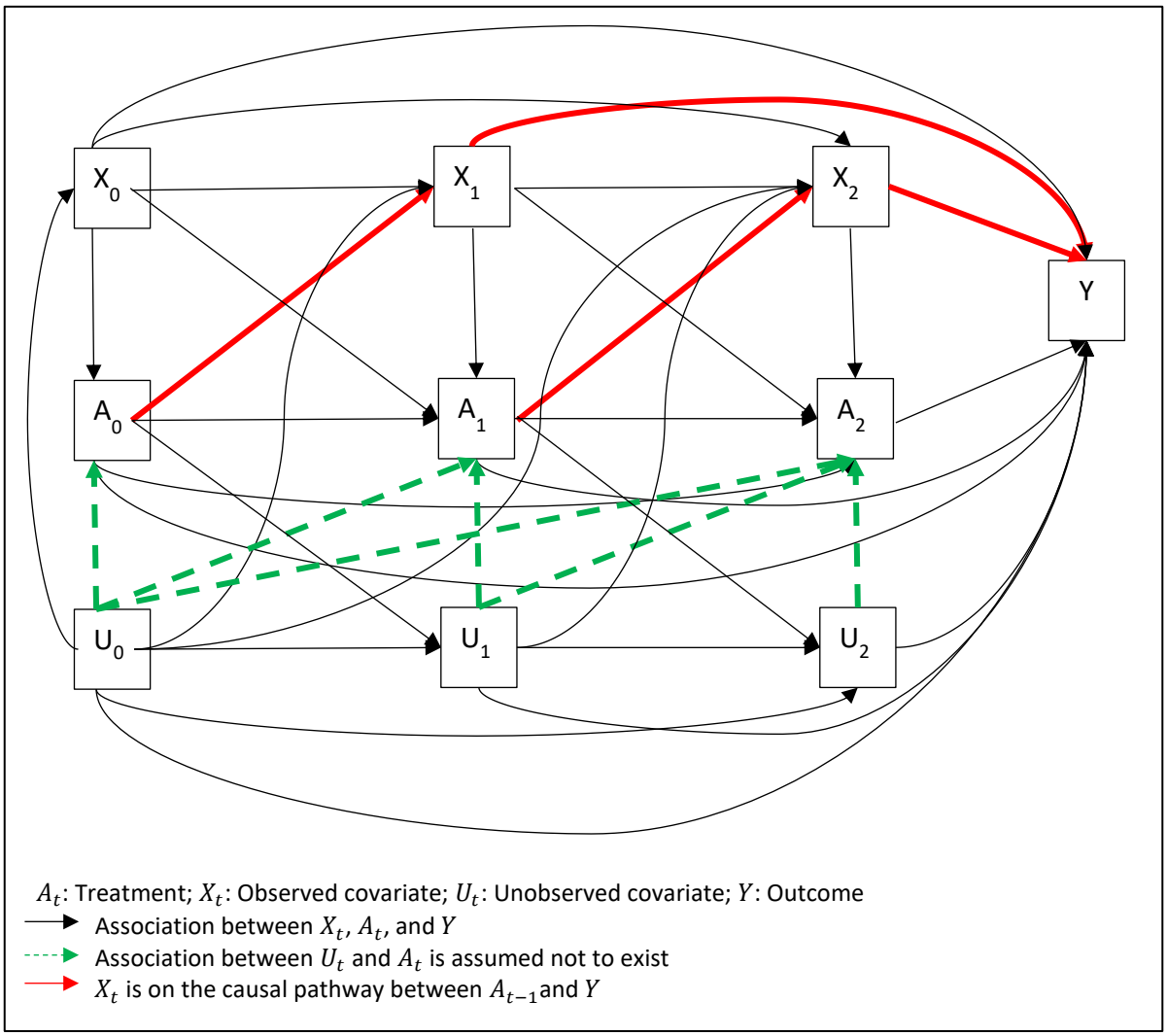


In a repeated measures set-up, confounding is observed at all time points and the strength of confounding effect may vary over time, which may be due to, for example, patient attrition, missing data, and noncompliance with or changing treatment. A covariate X_t measured at time t, is a time-varying confounder if its value changes over time and predicts future treatment and outcome. If past treatment also predicts X_t, time-varying confounding is affected by prior treatment.

Time-varying associations between covariates and treatment are essentially as depicted in Figure 5.1 but repeated over time. Figure 5.2 illustrates an example using three time points $(t = 0, 1, 2)$, but by analogy this can be extended to include further time points. As in Figure 5.1, the assumption of no unmeasured confounding is illustrated by dashed green arrows. In addition, treatment A_t and a single covariate X_t at time t, predict subsequent treatment and covariate (represented by the black arrows). Covariate X_1 is a time-varying confounder as it predicts treatment A_1 and outcome Y, but is associated with past treatment A_0 thus mediates the association between A_0 and Y. Similarly, X_2 is a time-varying confounder as is predicts A_2 and Y, but is also associated with A_1, thus mediates the association between A_1 and Y.

Adjustment for covariates that have a dual role of confounders and mediators in regression models will bias the estimated treatment effect as the effect of previous treatment on outcome will be adjusted out (Robins et al., 2000).

Figure 5.2: Illustration of confounding: treatment and covariates measured at multiple time points



5.3 Common methods to control for confounding

Choice of methodology adopted to control for confounding will depend on different aspects of the study, such as the available sample size and number of observed covariates, complexity of the data, whether control of time-invariant or time-varying confounding is of interest, analyst's preferences and expertise, and available software.

This section outlines some of the most common approaches that are used in practice to control for confounding, and furthermore, rationale is provided why propensity score (PS) subclassification and marginal structural models (MSM) were chosen to estimate treatment effectiveness of allopurinol using the Clinical Practice Research Datalink (CPRD).

5.3.1 Subclassification and matching

Subclassifying involves placing patients into homogenous mutually exclusive subclasses based on the values of observed covariates of interest. Difference in outcome between the treated and untreated patients is calculated in each subclass, then averaged across all subclasses. However, as the number of covariates considered increases the sample size within each subclass would decrease rapidly or become zero, depending on the sample size and distribution of covariates (Anderson et al., 1980).

In matching, a treated patient is matched to n untreated patients based on them having the same values of observed covariates of interest thus ensuring comparability of treatment groups. There are several decisions that need to be made in order to perform matching. These include choice of covariates to match on; whether to match on the exact same covariate value (exact matching) or to match on similar covariate values within a predefined caliper width (interval matching); and specifying the number of untreated patients to be matched to a treated patient. Treatment effect is subsequently estimated in the matched sample with patients who are not matched are excluded from the analysis. An obvious disadvantage is that there is an increasing difficulty in finding matches for each of the treated patients as the number of covariates included in the matching process increases (Anderson et al., 1980, Greenland and Morgenstern, 1990).

5.3.2 Generalised linear models

Generalised linear model (GLM) is the most common method used to quantify the effect of treatment on outcome adjusting for covariates. GLM refers to a larger class of models including linear, logistic, and Poisson regression among others. Let $\boldsymbol{X} = X_0, X_1, X_2, \dots, X_K$ be a vector of covariates with $X_0 = 1$ and $\boldsymbol{\beta} = \beta_0, \beta_1, \beta_2, \dots, \beta_K$ is a vector of unknown regression parameters to be estimated. For treatment A, the regression parameter to be estimated is denoted α. The GLM consists of two components: (1) the linear predictor (or function) of treatment and covariates (Equation 5.6), and (2) the link function $g[.]$ that transforms the expectation of outcome to the linear predictor (Equation 5.7).

$$E(Y|A, \boldsymbol{X}) = \alpha A + \sum_{k=0}^{K} \boldsymbol{\beta}_k \boldsymbol{X}_k \qquad 5.6$$

$$g[E(Y|A, \boldsymbol{X})] = \alpha A + \sum_{k=0}^{p} \boldsymbol{\beta}_k \boldsymbol{X}_k \qquad 5.7$$

The specification of $g[.]$ will depend on the distribution of the outcome; for example, for a binary outcome, the link function is a logit function $g[E(Y|A, \boldsymbol{X})] = log\left[\frac{E(Y|A,\boldsymbol{X})}{1-E(Y|A,\boldsymbol{X})}\right]$ and for a Poisson count distributed outcome it is $g[E(Y|A, \boldsymbol{X})] = log[E(Y|A, \boldsymbol{X})]$.

If the GLM is specified correctly such that all relevant covariates are included in $\boldsymbol{X}$, and the assumptions associated with the outcome distribution are met, an unbiased conditional estimate of the ATE can be obtained (Austin et al., 2007b).

GLM however may produce unreliable treatment effect estimates in a range of situations. For example, issues are encountered when the model is over-fitted, i.e., when the number of included covariates is large compared to the number of observations and rate of outcome; this can lead to certain combinations of covariates not being observed in the study sample

and poor comparability between treatment groups. Consequently, the estimated regression parameters are extrapolated. The fact that outcome is always modelled when determining which covariates to adjust for increases the risk of model over-fitting. Furthermore, model assumptions often go untested in practice, particularly as the number of covariates increases.

5.3.3 Propensity scores

Rosenbaum and Rubin (1983) introduced the notion of PS. Their seminal paper described the theory and application of PS to estimate treatment effect in observational studies accounting for confounding effects from observed covariates. Since then, PS methodology has become a popular choice for treatment effect estimation, and extensions of the standard PS methodology have been examined via both simulation studies and real data, including for example comparisons with traditional multivariable regression models (Shah et al., 2005), assessment of different PS adjustment methods, choice of covariates for inclusion in PS (Austin et al., 2007a), extensions to treatments of non-binary form (Imai and van Dyk, 2004), development within Bayesian framework (McCandless et al., 2009), as well as estimation of time-varying PS when treatment and confounding variables are time-varying (e.g., risk set matching (Li et al., 2001, Lu, 2005)) and although developed separately MSM (Robins et al., 2000)).

PS methodology makes two core assumptions, (1) the ignorable treatment assignment assumption, and (2) the stable unit treatment value assumption (SUTVA). These two assumptions are a mixture of the exchangeability, consistency and positivity assumptions described in Section 5.2 above.

The ignorable treatment assumption implies that treatment assignment A and potential outcomes $Y^{a=1}$ and $Y^{a=0}$ are conditionally independent given covariates X (*exchangeability*) and that each patient has a positive probability of receiving each treatment (*positivity*).

SUTVA is made up of two assumptions. Firstly, it is assumed that the potential outcome of a patient who received treatment would remain the same regardless of how treatment was assigned to that patient and regardless of treatment assignment to other patients, essentially ruling out interference between patients and effect of treatment assignment (i.e., association between covariates and treatment) on outcome. Secondly, it is assumed that treatment is well defined such that there is only one version of treatment; there cannot be various versions of treatment, such as different doses of treatment that are unaccounted for in statistical analyses (Rubin, 1980, Rubin, 1986). The SUTVA assumption includes the consistency assumption above.

PS is a balancing score $e(\boldsymbol{X})$, defined as the probability that lies between 0 and 1 of receiving treatment conditional on observed covariates $\boldsymbol{X}$ (Equation 5.8).

$$e(\boldsymbol{X}) = pr(A = 1 \mid \boldsymbol{X}), 0 < e(\boldsymbol{X}) < 1 \qquad 5.8$$

In RCTs the PS is known as it is based on the allocation ratio. Due to randomisation the PS is conditional on covariates $\boldsymbol{V}$ which consists of both observed ($\boldsymbol{X}$) and unobserved ($\boldsymbol{U}$) covariates and all patients have a non-zero probability of receiving each treatment. Also, treatment assignment A is independent of outcome Y given covariates $\boldsymbol{V}$ (Equation 5.9).

$$e(\boldsymbol{V}) = pr(A = 1 \mid \boldsymbol{V}), 0 < e(\boldsymbol{V}) < 1$$
$$Y \perp\!\!\!\perp \mathrm{A} \mid \boldsymbol{V} \qquad 5.9$$

These properties do not hold in observational studies hence, at best, $e(\boldsymbol{X})$ can be estimated, often using logistic regression in the case of a binary treatment. Treatment assignment is assumed to be strongly ignorable if $\boldsymbol{V} = \boldsymbol{X}$, i.e., all relevant covariates are included in $\boldsymbol{V}$ and there is no unmeasured confounding. It follows that if treatment assignment is strongly ignorable given $\boldsymbol{X}$ then it is strongly ignorable given the balancing score (i.e., the PS) as it is a function of $\boldsymbol{X}$. On this basis Y is independent of treatment assignment given the PS (Equation 5.10).

$$Y \perp\!\!\!\perp \mathrm{A} \mid e(\boldsymbol{X})$$
$$0 < pr[A = 1 \mid e(\boldsymbol{X})] < 1 \qquad 5.10$$

At a specific PS value, the distribution of PS would be the same amongst treated and untreated patients although specific covariate values will differ. Here, the PS balances covariates between treated and untreated patients. Consequently, direct adjustment for PS in a regression model, subclassification and matching on PS, and inverse probability treatment weighting (IPTW) would produce unbiased estimate of the treatment effect.

The most straightforward method of using estimated PS is to include it as a covariate in the regression model estimating the association between treatment and outcome (with conditional treatment effect estimated (Austin, 2013, Austin et al., 2007b). Such direct model adjustment approach is the only method to model the association between PS and outcome and requires that association to be correctly modelled.

Subclassification on the correctly specified and estimated PS divides treated and untreated patients into J subclasses based on percentiles of the PS distribution; thus each subclass would have equal number of patients. The treatment effect is calculated within each subclass generating subclass specific estimates of the treatment effect which can then be pooled

generating an overall conditional ATE estimate weighted by the inverse of the treatment effect's standard error.

PS matching involves matching each treated patient to n untreated patients based on having the same or similar PS. Consequently, within the matched sample the distribution of covariates is similar between the treatment groups. Subsequently comparison of outcomes between the treatment groups would yield a marginal ATE estimate that is less biased and more precise (i.e., smaller standard error) compared to estimate based on an unmatched sample. There are several decisions that need to be made to perform matching and there are ample simulation studies providing guidance. These include: choosing the most appropriate/suitable matching algorithm, with popular choices including the nearest neighbour matching and optimal matching (Austin, 2014); specifying the number of untreated patients to be matched to a treated patient (Austin, 2010); specifying caliper width (or the range) within which the PS must fall in order to be considered a valid match (Austin, 2009b, Austin, 2011b); and choosing whether matching is performed with or without replacement (Austin, 2014).

Rosenbaum (1987) introduced IPTW as a form of model-based direct standardisation. IPTW belongs to a larger class of MSM used to estimate treatment effect in the presence of time-varying treatment and time-varying covariates that are affected by prior treatment (Robins et al., 2000). The weight is defined as the inverse of the probability of receiving treatment that the patient was given (or observed). For a treated patient, their weight is simply the inverse of their PS; for an untreated patient, their weight is the inverse of the PS subtracted from one (Equation 5.11).

$$w = \frac{A}{e(\boldsymbol{X})} + \frac{1 - A}{1 - e(\boldsymbol{X})} \qquad 5.11$$

The weight value determines the number of times a patient is represented in the pseudo-population. Intuitively, untreated patients with high probability of treatment, and treated patients with low probability of treatment will be assigned larger weights. Using weights in analysis reweights the treated and untreated groups up to the population level making them representative of the population, where there is no association between covariates and treatment, thus allowing marginal ATE to be estimated.

Different PS methods have various advantages and disadvantages. Direct adjustment for PS in the outcome model is the simplest approach however it is generally not considered in practice because it does not allow one to assess comparability nor create comparable treatment groups. Matching is a popular approach however, as mentioned above, several decisions need to be made during the modelling process and treatment effect estimates may be dependent on these different choices. Furthermore, generalisability of findings may be affected if many patients are discarded due to lack of good matches. If poor matches are found in matching, alternative approaches are IPTW and subclassification that analyses all patients. With subclassification on PS, one has to ensure that each subclass has high enough sample size in both treatment groups, while considering that bias in treatment effect estimate is reduced the greater the number of subclasses used. One methodological consideration of IPTW is the estimation of extreme weights from very small PS; such weights may bias and increase the variability of the estimated treatment effect.

These different PS approaches estimate either conditional or marginal treatment effects. Direct PS adjustment and PS subclassification allows one to estimate patient-level treatment effect conditional on PS (subclasses); whereas PS matching and IPTW allows one to estimate a marginal treatment effect at the population level.

PS methodology is an appealing approach as it is inherently a more natural method of mimicking RCT design compared to traditional regression models. PS approaches avoid the

use of covariates themselves to achieve comparable treatment groups, thus minimising the effect of model selection and over-fitting sometimes seen in regression models.

5.3.4 Controlling for time-varying confounding

Conventional statistical methods such as time-varying Cox regression (Fisher and Lin, 1999, Cox, 1972) and generalised estimating equations (GEE) regression (Zeger et al., 1988) are used to model longitudinal data where treatment and covariates (and also outcome) may be observed repeatedly over time. The time-varying Cox model compares the risk of outcome between treatment groups and re-evaluates that risk with each change to treatment status. GEE regression, an extension of GLM, accounts for the correlation between repeated measurements within a patient and estimates the population average effect of treatment. To naively adjust for time-varying confounding variables and past treatment will estimate biased treatment effect (as previously shown in Figure 5.2) (Robins et al., 2000, Cole et al., 2005).

MSM, parametric G-computation formula and G-estimation, collectively known as G-methods (with 'G' standing for `generalised'), were specifically developed to estimate time-varying treatment effect in the presence of time-varying confounding.

More direct approaches of utilising PS estimated over time in matching and subclassification have also been proposed.

Some of these different methods are briefly described below, and further outline of the associated advantages and disadvantages is provided in Section 5.3.5.

Marginal Structural models

Robins et al. (2000) developed a class of casual models known as MSM that directly models counterfactual outcomes. MSM are termed "marginal" because they use the joint distribution of treatment and covariates via IPTW to estimate a weighted treatment effect on outcome

across all time points i.e., treatment effect is unconditional on covariates. They are termed "structural", as they were originally developed within the context of economic and social sciences, where historically models estimating causal associations are referred to as structural.

Fitting MSM is a two-step process. The first step involves correctly specifying the PS model such that the patient's probability of receiving observed treatment at time t, given their treatment and covariate history up to that point, is estimated and subsequently used to estimate weights. Various weights can be estimated. The simplest are unstabilised weights defined in Equation 5.12, where $\bar{A}_t$ represents treatment history up to and including time $t-1$; $\bar{\boldsymbol{X}}_t$ represents covariate history (of time-varying and time-invariant covariates) up and including time $t-1$.

$$W = \prod_{t=0}^{T} \frac{1}{pr(A_t|\bar{A}_t, \bar{\boldsymbol{X}}_t)} \qquad 5.12$$

The second step uses the estimated weights to reweight the study sample to create a pseudo-population such that the association between covariates and treatment no longer exists thus removing effects of confounding from treatment effect estimation.

G-computation formula

The G-computation formula, introduced by (Robins, 1986), compares outcomes under different treatment histories as if they were derived from a randomised study, i.e., outcome that would have been observed if all patients in the study sample followed a particular treatment history. G-computation calculates the expectation of outcome for a particular treatment history by the sum over all covariates, the probability of outcome conditional on

treatment and covariate histories, and the probability of covariate conditional on treatment and covariate histories (Equation 5.13).

$$E[Y(\bar{A})] = \sum_{X}[E(Y|\bar{A},\bar{X}) \prod_{t=0}^{T} pr(X_t|\bar{A}_{t-1},\bar{X}_{t-1})] \quad 5.13$$

G-computation is performed in two stages. Firstly, the associations between outcome, treatment and covariates are modelled using GLM thus estimating the joint distribution of outcome and covariates. Secondly, the estimated joint distributions are then used to simulate the risk of outcome at each time point under hypothetical treatments to be compared (Daniel et al., 2011, Daniel et al., 2013).

G-estimation of structural nested models

Structural nested models (SNM) estimate the effect of treatment conditional on different values of a time-varying effect modifier. An effect modifier is when a covariate increases or decreases the effect of treatment. Typically, effect modifiers are included in GLM by fitting an interaction term between treatment and covariate. SNM estimates the average effects of treatment at each time point as a function of effect moderators prior to that time point (Robins et al., 1992).

There are different types of SNM based on the distribution of outcome; for continuous outcome there is the structural nested mean models and for time-to-event outcomes the structural nested failure time model.

Risk set matching

Li et al. (2001) proposed balanced risk set matching that minimises imbalance between treatment groups on specified covariates. A patient receiving treatment at time t is matched to n patients with a similar history of covariates but who has not received treatment up to

time t. Rather than matching on several covariates, Lu (2005) proposed matching on the PS instead. The PS is estimated via the Cox proportional hazards (PH) model for the instantaneous probability of receiving treatment given covariate history. Various matching algorithms exist including sequential matching and simultaneous pair matching (Lu, 2005). Matching on the prognostic score is another approach (Smith and Schaubel, 2015). Treatment effect estimation is subsequently performed on the matched sample.

Time-varying PS subclassification

As previously stated in Section 1.3, Leon (2011b) extended PS subclassification to the repeated measures setting. Leon (2011b) estimated PS over time by estimating the probability of treatment given covariates at each time point using multilevel logistic regression. Observations are then stratified into subclasses based on the PS, typically using quintiles, with treatment groups comparable within each subclass. As the PS can vary within a patient over time, patients may contribute observations to more than one subclass, for example if gout becomes worse propensity for treatment may increase. Treatment effects are then estimated within each subclass using the complementary log-log model and pooled. More detail on this approach is given in Section 5.4.5.1.

5.3.5 Methods used in the PhD project

Some of the most common approaches used to control for time-varying confounding have been described above. The implementation of any of these methods will be complicated by the large number of follow-up time points and large sample size, which are likely to be seen in studies using CPRD. Furthermore, availability of suitable and easy to use software is also an issue that hinders a more widespread application of these approaches in practice. For example, Daniel et al. (2013) discuss various limitations of G-methods and report that

implementation of these methods computationally intensive and not widely available in standard software.

G-computation models the association between all covariates which can become burdensome if the number of covariates is large; CPRD offers the opportunity for large number of covariates to be explored. Furthermore, G-computation models the effect of treatment history on outcome. Within CPRD, there will be many treatment histories of allopurinol; some patients may never start treatment, whilst others may start and remain on treatment, or discontinue and restart treatment (Scheepers et al., 2018). Therefore, evaluating effect of all treatment histories on outcome may not be feasible or practical. It would be easier if effect of treatment was evaluated over a short period of time, say a few years. However in relation to the topic of this PhD project, only a small proportion of patients with gout would be prescribed urate-lowering therapy shortly from diagnosis (Kuo et al., 2014), which may result in a small number of patients with distinct treatment histories in the few years post diagnosis. It was only more recently that it was shown the length of time patients adhered with allopurinol treatment using CPRD data (Scheepers et al., 2018) that may have aided this decision.

Out of the three G-methods, MSM are the most commonly used method and is considered the gold standard for modelling time-varying treatment effect in the presence of time-varying confounding. MSM are considered for use in this PhD project as they can be fitted using standard software, such as Stata, and the procedure behind them is more intuitive to other G-methods. Furthermore, the risk of model misspecification is smaller as only two models are fitted (one to model the multivariable association between covariates and treatment, and the other to model the effect of treatment on outcome), compared with G-computation which fits a separate model for each covariate. The disadvantage of MSM is that weights can become extremely large due to near violations of the positivity assumption, i.e., PS is close to zero thus taking the inverse yields large weights. Large weights often occur if too many covariates are

used in the estimation of PS, which may result in combinations of covariates having small PS i.e., only a few patients with very low PS actually receive treatment. Large weights can become amplified in patients with long follow-up times due to cumulative multiplication of small PS over time. Use of stabilised and basic weights, and truncating weights may prevent this (described further in Section 5.5.5.1).

A common issue across GLM, matching and subclassification on the covariates is that they are susceptible to poor model fit when the number of covariates is large, the study sample and number of outcomes is not large enough for all relevant covariates to be adjusted for. PS methodology aims to overcome the most common issues encountered in these three approaches. PS methods are easily understood and widely used with matching being the most popular method. However, it is currently unclear how straightforward it would be to implement any methods based on direct extension of PS to complex electronic health records (EHR) data involving repeated measures of treatment and covariates. Risk set matching (Lu, 2005), as described above, is restricted to the simple setting of comparing patients that were never treated to those that initiated treatment, who share similar PS; this approach ignores the issue of treatment adherence, and may be impracticable to find valid matches. A less restrictive approach would be to use time-varying PS subclassification proposed by Leon (2011b). This method is considered for use in this PhD project as it allows patients to repeatedly initiate and discontinue treatment over time, stratification attempts to achieve approximately homogenous subclasses where matching on the same PS may be infeasible,

5.4 Time-invariant PS subclassification

This section describes components of the analysis taken to specify, estimate and assess the PS, and how it may be used via subclassification and in subsequent estimation of treatment effect.

5.4.1 Propensity score estimation

Logistic (or logit) regression is the most frequently used model to estimate PS in the case of binary treatment. A systematic literature review conducted in 2004 found 47 out of 48 research articles relating to PS used logistic regression (Weitzen et al., 2004). Logistic regression is given in Equation 5.14 and allows control for multiple covariates through the regression component. The estimated regression coefficients measuring the association between treatment A and baseline covariates $\boldsymbol{X}$ are measured on the logit scale. For each patient, their linear prediction on the logit scale is converted into the probability of treatment i.e., the PS, that ranges from 0 to 1 (Equation 5.15).

$$ln\left[\frac{pr(A=1)}{1-pr(A=1)}\right] = \sum_{k=0}^{K} \beta_k \boldsymbol{X}_k \qquad 5.14$$

$$pr(A=1|\boldsymbol{X}) = \frac{exp(\sum_{k=0}^{K} \beta_k \boldsymbol{X}_k)}{1+\exp\left(\sum_{k=0}^{K} \beta_k \boldsymbol{X}_k\right)} \qquad 5.15$$

5.4.2 Covariate specification

Ideally covariate selection should be pre-specified based on prior clinical knowledge on the associations between covariates, treatment, and outcome. In reality this knowledge may be limited or unknown. Alternatively, statistical tests on the associations between covariates and treatment and outcome are used and covariate selection is based on pre-specified criteria.

General consensus is to include all covariates associated with outcome in the PS model (Austin et al., 2007a, Cuong, 2013, Rubin and Thomas, 1996). Brookhart et al. (2006) performed a simulation study and had shown all covariates associated with outcome, regardless of their association with treatment, should be included in the PS model; inclusion of covariates only associated with outcome in addition to confounding variables (covariates associated with both outcome and treatment) was found to have higher precision of the treatment effect without affecting bias. Inclusion of covariates only associated with treatment were found to lower precision of the treatment effect although bias was unaffected.

Consideration of including interaction terms and functional form specification via higher order linear or non-linear terms for continuous covariates is needed. Rosenbaum and Rubin (1984) and similarly Dehejia and Wahba (1999) suggested including such terms if balance was not achieved on the main effects of covariates. Assessment for balance is described further in Section 5.4.4.

Attempts to improve PS prediction are discouraged as goodness-of-fit tests and discrimination tests do not necessarily yield the best PS model that improves balance between treatment groups, and may lead to lack of common support (Brookhart et al., 2006, Westreich et al., 2011, Patrick et al., 2011, Weitzen et al., 2005).

5.4.3 Number of subclasses

Subclassification on the correctly specified estimated PS divides treated and untreated patients into J subclasses based on percentiles of the PS distribution thus all subclasses would have equal number of patients. Patients within each subclass will have similar distribution of covariates. There is no general consensus on the optimal number of subclasses. Cochran (1968) had shown a 90% reduction in bias in the treatment effect when patients were divided into quintiles, i.e., five subclasses on the distribution of a continuous covariate. Rosenbaum

and Rubin (1984) showed that this finding is still applicable when patients were spilt into five subclasses based on the PS distribution. Thus, it has become standard practice for studies to use five subclasses. This is not of course an optimal approach across board, as the number of subclasses is dependent on sample size, covariate balance, and associated bias reduction in the treatment effect.

Various simulation studies have shown increasing the number of subclasses reduces bias in the treatment effect as the distribution of covariates within subclasses become more homogenous (Lunceford and Davidian, 2004, Hullsiek and Louis, 2002, Neuhäuser et al., 2018). However, the trade-off is that variance may increase due to smaller sample size in subclasses (Lunceford and Davidian, 2004). Hullsiek and Louis (2002) recommend that a large number of subclasses as possible should be used (until subclass-specific treatment effects can no longer be estimated) in order to achieve maximal bias reduction, this study as well as Neuhäuser et al. (2018) demonstrated the use of more than five subclasses led to much smaller reductions in bias.

Note however that despite subclassification removing a significant amount of bias from confounding effects, some residual confounding is expected to remain as each subclass would contain a range of PS, thus small imbalances in distribution of covariates between the treatment groups is expected to remain (Rosenbaum and Rubin, 1983). Lunceford and Davidian (2004) suggested to fit regression analysis evaluating the effect of treatment on outcome adjusted for covariates within each subclass to further reduce within-subclass confounding.

5.4.4 Common support and covariate balance evaluation

In evaluating covariate balance, two checks are undertaken:

1) Common support, to assess whether the distribution of PS overlaps between the treated and untreated groups.
2) Covariate balance, to assess whether the distribution of covariates is similar across the treated and untreated groups after PS subclassification.

If adequate common support and covariate balance has been achieved, the treatment groups are assumed comparable. Evaluation of common support and covariate balance can be assessed either graphically or descriptively. Austin (2009a) outlines different methods that can be used to evaluate covariate balance.

Common support

Once PS is estimated, common support is evaluated before and after PS subclassification. This involves assessing the amount of non-overlap between treatment groups. This can be investigated using box plots with descriptive summaries e.g., minimum, first quartile, median, third quartile, and maximum values, or using kernel density plots e.g., histograms. Where there is overlap on the PS distribution, this implies similarity of PS between treated and untreated patients. Often it is assumed treated patients would have higher PS than untreated patients, and untreated patients would have lower PS than treated patients. If there is little overlap at the tails (i.e., the lower and upper ends) of the PS distribution, this suggests there could be a combination of covariate values that are not shared between treated and untreated patients. In case of such poor covariate balance in the tails of the PS distribution, it may be sensible to remove these patients, a process known as trimming, prior to outcome analysis (Patorno et al., 2013). However, trimming may lead to the study sample no longer being representative of the population it was derived from.

After PS subclassification, each subclass requires a sufficient number of treated and untreated patients with outcome occurring in both treatment groups to enable calculation of subclass-

specific treatment effects. In subclasses that only contain either treated or untreated patients there would be no common support and the treatment effect cannot be estimated within that subclass. Similarly, treatment effect cannot be estimated if outcome did not occur within a treatment group. Common support can be improved by reducing the number of subclasses however, covariate balance may decrease within subclasses because the range of PS would increase (Stuart, 2010).

Covariate balance

Covariate balance is usually evaluated using standardised mean difference (SMD), also known as standardised bias. SMD is used to compare means and proportions in units of the pooled standard deviation across the treatment groups. SMD is evaluated before and after PS subclassification to investigate whether SMD reduces in size for each covariate. SMD is not influenced by sample size and allows covariates measured in different units to be compared to one another on the same scale. The formula to calculate SMD for continuous covariates is shown in Equation 5.16 where $\bar{x}$ is the mean and s is the standard deviation of the covariate in each treatment group.

$$\frac{(\bar{x}_{treated} - \bar{x}_{untreated})}{\sqrt{\frac{s^2_{treated} + s^2_{untreated}}{2}}} \qquad 5.16$$

SMD for binary covariates can be derived using Equation 5.17 where $\hat{p}$ is the proportion of covariate of interest in each treatment group (Austin, 2009a).

$$\frac{(\hat{p}_{treated} - \hat{p}_{untreated})}{\sqrt{\frac{\hat{p}_{treated}(1 - \hat{p}_{treated}) + \hat{p}_{untreated}(1 - \hat{p}_{untreated})}{2}}} \qquad 5.17$$

No acceptable size of SMD has been suggested however, a common rule of thumb is SMD larger than 0.10 indicates imbalance on the covariate (Austin, 2011a). Other studies have used a more relaxed criterion of 0.25 (Stuart, 2010, Stuart and Rubin, 2008). Ho et al. (2017) argued covariates with stronger associations with outcome need better balance than covariates with weaker associations.

After PS subclassification, subclass-specific balance should be evaluated for each covariate and then averaged across all the subclasses, weighted by the proportion of patients in each subclass from the total study sample (Harder et al., 2010).

There are many other measures of covariate balance such as the variance ratio for continuous covariates (Austin, 2009a), and use of prognostic scores to ensure the risk of outcome is similar between the treatment groups (Stuart et al., 2013b). The use Kolmogorov-Smirnov test statistic and t-tests are discouraged as these measures are dependent on sample size and small imbalances are likely to be found statistically significant due to increased power in larger datasets (Austin, 2009a, Imai, 2008). If large number of covariates are entered into the PS model, multiple testing would be rife and subjected to type I error. Covariate balance was assessed using SMD in this PhD project.

5.4.5 Treatment effect estimation

Once covariate balance has been achieved, treatment effect estimation can proceed. Treatment effects are estimated within each subclass by comparing outcome directly between the treatment groups. Subclass-specific estimates are derived using GLM which may also adjust for covariates to account for residual differences between treatment groups (Equation 5.6 and Equation 5.7). The subclass-specific treatment estimates a_j are pooled as the sum of

weighted subclass-specific treatment estimates (Equation 5.18). The pooled weighted variance s^2 is shown in Equation 5.19.

$$a = \sum_{j=1}^{J} w_j a_j \quad 5.18$$

$$s^2 = \sum_{j=1}^{J} w_j^2 s_j^2 \quad 5.19$$

Two methods are commonly used to weight subclass-specific estimates. Firstly, one could weight by the subclass-specific sample size, which involves calculation of subclass-specific weight w_j as shown in Equation 5.20, where N is the total number of patients and N_j is the number of patients in subclass j.

$$w_j = \frac{N_j}{N} \quad 5.20$$

However, given that each subclass would have approximately the same number of patients, each subclass would have the same, or very similar, weight (Rosenbaum and Rubin, 1984, Cochran, 1968). Note that this may not always be the case, if for example two (or more) subclasses are collapsed into one due to low number of treated or untreated patients within particular subclasses.

Secondly, treatment effect estimates may be weighted by the inverse variance, with the Mantel-Haenszel (MH) method (Mantel and Haenszel, 1959, Mantel, 1963) being commonly used. The weights are represented as the inverse of the estimated variance of the subclass-specific treatment effect (Equation 5.21).

$$w_j = \frac{\frac{1}{s_j^2}}{\sum_{j=1}^{J} \frac{1}{s_j^2}} \quad 5.21$$

The advantage of using inverse variance weights is that subclass-specific estimates with larger variance would have smaller weights thus contributing less to the overall treatment effect estimate.

Rudolph et al. (2016) simulation study had shown that using inverse variance weights outperformed weighting by the sample size in PS subclassification but only when treatment effects across the subclasses are homogenous. When the treatment effect varied across subclasses, the weighting by sample size performed better, especially if over 10 subclasses were used. In this PhD project, both approaches were utilised, as appropriate.

The assumption of constant treatment effects across subclasses can be verified using all observations from all subclasses. A GLM model initially regresses outcome on treatment and the PS subclasses. The incremental contribution of the interaction term between PS subclasses and treatment is then tested using the likelihood ratio test (LRT) by comparing the -2 difference in log-likelihood of the GLM models with and without the interaction. An insignificant test would indicate that the effect of treatment does not vary significantly between subclasses, hence results can be pooled. If however the LRT is significant, this would indicate that the treatment effect varies across subclasses and the subclass-specific estimates should be reported, alongside with the distribution of covariates (Leon, 2011b).

One should note that differences between treatment groups may remain within PS subclasses and may be difficult to achieve balance on particular covariates, even after increasing the number of PS and re-estimating PS. One solution would be to adjust for imbalanced covariates within subclass-specific treatment effect estimation. However, the number of covariates to

adjust for may be restricted by sample size and frequency of outcome (Vittinghoff and McCulloch, 2006). One should aim to achieve balance for the majority of covariates prior to adjusting for them.

5.4.5.1 Survival analysis

Survival analysis, the method of interest to estimate treatment effectiveness in this PhD project, is the analysis of times from some time origin (such as baseline) to occurrence of outcome of interest, for example death.

Survival data, though inherently continuous, cannot be processed in the usual manner based on assumed normality, due to their tendency to follow non-symmetrical distributions and a substantial proportion of survival times are usually censored. Censored observations are those that have not been fully observed, and this can take the form of right, left or interval censoring. Such observations cannot be ignored as they carry potentially important information about the effect of treatment and covariates. The form of censoring that is easiest to model, and occurs most commonly, is right censoring. An observation is right censored if at the end of study, say t_c, the exact survival time, t , is not known and all we can state is that $t > t_c$. This may occur if, for example, a patient is lost in the follow-up period or the outcome of interest has not been observed at the end of the study.

Left censoring is rare and occurs when the outcome is observed prior to the start of study; interval censoring occurs when the outcome is observed between two specified time points.

Only right censoring was considered in this PhD project. Furthermore, it was assumed that the censoring mechanism is non-informative, i.e., the censoring is not related to any factors associated with the actual survival time.

There are three key functions of interest; the survival function, probability density function, and hazard function, which are now defined: Let T be a non-negative random variable denoting survival time. The survival function $S(t)$ measures the proportion of patients who have not experienced outcome beyond time t (Equation 5.22).

$$S(t) = P(T \geq t) \tag{5.22}$$

The survival function is a monotone, non-increasing function with boundary conditions $S(0) = 1$ and $S(\infty) = 0$. The probability density function $f(t)$ is the unconditional probability of outcome occurring before time t (Equation 5.23).

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} \tag{5.23}$$

The hazard function $h(t)$ is defined in Equation 5.24 and can be rewritten as the ratio between the probability density function and survival function.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{5.24}$$

So $h(t)\Delta t$ is approximately the probability that the outcome occurs in the interval $(t, t + \Delta t)$ given that the patient has not experienced outcome up to time t. Thus $h(t)$ is a conditional function and can also be thought of as the risk of outcome occurring immediately after t. Note that $h(t)$ is a rate, not a proportion, so it can take on any value between zero and infinity.

$h(t)$ can assume many different forms (increasing, decreasing or monotone over time) and is useful in explaining the way in which the risk of an outcome changes over time.

5.4.5.2 Cox proportional hazards model

Cox survival model (Cox, 1972) is used in this thesis and its properties are summarised in this section. The hazard rate $h(t)$ is given in Equation 5.25: $\boldsymbol{X} = X_0, X_1, X_2, \dots, X_K$ is a vector of covariates with $X_0 = 1$, and $\boldsymbol{\beta} = \beta_0, \beta_1, \beta_2, \dots, \beta_K$ is a vector of unknown regression parameters to be estimated; for treatment A, the regression parameter to be estimated is denoted α; and $h_0(t) > 0$ is some arbitrary function of time, also known as the baseline hazard function, describing the risk when $A = 0$. $h_0(t)$ represents a reference point that depends on time, just as the intercept denotes a reference point in other types of regression models. Cox regression parameters are estimated using the partial likelihood, which depends only on the parameters of interest.

$$h(t) = h_0(t) exp(\alpha A + \sum_{k=0}^{K} \boldsymbol{\beta}_k \boldsymbol{X}_k) \qquad 5.25$$

The hazard ratio between two treatment groups when all covariates are fixed is simply $exp(\alpha)$, interpreted as the relative risk of outcome.

This model is termed PH because the hazard ratio for two patients with time-invariant covariates is constant over time and has a relative risk interpretation. The Cox model is therefore a semi-parametric model because the exact form of $h_0(t)$ is left unspecified and a strong assumption of proportionality of hazards is made. It is this semi-parametric property of the Cox model that has made it the most popular model used in survival analysis when there is doubt about correct parametric specification. Note that specification of $h_0(t)$ in Equation 5.25 would lead to parametric models, such as Weibull.

It is important that the PH assumption is tested if the Cox model is to be used; there are numerous methods of doing this, both numerically and graphically. Schoenfeld residuals, defined as the difference between the observed covariate value minus its expected value for

that patient at its failure time, are commonly used to assess the PH assumption. The residuals are independent of time, therefore plotting the residuals against time would have a random pattern. Also, one can test whether the slope is equal to zero in a linear regression model of the Schoenfeld residuals on time. If the p-value <0.05, the slope is not equal to zero and indicates deviation from PH and the log-hazard ratio changes over time (Grambsch and Therneau, 1994).

Alternatively, the Kaplan-Meier survival curves (or survival times) can be plotted for each treatment (or covariate) group (Sedgwick, 2014). If the two survival curves crossover, it indicates violation of the PH assumption. Violation of the PH assumption may be resolved by including an interaction term between the particular covariate and time in the Cox model.

5.4.5.3 Anderson & Gill method

The Cox PH model considers time to a single outcome but cannot be used in the case of repeated outcome (Andersen and Gill, 1982). The Anderson and Gill model is a common method used in the analysis of repeated outcome and is a simple extension of the Cox PH model. The Anderson and Gill model assumes repeated outcomes times are independent to one another and ignores the order of occurrence of the outcomes. Therefore, the hazard of experiencing an outcome at time t, is the same regardless if outcome occurred previously. Each recurrent outcome is assumed to follow the Cox PH model (Equation 5.25).

5.5 Time-varying PS subclassification

Leon et al. (2001) extended the approach of PS subclassification to the setting of repeated treatment and covariates over time. In theory, time-varying covariates and time-varying treatment may be measured in continuous time thus it is possible to know the exact date

when covariate and treatment status change, such as in EHR where patient observations are recorded continuously. This would allow PS to be estimated and updated at each change of covariate and treatment status. However, in practice, this set-up may yield an unwieldly large dataset; consequently, issues with fitting an effective PS model would arise due to large number of repeated measurements for each patient, and potentially high correlations between repeated measurements resulting from lack of change between a high proportion of time points. Calculating the PS in this way is computationally intensive. As a compromise, treatment and covariate values may be ascertained in pre-set time intervals, but the specifics would be dependent on the length of follow-up and clinical relevance. Such discrete time approach commonly assumes treatment and covariates do not change within an interval. It is therefore important, as well as challenging, that both the ease of statistical modelling and clinical relevance are considered when making decisions about the width of time intervals.

Published methods of subclassifying patients on time-dependent PS usually assume such discrete time. For example, in Leon et al. (2003) study, patients were followed up semi-annually for the first five years and annually thereafter up to 20 years. Follow-up was divided into intervals of variable length corresponding to which treatment intensity the patient was prescribed; median number of intervals of 8 (range 1, 65) per patient. Thus, a patient had a mixture of discrete intervals taking various treatment intensity. This created a two-level hierarchical data structure with repeated measurements (level 1) clustered within a patient (level 2).

Since their initial paper, subsequent publications by Leon et al. have shown the methodology evolve. Various simulations studies have evaluated the performance of time-varying PS subclassification in various settings; this method is capable of adjusting for time-varying covariates and estimating unbiased treatment effect for survival outcomes (Leon and Hedeker, 2005), continuous outcomes (Leon and Hedeker, 2007a) and repeated binary

outcomes (Leon, 2011b). Leon et al. (2012b) also proposed full matching, a type of subclassification. This involves creating a series of matched sets, where a matched set contains at least one treated interval and one untreated interval that have the same PS.

Application of these methods have commonly been applied to the evaluation of anti-depressants on outcome (Leon et al., 2001, Leon et al., 2003, Leon et al., 2011, Leon, 2011b) but has received little attention elsewhere.

5.5.1 Propensity score estimation

Following specification of hierarchical data structure as described above, PS needs to be estimated in each time interval.

Use of logistic regression as specified in Section 5.4.1 and Equation 5.14 would be erroneous as the model would make an unrealistic assumption that repeated measurements within a patient are independent given covariates. Instead, mixed effects (also termed multi-level) models can be used to analyse such hierarchical data where the variability in outcome is attributable to both repeated measurements within patients and between patients (Rabe-Hesketh and Skrondal, 2012). Mixed effects models incorporate both fixed and random effects. Leon et al. (2001) used such models, specifically random intercept ordinal logistic regression model, to estimate PS for each dose of treatment (ordinal treatment dose) over time. For binary treatment, random intercept logistic regression would be used.

Extending Rosenbaum and Rubin (1983) notation to the repeated measures set-up, the PS at time t is defined in Equation 5.26; r is the patient-specific random intercept that is assumed to be normally distributed with mean zero and variance σ_r^2 and $\overline{\boldsymbol{X}}_t$ denotes covariate history (that include time-invariant and time-varying covariates) up to and including time $t-1$ to ensure temporal ordering between covariates and treatment.

$$e(\bar{\boldsymbol{X}}_t, r) = pr(A_t = 1 \mid \bar{\boldsymbol{X}}_t, r),\ 0 < e(\bar{\boldsymbol{X}}_t, r) < 1 \qquad 5.26$$

The random intercept logistic regression model is specified in Equation 5.27 which is an extension of Equation 5.14; it now estimates the probability of treatment at time t given the random intercept and covariate history.

$$ln\left[\frac{P(A_t = 1)}{1 - P(A_t = 1)}\right] = \sum_{k=0}^{K} \beta_k \bar{\boldsymbol{X}}_{tk} + r \qquad 5.27$$

$$pr(A_t = 1|\bar{\boldsymbol{X}}_t) = \frac{\exp\left(\sum_{k=0}^{K} \beta_k \bar{\boldsymbol{X}}_{tk} + r\right)}{1 + \exp\left(\sum_{k=0}^{K} \beta_k \bar{\boldsymbol{X}}_{tk} + r\right)} \qquad 5.28$$

Within a patient, PS can change over time due to time-varying covariates included in $\bar{\boldsymbol{X}}$. The random intercept can be thought of as patient-specific regression coefficient; inclusion of the random intercept in PS estimation allows each patient to have a separate intercept so the PS at baseline $(t = 0)$ would be higher or lower compared to the estimated intercept b_0.

Measurements within patients may be correlated, which is accounted for in mixed models via σ_r^2. The estimation of the covariance structure is usually performed under certain specific structure assumptions. Independent, exchangeable, and unstructured covariance structures are commonly used in practice. The independent structure is the most simplistic and assumes repeated measurements within a patient are independent. The exchangeable structure assumes correlations between subsequent measurements are the same, irrespective of time. The least restrictive is the unstructured structure where all correlations are assumed to be different. In practice, there is no strict rule on which covariance structure is optimal, and in this PhD project, various options have been considered.

5.5.2 Covariate specification

As discussed in Section 5.4.2, covariates associated with outcome should be included in estimation of the PS at a single time point. However, it is unclear whether the same applies in the case of estimation time-varying PS.

In a repeated treatment and covariates setting, Leon and Hedeker (2007b) performed a simulation study on the impact of a misspecified PS model in treatment effect estimation. That study found omitting confounding variables from the PS model yielded biased treatment effects. Omitting continuous time-varying confounding variable increased bias the most followed by time-varying binary confounding variables, and then baseline confounding variables. Leon and Hedeker (2007b) also found omitting a confounding variable in PS estimation that were highly correlated with other confounding variables yielded less bias compared with omitting a confounding variable with lower correlation. Therefore, it was suggested to adjust for time in treatment effect estimation to lessen the impact of omitting a time-varying confounding variable in PS estimation to the extent that confounding variable is associated with time.

Many of Leon's research articles hypothesised that the included covariates in the PS were associated with treatment however, there was no discussion whether these covariates were also associated with outcome (Leon et al., 2003, Leon et al., 2001, Leon, 2011a). This differs to time-invariant PS where Brookhart et al. (2006) simulation study suggested all covariates associated with outcome should be included in PS estimation (Section 5.4.2).

As described above, the PS at time t is estimated using covariate history, therefore the PS should contain lagged covariates representing covariate values in earlier time intervals. For example, Leon et al. (2003) included history of disease (e.g., prior number of episodes), trajectory of symptom severity prior to treatment (stable, increasing, decreasing), and prior

treatment use in the PS model. Leon and Hedeker (2007a) recommended including time as a covariate in the PS model if it was hypothesized there was temporal trend in being treated.

5.5.3 Number of subclasses

Once PS are estimated within intervals, the intervals are then divided into subclasses J based on the PS distribution leading to equal or very similar number of intervals in each subclass. A patient may contribute to multiple subclasses if their PS changes over time.

Similarly to PS subclassification at a single time point (Section 5.4.3), many studies had used quintiles based on the time-varying PS distribution (Leon et al., 2001, Leon et al., 2003, Leon and Hedeker, 2005). Later studies by Leon and Hedeker (2007a) and Leon (2011b) performed simulation studies and determined that between four and five subclasses were required to remove 80-90% of bias in the treatment effect in this repeated measures setting. They noted that statistical power decreased as more subclasses were used; these studies did not consider more than five subclasses.

As previously noted in Section 5.4.3, the number of subclasses is dependent on sample size, covariate balance, and bias reduction in the treatment effect when PS is performed at baseline. Logically the same is assumed here.

5.5.4 Common support and covariate balance evaluation

Similarly to PS subclassification at a single time point (Section 5.4.4), common support is evaluated before and after PS subclassification. It is first necessary to check there is sufficient number of treated and untreated intervals with outcome occurring in both treatment groups to enable calculation of subclass-specific treatment effects. There is no consensus on the minimum number of observations needed within a cell. Leon et al. (2001) and Leon et al.

(2003) had a minimum of 4% for each treatment within a subclass in their analyses whilst Leon and Hedeker (2007a) suggests a minimum of 5-10 observations in each cell of the PS subclass and treatment contingency table.

Covariate balance was assessed before and after PS subclassification. Leon and Hedeker (2007a) and Leon (2011b) assessed balance by using mixed models to regress treatment on each covariate (that was included in the PS model) individually; the estimated regression coefficient and statistical significance was noted for each covariate. Subsequently, each mixed model was adjusted for PS subclasses. If the addition of the PS subclasses to the model attenuated the regression coefficient for the covariate towards the null and had become statistically insignificant, then it was assumed that balance was achieved on that covariate. If balance was not achieved on that covariate, it was adjusted for in the estimation of subclass-specific treatment effect.

Note however that as stated above in Section 5.4.4, use of significance testing for covariate imbalance is dependent upon sample size and its use is discouraged as small imbalances are likely to be found statistically significant due to increased power in larger datasets, for example when using CPRD (Austin, 2009a, Imai, 2008); intuitively the same applies here. Therefore, assessment of change in the covariate regression coefficient alone is better placed to ascertain whether balance was achieved.

A more appropriate approach to examining the magnitude of imbalance after PS subclassification would be to use SMD as described in Section 5.4.4. In the repeated measures setting, covariates measured in each interval would contribute to calculating SMD that assumed the repeated measurements were independent; $\bar{x}$ in Equation 5.16 represents the mean covariate across all intervals, and $\hat{p}$ in Equation 5.17 represents the proportion of the covariate of interest across all intervals. Any covariates which were not successfully balanced

across treatment groups should be adjusted for when estimating subclass-specific treatment effects.

5.5.5 Treatment effect estimation

Once balance has been achieved, evaluation of treatment effect can proceed. There are two key ways in which PS subclasses can be accounted for in treatment effect estimation. Firstly, one could directly adjust for the PS subclasses as covariate in the outcome model that estimates the overall association between treatment and outcome (Leon et al., 2001).

Alternatively, subclass-specific treatment effect estimates could be obtained and pooled via the MH method as described in Section 5.4.5 using Equations 5.18, 5.19 and 5.21. To enable pooling, the assumption of homogeneity of treatment effects across subclasses would first need to be tested, as explained before.

5.5.5.1 Discrete-time survival model

Discrete time survival models can be used when continuous time line is split into intervals, thus several patients would share the same analysis time (Cox, 1972). Each interval contains occurrence of outcome which allows one to model the probability that outcome occurred in each interval, conditional that the outcome had occurred until then (Rabe-Hesketh and Skrondal, 2012).

The survival function in discrete time at time t_j is defined as the probability that survival time T is greater than t_j, or in other words the probability of not having outcome by time t_j (Equation 5.29).

$$S(t_j) = pr[T \geq t_j] = \sum_{k=j}^{\infty} f(t_j)$$

$$f(t_j) = pr[T = t_j] \qquad 5.29$$

The hazard of failure (outcome) is defined as the probability of failure at time t_j given that the patient has not had outcome to that point (Equation 5.30).

$$h(t_j) = pr[T = t_j | T \geq t_j] = \frac{f(t_j)}{S(t_j)} \qquad 5.30$$

The survival function can be written in terms of the hazard function of all previous time points such that at time t_j, a patient cannot have outcome at time t_1, t_2, up until t_{j-1} (Equation 5.31). Equation 5.31 is equivalent to the survival function in continuous time (Equation 5.22).

$$S(t_j) = [1 - h(t_1)][1 - h(t_2)], \dots, [1 - h(t_{j-1})] \qquad 5.31$$

If continuous time is divided into discrete intervals with constant hazard within each interval, the Cox PH model is analogous to the binomial regression model with a complementary log-log link (Equation 5.32). Complementary log-log regression allows one to estimate the hazard ratio in discrete time, and are the same as if the Cox model was fitted in continuous time and assumes PH. In the case of time-varying treatment, the PH assumption does not need to be satisfied as the hazard function is estimated within each interval. As treatment changes over time, the association between treatment and outcome estimated within each interval may differ across intervals, therefore the PH assumption may not be satisfied. Interval-specific treatment effect estimates are pooled together to obtain an overall estimate.

$$\ln\left[-\ln\left(1-h(t_j)\right)\right] = \alpha A + \sum_{k=0}^{K} \boldsymbol{\beta}_k \boldsymbol{X}_k \qquad 5.32$$

One should note that this particular model does not account for duration of time within an interval and simply models whether outcome had occurred in each interval. To account for repeated correlated intervals within a patient, one can include patient-specific random intercept r (as previously described in Section 5.5.1) to Equation 5.32, and shown in Equation 5.33 (Hedeker et al., 2000, Rabe-Hesketh and Skrondal, 2012, Austin, 2017).

$$\ln\left[-\ln\left(1-h(t_j)\right)\right] = \alpha A + \sum_{k=0}^{K} \boldsymbol{\beta}_k \boldsymbol{X}_k + r \qquad 5.33$$

The random intercept complementary log-log regression model has frequently been used to estimate subclass-specific treatment effects after time-varying PS subclassification (Leon, 2011b, Leon, 2011a, Leon et al., 2012a).

5.6 Marginal structural models

As stated in Section 5.3.4, MSM are used to model the effect of time-varying treatment on potential outcomes in the presence of time-varying confounding effects (Robins, 2000). At the start of this PhD project (in 2013), very few EHR studies had used MSM to estimate treatment effect however, its use has increased since; examples include warfarin use and risk of bleeding using GPRD data (Platt et al., 2012), effectiveness of beta-blockers on mortality using GPRD data (Delaney et al., 2009), and bisphosphonates use and risk of infection using claims data (Xue et al., 2017).

Two modelling steps are involved in MSM. The first step requires the correct specification of the PS model for estimation of weights; the second step uses the estimated weights to create a pseudo-population (where there is no confounding) within which an unbiased treatment effect can be estimated.

Time-varying covariates and treatment can be measured in either continuous or discrete time. The same data structure as described for time-varying PS subclassification in Section 5.4.5.1 is assumed; treatment and covariates are measured repeatedly in pre-set time intervals creating a two-level hierarchical data structure with repeated measurements (level 1) clustered within a patient (level 2).

5.6.1 Weight estimation

The purpose of weights is to balance the distribution of covariates across the treatment groups in the pseudo-population in all time intervals. Weights w_t are assigned to each time interval t resulting in time-varying weights. As previously described in Section 5.3.3, the weights reflect by how much observations are under-represented or over-represented in the study sample compared to the pseudo-population in the absence of no confounding. The ideal properties of the weight distribution are mean equal to 1, normally distributed, and a narrow range at all time points. In practice, this may not be feasible if the number of time points is large, therefore the overall weight distribution can be examined instead (Cole and Hernan, 2008).

Section 5.3.4 described the estimation of unstabilised weights (Equation 5.12). The distribution of unstabilised weights is likely to be skewed, with mean weight is likely to deviate from one substantially and have large variance, inflated by the presence of extremely large weights. The mean and standard deviation of weights with minimum and maximum weight

values should be reported. These issues are indications that the positivity assumption is violated (i.e., PS is very close to zero), the PS model could be misspecified (Cole and Hernan, 2008) or that residual confounding may be present (Jackson, 2016).

Using the intervals as the unit of analysis, pooled logistic regression is used to estimate PS in each interval, given baseline and time-varying covariates (Fewell et al., 2004). Logistic regression is 'pooled' as a patient contributes multiple times to the model which pools treatment and covariates across intervals into a single sample however, the model does not account for when treatment and covariates occur within an interval (Ngwa et al., 2016).

Stabilised weights

Robins (2000) recommended the use of stabilised weights, a modification of unstabilised weights. The formula for stabilised weights is shown in Equation 5.34. The numerator of the stabilised weight is the probability of observed treatment conditional on past treatment history $\bar{A}_t$ (treatment history up to and including time $t-1$) and time-invariant covariates $\boldsymbol{Z}$. $\boldsymbol{Z}$ is a subset of $\boldsymbol{X}$ that includes both time-invariant and time-varying covariates. The denominator of the stabilised weight is the same as the denominator of the unstabilised weight.

$$sw_i = \prod_{t=0}^{T} \frac{pr(A_t|\bar{A}_t, \boldsymbol{Z})}{pr(A_t|\bar{A}_t, \bar{\boldsymbol{X}}_t)} \qquad 5.34$$

By including treatment history and baseline covariates in the numerator and denominator, the stabilised weights reflect the incremental effect of time-varying covariates on the probability of treatment independently of other covariates (Equation 5.34). Therefore, the stabilised weights are less variable and skewed, with smaller variance and range, with mean weight closer to one, than unstabilised weights. Stabilised weights tend to decrease over time

as the denominator accounts for time-varying covariates the probabilities should be on average higher (as more recent covariate status would have stronger association with treatment than time-invariant covariates) than the probabilities in the numerator for each time interval (Xiao et al., 2010). Use of stabilised weights also mean the estimated treatment effect would be more precise than using unstabilised weights. The disadvantage of stabilised weights is that confounding would remain from baseline covariates and treatment history due to its inclusion in the numerator; in the pseudo-population as treatment is only randomised within levels of the time-invariant covariates and treatment history, one needs to adjust for these covariates in the outcome model estimating treatment effect (Cole and Hernan, 2008).

Alternatively, one could estimate basic stabilised weights where the numerator is the proportion of treated patients at time t (Equation 5.35). Similarly, basic stabilised weights are less variable than unstabilised weights and leads to more precision in the estimated treatment effect. Use of basic stabilised weights avoids the need to condition on treatment history and baseline covariates in the outcome model (Talbot et al., 2015).

$$bsw_i = \prod_{t=0}^{T} \frac{pr(A_t)}{pr(A_t|\bar{A}_t, \bar{X}_t)} \qquad 5.35$$

The PS for the numerator and denominator for all types of weights are estimated separately via pooled logistic regression models.

Normalised weights

Distribution of stabilised weights may still be undesirable if the mean deviates from one and the variance is large. Xiao et al. (2010) proposed normalised weights that ensure the mean weight of one in all intervals; weights would be less extreme resulting in reduced variability of weights. Both stabilised and unstabilised weights can be normalised. Calculation of normalised

unstabilised weights nw_i is shown in Equation 5.36 where $N(t)$ is the total number of patients in interval t and $R(t)$ is the number of patients at risk of outcome in interval t. Calculation of normalised stabilised weights is shown in Equation 5.37.

$$nw_i = \frac{w_i N(t)}{\sum_{i \in R(t)} w_i(t)} \quad 5.36$$

$$nsw_i = \frac{sw_i N(t)}{\sum_{i \in R(t)} sw_i(t)} \quad 5.37$$

Xiao et al. (2010) performed a simulation study evaluating effect of different weights on treatment effect estimation. Unstabilised weights yielded the largest standard error of the treatment effect whilst use of stabilised weights significantly reduced standard error as expected. Use of normalised weights (both unstabilised and stabilised) yielded the smallest standard errors of the treatment effect.

Censoring weights

Longitudinal studies may be susceptible to selection bias due to loss of follow-up. If patients were censored prior to outcome occurring, for example if they died, bias in the treatment effect is introduced if covariates differ between censored patients and those who remain in the study and when these covariates are also associated with outcome.

Censoring can be viewed as a time-varying confounding variable. Censoring weights are defined as the probability of being censored at each time point given time-varying and time-invariant covariates. Censoring weights reweight the study sample and create a pseudo-population where censoring did not occur. Multiplying treatment weights with censoring weights at each time point creates a pseudo-population where there is no loss of follow-up and no confounding.

Weight truncation

Extreme weights may persist in some situations, regardless of which weight estimation approach is used. There are a few approaches that can be adopted to reduce such persistent extreme weights. One could remove covariates from the PS model that are weakly associated with outcome and cause extreme weights however, this may not always be possible if all covariates are strongly associated with outcome. Another approach is exclusion of observations that have large weights (known as trimming) however, one would need to define a threshold at which weight is considered high enough for an observation to be removed; furthermore, generalisability of the sample may be compromised.

The most common approach is weight truncation. Truncation involves setting the value of weights greater than and lower than a certain percentile cut-off to the values of these percentiles cut-off points. As the weights are progressively truncated, the treatment effect estimate progressively becomes more biased however the standard error reduces. Typically, weights are truncated at 1% and 99% of their distribution and typically removes the majority of extreme weights (Cole and Hernan, 2008). Cole and Hernan (2008) described weight truncation as the trade-off between bias and precision in the treatment effect.

There is little guidance in the optimal level of truncation. Xiao et al. (2013) suggested the optimal level of truncation should be based on minimising the mean squared error (MSE, defined as the mean squared difference between the estimated and actual values). A simulation study showed truncating weights fixed at 99.5% and 99% percentiles performed similarly in terms of bias and variance on the treatment effect with truncation based on MSE (Xiao et al., 2013).

5.6.2 Intention-to-treat, per-protocol, and as-treated principles

Weight estimation is dependent on whether treatment effect is estimated under the intention-to-treat, per-protocol, or as treated principles. For this PhD project, interest is in estimating the actual effect of treatment via as-treated analysis.

5.6.2.1 Intention-to-treat

Intention-to-treat analysis assumes that once a patient has initiated treatment they remain on that treatment throughout the rest of follow-up, therefore the effect of initiating treatment is estimated (Cole and Hernan, 2008). Intention-to-treat approach is most appropriate in situations where treatment adherence after initiation is low. Weight estimation would be simplified as covariates associated with treatment initiation only need to be considered as the assumption of exchangeability applies up to treatment initiation; weights are estimated up until treatment is initiated, thereafter the probability of treatment is assigned a value of one for the remainder of the follow-up.

5.6.2.2 Per-protocol analysis

The magnitude of the treatment effect estimate and its precision resulting from intention-to-treat analysis is dependent on whether patients adhered with treatment. Per-protocol analysis may be performed by restricting analysis to patients who adhered with treatment by censoring patient follow-up time at the point they stop treatment. This artificial censoring may be dependent on covariates and outcome and is a type of selection bias. Analysis would have to account for both confounding by indication and confounding from selection bias by estimating weights separately when patients initiate treatment and when patients continue with treatment (Yang et al., 2014, Danaei et al., 2013, Yang et al., 2015b).

5.6.2.3 As-treated analysis

As-treated analysis is where patients are analysed according to whether they receive treatment or not. This approach estimates the actual treatment effect. Different scenarios may occur where patients (1) may never initiate treatment, (2) initiate and remain on treatment until the end of follow-up, (3) discontinue treatment once initiated, (4) initiate, discontinue and then restart treatment; treatment discontinuation and/or restarting treatment may occur repeatedly. This scenario is expected to be seen within CPRD data.

Weight estimation becomes more complex as weights would need to be estimated separately if confounding effects were found to differ for each scenario. Studies rarely consider these complexities of as-treated analyses and assume the effect of confounding variables are the same for all scenarios (Yang et al., 2014).

Yang et al. (2014) conducted a systematic review identifying pharmaco-epidemiologic studies published in 2012 that attempted to account for patients not adhering with treatment. The authors extracted information from 20 eligible studies and found eight studies conducted as-treated analyses, six intention-to-treat analyses, and three per-protocol analyses.

Yang et al. (2015b) performed a simulation study comparing impact on treatment effect estimate when weight estimation was considered under four different scenarios: (1) intention-to-treat analysis, (2) acknowledged confounding effects differed between patients initiating treatment and patients continuing treatment, with two separate weights estimated, (3) confounding variables had the same effect on initiating treatment and continuing treatment, and (4) only time-invariant confounding variables were considered.

Approach 1 performed poorly in the presence of non-adherence and when treatment effect was non null; bias increased approximately by 1% for each 1% increase in the proportion of patients discontinuing treatment. Standard error of the treatment effect was larger compared with the other approaches. Approach 1 should only be used if adherence to treatment is high.

Incorrectly assuming confounding effects are the same for patients initiating and continuing treatment (approach 3), resulted in biased treatment effect due to incomplete control of confounding effects. Modelling weights separately for treatment initiation and adherence (approach 2) yielded unbiased treatment effect and lower standard errors compared with other approaches. Approach 4 was found to estimate biased treatment effect with larger standard errors than the other three approaches.

A limitation of Yang et al. (2015b) study is that it assumed only two time points and only a continuous outcome was considered. Graffeo et al. (2018) simulation study focused on estimating weights for patients who repeatedly initiated and discontinued treatment with time-to-event outcome. Weights were estimated in continuous time using Cox regression models. The authors found that bias in the treatment effect reduced when weights were modelled separately for treatment initiation and discontinuation compared with an intention-to-treat analysis.

5.6.3 Covariate specification

Yang et al. (2014) systematic review of pharmaco-epidemiological studies that used MSM found studies generally selected covariates based on either previous knowledge or using statistical criterion. Within MSM, Lefebvre et al. (2008) performed a simulation study and found the PS model should include covariates that are associated with outcome to increase precision in the estimated treatment effect however, including covariates that were only associated with treatment increased bias and standard error in the estimated treatment effect. These findings were similar to what Brookhart et al. (2006) found in estimating time-invariant PS (Section 5.4.2) although inclusion of covariates only associated with treatment did not affect bias.

As discussed for time-varying PS subclassification approach, the PS estimation is conditional on covariate history therefore, the PS model should contain lagged covariates representing covariate values in earlier time intervals (Robins et al., 2000). Neugebauer et al. (2007) proposed history restricted MSM where a shorter history of covariates is considered assuming it adequately captures covariate history. The PS model needs to adjust for treatment history; this can be achieved by including a counter of previous treatment as a covariate (Graffeo et al., 2018).

The functional form of covariates needs to modelled accurately. Cole and Hernan (2008) found modelling continuous covariates as categorical variables affected the bias-variance trade-off in the treatment effect. As the number of categories for a continuous covariate increased, this led to better control of confounding (as more information on the covariate was available). However, the standard error of the treatment effect estimate increased which may have resulted from increased range and standard deviation of weights, mean weight deviating from one, and a small number of patients that have a certain combination of covariates.

Instead, one could directly model continuous covariates as a non-linear function as it reduces the number of parameters estimated by assuming the intercept is a smooth function thus avoiding the need to categorise covariate, and better control for confounding without assuming linearity.

Cole and Hernan (2008) recommended the use of restricted cubic regression splines. Regression splines offers a way to examine the functional form of expected treatment with a function of the linear predictor, where a spline is a smoothed curve. The amount of smoothness is dependent on the number of parameters (or degrees of freedom (df)) used by the spline. Estimation of one parameter for the covariate uses up one df thus imposes linearity. Estimation of two parameters for the covariate uses up two df and so on. Increasing the number of parameters allows one to capture more complicated trends. The number and

location of knots need to be specified; a knot is where two regression splines meet; knots are commonly placed at the 5% and 95% percentiles of the distribution (Rutherford et al., 2015).

Alternatively, one could model non-linear functions using fractional polynomials (FP). This is where covariate X is transformed to X^p where p is chosen from a set of candidates $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where X^0 denotes the logarithm of X; this is a first degree FP. To model more complex non-linear functions, second degree FP where the covariate is represented by two power transformations (p and q) can be fitted. The linear predictor takes the form $\beta_o + \beta_1 X^p + \beta_2 X^q$ or $\beta_o + \beta_1 X^p + \beta_2 X^p \log(X)$ when $p = q$. The best fitting combination of powers is the regression model that has the lowest deviance (defined as twice the negative log likelihood). Only a small number of power transformations for p was considered as they offer considerable flexibility to capture non-linear patterns, and inclusion of other powers, for example p=-3, may yield extreme observations and small improvement in model fit (Royston et al., 1999).

A simulation study had shown that failure to correctly model the functional form of a time-varying covariate resulted in bias due to imbalance in that covariate between treatment groups. Both FP and regression splines performed well yielding small bias in the treatment effect (Kyle et al., 2019).

5.6.4 Covariate balance evaluation

The purpose of weighting is to ensure balance is achieved on history of treatment and covariates between treatment groups at each time point. Assessment of balance is often ignored in MSM (Vandecandelaere et al., 2016).

SMD as described in Section 5.4.4 and defined in Equations 5.14 and 5.15 can be used to assess balance for covariates included in the PS model between treatment groups at each time point.

Various rules of thumb had been used to define imbalance for MSM. Vandecandelaere et al. (2016) considered SMD greater than 0.25 as unacceptable imbalance between treatment groups; SMD greater than 0.10 has also been used as a cut off to indicate imbalance (Lavikainen et al., 2016). Jackson (2016) used trellis plots that illustrated SMD of covariates for each pattern of treatment history e.g., SMD between treatment groups at time point 3 is evaluated for covariates measured at baseline, first, second and third time points. Assessing balance in this way can become cumbersome particularly if the follow-up is long and if there are more than two treatment regimes.

Unlike PS subclassification, common support of weights between treatment groups is not a requirement as balance can still be achieved even if there is little overlap in the weight distributions of treated and untreated patients.

5.6.5 Treatment effect estimation

MSM allow one to correctly estimate the effect of time-varying treatment in the presence of time-varying confounding variables that are affected by past treatment. They estimate the treatment effect that is composed of the direct effect of current treatment on outcome (where treatment is not mediated through covariates) and past treatment effect that is mediated through covariates on outcome. In this thesis, marginal structural Cox model is used to estimate treatment effect.

5.6.5.1 Marginal Structural Cox model

The marginal structural Cox model is specified in Equation 5.38 where $\bar{A}_t$ denotes treatment history up to time t, $T_{\bar{A}}$ represents patient time to outcome had they followed a particular

treatment history α is the regression parameter to be estimated for treatment history, and $g(.)$ is some function of treatment history.

$$h_{T_{\bar{A}}}(t) = h_0(t)exp\ (\alpha g(\bar{A}_t)) \qquad 5.38$$

Specification of $g(.)$ allows one to estimate different treatment effect estimates. One can evaluate the effect of all treatment histories on outcome; this is known as a saturated MSM. There would be 2^T treatment histories, therefore saturated MSM model would be complex to fit if the number of intervals is large. Alternatively, the effect of treatment at each time point could be estimated, which too may be complex if the number of time points is large. Effects of particular treatment regimens (never treated vs. always treated), or a summary measure of treatment, for example total number of treatment occurrences over time, could be evaluated.

However, use of summary measures to capture exposure to treatment may not be optimal; an assumption would be made that effect of allopurinol among those who are continuously treated (without periods of discontinued use) is the same as the effect of allopurinol among patients with the same total number of periods they were treated in but who also had periods of discontinued treatment. Furthermore, it would also be assumed that total past treatment has the same effect on outcome regardless of how long ago that treatment was given.

Hernan et al. (2000) estimated the effect of current treatment at each time point on outcome in intention-to-treat analysis. Equation 5.38 becomes Equation 5.39 to reflect estimation of current treatment effect.

$$h_{T_{\bar{A}}}(t) = h_0(t)exp\ (aA) \qquad 5.39$$

In this thesis, focus was on estimating current treatment use via as-treated analyses thereby estimating the effect of actual treatment. However, as allopurinol use is intermittent, this approach would implicitly assume that there is no cumulative effect of treatment on outcome i.e., it does not model the effect of past treatment that is mediated through covariates on outcome, only the direct effect (not mediated through covariates) is modelled (Yang et al., 2014).

Hernan et al. (2000) used weighted pooled logistic regression on discrete time data to approximate the weighted Cox model, mainly due to software constraints that did not allow for use of time-varying weights with the Cox model. Fitting MSM in this way is popular however, fitting the marginal structural Cox model in this way may lead to biased treatment effect when outcome is frequent as shown in a simulation study (Xiao et al., 2010). It has been shown that marginal structural Cox model should be fit directly using the Cox model, which is possible in R software (Xiao et al., 2010, van der Wal and Geskus, 2011).

Robust standard errors of the treatment effect estimates should be obtained, in order to account for repeated measurements within a patient introduced from the weights (Hernan et al., 2000).

5.7 Summary

In summary, this chapter gave an overview of statistical methods that could be used to control for confounding. Baseline and time-varying PS subclassification and MSM are the methods of choice that will be applied to the estimation of effect of allopurinol on outcome using data from CPRD. The next chapter describes how these models were applied to CPRD data described in Chapter 4.

6 Statistical analysis plan

This chapter describes the specifics of how PS methods outlined in Chapter 5 were applied to CPRD data described in Chapter 4. Four statistical approaches to control for confounding were considered: time-invariant (or baseline) propensity score (PS) subclassification (Section 6.2); time-varying PS subclassification (Section 6.3); marginal structural models (MSM) assuming simple associations between treatment and covariates (Section 6.4); MSM assuming complex associations between treatment and covariates (Section 6.5).

For each method, the main analysis was to evaluate the effectiveness of allopurinol on a range of gout outcomes. Various sensitivity analyses were conducted within each method to assess robustness of estimated treatment effects.

Time-invariant PS subclassification

1) To evaluate whether the effectiveness of allopurinol differs between patients with and without renal disease.
2) To evaluate whether the effectiveness of allopurinol differs between patients with severe (>480μmol/L) and non-severe hyperuricaemia.
3) Assess the impact of missing data, unmeasured confounding and landmark period on treatment effect estimation.

Time-varying PS subclassification

1) Assess the impact of missing data on treatment effect estimation.

Simple mechanisms of allopurinol use via MSM

1) Assess the impact of normalised weights, weight truncation, truncating follow-up, and performing intention-to-treat analysis, on treatment effect estimation.

Complex mechanisms of allopurinol use via MSM

1) Assess the impact of different PS models, weight truncation, missing data, and unmeasured confounding on treatment effect estimates.

6.1 Missing data

As seen in Chapter 4, missing data was present for body mass index (BMI), alcohol consumption, smoking status, and serum urate (SU) level. When these covariates were assumed time-invariant, the missing indicator method (MIM) was utilised which sets missing values to a fixed value (indicating missingness) creating an extra dummy variable.

When BMI, alcohol consumption, and smoking status were considered time-varying, first, the MIM approach was applied from start of follow-up (if missing) until these covariates were measured. Secondly, the last observation carried forward (LOCF) approach was used. This approach was adopted for primary analysis in favour of complete case analysis to preserve sample size. Use of the MIM within PS based methods ensures that the distribution of the missingness category is balanced across the treatment groups (Rosenbaum and Rubin, 1984).

Use of these methods (MIM and LOCF) were used first to understand the data and any issues that may arise from modelling that data.

6.2 Time-invariant PS subclassification

6.2.1 Description of study sample

The primary analysis utilised a one year-landmark period (Section 4.3.1). The number of patients consulting for gout and meeting the inclusion and exclusion criteria were described in a flow diagram. Description of baseline covariates, overall and stratified by treatment status

were provided. Assessment of covariate balance was determined using standardised mean difference (SMD) (Section 5.4.4).

6.2.2 Propensity score model

Logistic regression was used to estimate the propensity of treatment at baseline. The general consensus is to include all covariates into the logistic model if they are associated with outcome (Austin et al., 2007a, Cuong, 2013, Rubin and Thomas, 1996) (Section 5.4.2). Covariates were chosen based on statistical significance or clinical justification. To determine statistical significance, univariable Cox proportional hazards (PH) models were used to determine the association between each covariate and time to outcome. Statistical significance was achieved when p-value <0.05. Covariates that have previously been shown to be associated with outcome were also included in PS estimation, regardless of statistical significance: age, sex, deprivation, renal disease, colchicine, non-steroidal anti-inflammatory drugs (NSAIDS), diuretics, and SU level.

The initial PS model included main effects of covariates and assumed continuous covariates had a linear trend with the log odds of the PS. The distribution of PS was visually compared between the treatment groups to assess overlap in the tails of PS distribution. If there was considerable non-overlap, these patients were removed prior to PS subclassification.

Patients were then stratified into five mutually exclusive subclasses based on the quintiles of the PS. Overall SMD for each covariate across the subclasses was calculated. If overall SMD >0.10 in any of the covariates indicating imbalance between treatment groups, two approaches were considered to improve balance: (1) increasing the number of subclasses (Section 5.4.3), (2) re-estimating the PS including interactions between imbalanced covariates and/or non-linear terms for continuous imbalanced covariates. A cycle may occur where

overall SMD are continually assessed for each additional subclass or change to the PS model. Once overall SMD <0.10 across subclasses, covariate balance was considered to be achieved overall however, imbalance may remain within subclasses. Therefore, SMD was evaluated for all covariates within a subclass and any covariates with SMD >0.10 were adjusted for in the outcome analysis.

6.2.3 Estimating treatment effect

Cox PH regression was used to estimate the effect of allopurinol on time to first outcome. For repeated gout consultations, the Anderson & Gill method was used (Section 5.4.5).

The Cox PH model was fitted within each subclass to obtain subclass-specific hazard ratios (HR) and the associated robust standard errors, from which 95% confidence intervals (CI) were obtained. The HRs from each subclass were pooled together using the Mantel-Haenszel method if subclass-specific HRs were deemed to be homogenous, i.e., not statistically significantly different from each other (p-value <0.05), otherwise HRs were weighted by the inverse of the subclass-specific sample size if they are non-constant.

For outcomes where subclass-specific HRs were not homogenous, the estimated pooled HR, alongside with subclass-specific HRs and description of subclasses were presented for each subclass.

Departure from non-proportionality of hazards was assessed within each subclass (Section 5.4.5.2).

6.2.4 Sensitivity and stratified analyses

6.2.4.1 Stratification by renal disease

The one-year landmark analyses were stratified by presence of renal disease at baseline for all outcomes. Patients who did not have renal disease at baseline and went on to develop renal disease during the landmark period were removed from analysis. Patients who developed renal disease during follow-up had their follow-up censored at that date. The same set of covariates (with the exception of renal disease) was used to estimate PS as in the primary analysis.

6.2.4.2 Stratification by severity of hyperuricaemia

The one-year landmark analyses were stratified by severity of hyperuricaemia at baseline for all outcomes. Patients had severe hyperuricaemia if SU level was above 480µmol/L. Patients had non-severe hyperuricaemia is SU level was between 361-480µmol/L. Patients with SU <360µmol/L or no baseline information were removed from analysis. For the analysis of secondary outcomes, as SU level was infrequently measured over time, patients were not censored when SU level reached target. The same set of covariates (with the exception of SU level) was used to estimate PS as in the primary analysis.

6.2.4.3 Two-year landmark period

The landmark period was extended to two years to evaluate robustness of treatment effect estimates when more patients were classified as allopurinol users (Section 4.3.1). Patients with less than two years of follow-up were removed from analysis and follow-up commenced two years after the index date. Same set of covariates was used to estimate PS as in the primary analysis. Patients were classed as an allopurinol user if they had a three-month prescription in the first or second year of follow-up during the two-year landmark period.

6.2.4.4 Unmeasured confounding

The core assumption in treatment effect estimation is that there is no unmeasured confounding. In practice this is unlikely to be plausible especially with retrospective data collection where one cannot choose which covariates to measure. Therefore, there will be residual confounding in the treatment effect estimate. There are a large number of approaches one could use to evaluate the impact of unmeasured confounding (Streeter et al., 2017, Uddin et al., 2016).

For time-invariant PS subclassification, impact of an unmeasured binary covariate (or a vector of binary covariates) on treatment effect estimation was assessed for outcomes target SU level and mortality. The method by Lin et al. (1998) was used for ease of use and has been previously used in other CPRD studies (Blagojevic-Bucknall et al., 2019).

This is where the association between the unmeasured covariate and outcome was based on the hazard ratio (OHR) between observed covariates and outcome. Prevalence of unmeasured covariate is (P_1) and non-users (P_0).

The estimated adjusted hazard ratio (HR^*) from the one-year landmark analysis was corrected for the unmeasured covariate. The formula used for correction is shown in Equation 6.1.

$$Corrected\ HR = HR^* - ln\frac{OHR\ x\ P_1 + (1 - P_1)}{OHR\ x\ P_0 + (1 - P_0)} \qquad 6.1$$

The standard error from the adjusted hazard ratio was assumed to be the same for the corrected hazard ratio.

6.2.4.5 Missing data

The main analysis was performed using the MIM. Creating a separate category to group all observations with missing data on a particular covariate may result in grouping of dissimilar measurements; this in turn may lead to incomplete control of confounding effects from covariates with missing data. Complete case analysis was performed by restricting analysis to patients who had complete data on the lifestyle factors (BMI, smoking status, alcohol consumption, and SU level (for secondary outcomes)) excepting a much smaller sample size and reduced power although, it does not address the robustness of treatment effect estimates against missing data. This sensitivity analysis was performed for the primary outcome target SU level and mortality.

6.3 Time-varying PS subclassification

6.3.1 Descriptive statistics

Using the repeated measures structure described in Section 4.3.2, covariates were described over time by providing descriptive statistics at baseline and in each year of follow-up.

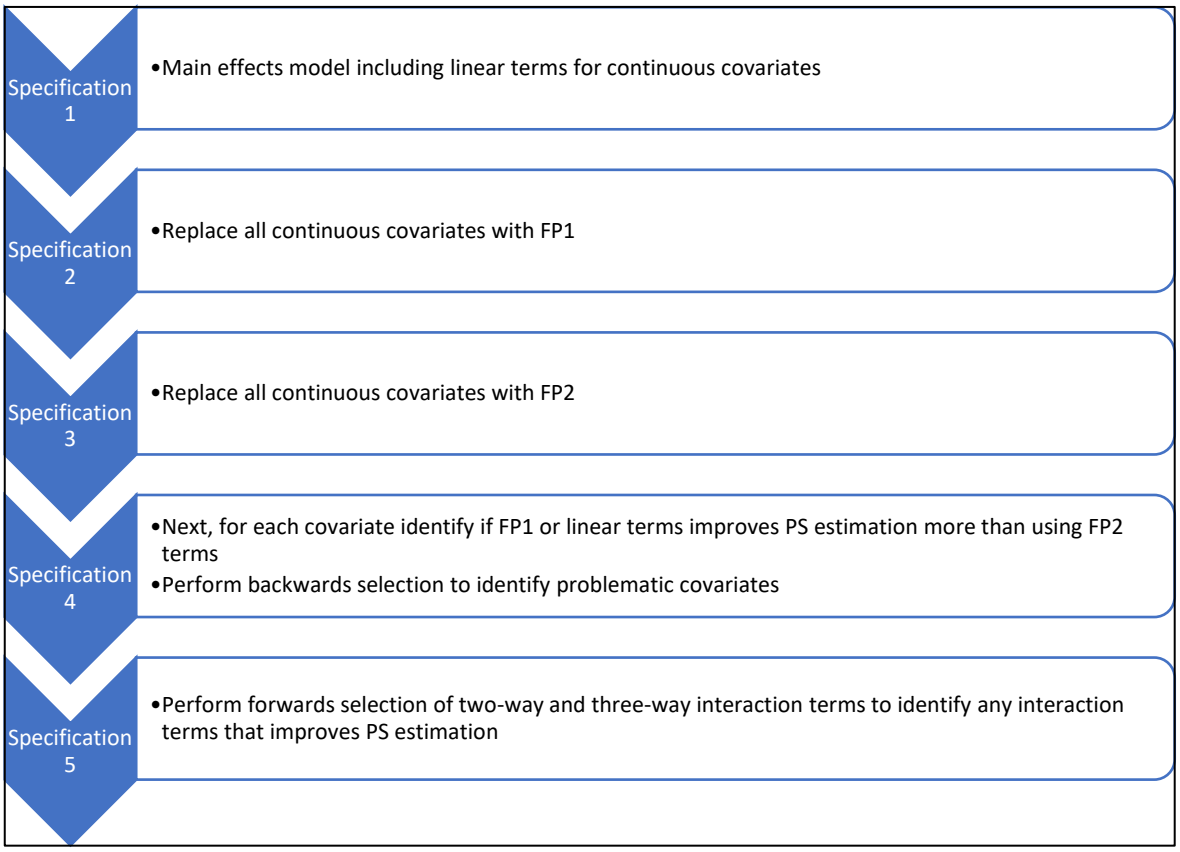
Using intervals as the unit of analysis, the distribution of covariates was described stratified by treatment, and SMD to assess covariate balance was calculated. The following characteristics of treatment patterns were described: the proportion of patients prescribed allopurinol at each time point, the number of times patients initiated and discontinued treatment; year of follow-up in which initiating and discontinuation of treatment occurred in; number of consecutive intervals in which patients were treated.

6.3.2 Propensity score model

The PS for each patient in each interval was estimated using mixed-effects logistic regression model with a random intercept (Section 5.5.1). Similarly to time-invariant PS subclassification, covariates were included in the PS model if they are associated with outcome based on statistical significance or clinical justification. The initial PS model contained main effects of all covariates and linear terms for continuous covariates. To determine statistical significance, univariable complementary log-log regression model were used to determine the association between each covariate and outcome. Statistical significance was achieved when p-value <0.05. Covariates that have previously been shown to be associated with outcome were also included in PS estimation regardless of statistical significance: age, sex, deprivation, renal disease, colchicine, NSAIDS, diuretics, SU level, gout consultation, cumulative allopurinol use, and follow-up time were included in the model regardless of significance.

During PS estimation, lack of common support was identified as an issue for all outcomes. The process followed to maximise common support for each outcome is summarised in Figure 6.1. With each modification of the PS model, common support was evaluated. Common support was assessed by calculating the total of (1) the number of allopurinol intervals for which the PS is above the maximum value of PS from the non-allopurinol intervals and (2) the number of non-allopurinol intervals for which the PS is below the minimum value of PS from the allopurinol intervals.

Figure 6.1: Process of selecting the best PS model



FP1: Fractional polynomials of dimension 1; Fractional polynomials of dimension 2

PS model specification 1 was the initial PS model as described above.

Non-linear transformations of continuous covariates (age, deprivation, SU level (if applicable), cumulative allopurinol use, and follow-up time) via fractional polynomials (FP) were next considered. The set of power transformations considered was described in Section 5.6.3. In PS model specification 2, all the continuous covariates were replaced with FP1 terms (fractional polynomial terms of dimension 1). In PS model specification 3, all the continuous covariates were replaced with FP2 terms (fractional polynomial terms of dimension 2). The FP1 and FP2 terms that best modelled the association between a continuous covariate and outcome was identified and included in the PS model. For each covariate, the best FP1 term was identified by fitting a complementary log-log regression model between each FP1 term and outcome. The model that yielded the lowest deviance indicated the best FP1 term; deviance was defined as twice the negative log-likelihood. Next, the same process was used

to identify the best FP2 terms for each covariate. Restricted cubic splines were initially considered however due to high collinearity they were not used.

In specification 4, for each covariate, it was identified whether its main effect, FP1 term or FP2 terms improved common support the most. Starting with PS model specification 3, three PS models were fitted where age was entered into the PS model as FP2 terms, then FP1 term, then as a linear term. The PS model that improved common support the most from model specification 3 was retained. This process was then repeated for the remaining continuous covariates until common support no longer improved. Next, backwards selection was performed removing one covariate from the model at a time to identify if there were problematic covariates that caused lack of common support.

In specification 5, there is the possibility that two-way interaction terms between covariates may improve common support. Interaction terms were considered for covariates included in the PS model except for non-linear continuous covariates. Interaction terms were only included in the PS model if it was associated with outcome. This was determined by fitting a complementary log-log model between outcome and any two covariates. The incremental contribution of the interaction term between the two covariates is then tested using the likelihood ratio test (LRT) by comparing the -2 difference in log-likelihood of the GLM models with and without the interaction. An insignificant test would indicate that that the interaction term is not associated with outcome thus cannot be included in the PS model. A significant test, p-value <0.05, indicates that the interaction term is associated with outcome thus may be considered to be included in the PS model. Once the relevant interaction terms were identified, each interaction term was fitted to the PS model specification 3, and common support was assessed. The interaction term that reduced lack of support the most was retained. The process was then repeated finding the next interaction term that reduced the

lack of common support the most, and so on. The same process was then repeated for three-way interaction terms.

Once this process was completed, the final PS model was estimated and the distribution of estimated PS was compared between the treatment groups.

Observations in each interval were initially stratified into quintiles based on the distribution of the PS. Subsequently, overall SMD for each covariate was calculated across the subclasses; if overall SMD was >0.10 for any covariate then the number of subclasses was increased until overall SMD was achieved. Once the number of subclasses was selected such that overall SMD is <0.10 for all covariates, SMD for each covariate was then evaluated within each subclass; covariates that had SMD >0.10 were adjusted for in subclass-specific treatment effect estimation.

6.3.3 Estimating treatment effect

The treatment effect was evaluated separately in each subclass by fitting a complementary log-log regression model (Section 5.5.5), regressing outcome on allopurinol and follow-up time; covariates where imbalance remained between treatment groups were adjusted for as well. Treatment effects were presented as HRs and the associated robust standard errors to account for clustering of patients within a subclass, from which 95% CI were obtained.

Subclass-specific hazard ratios were pooled together using the MH method. If the assumption of homogenous treatment effect across subclasses could not be verified, the subclass-class specific HRs were presented alongside with a summary of covariates of that subclass.

In case inclusion of random intercept in the complementary log-log model due to non-convergence of the model, robust standard errors were used to account for repeated measurements within patients.

6.3.4 Sensitivity analysis: missing data

Sensitivity analyses to missing data was performed. Similarly to baseline PS subclassification, treatment effect estimates from using the MIM and LOCF were compared to treatment effect estimates from complete case analysis.. Complete case analysis was restricted to patients who had complete data on the lifestyle factors (BMI, smoking status, alcohol consumption, and SU level (for secondary outcomes) in all follow-up intervals. This sensitivity analysis was performed for the primary outcome target SU level and mortality.

6.4 Simple mechanisms of allopurinol use via MSM

Initial weight estimation assumed the associations between covariates and treatment initiation and continuation were the same. However, the estimated weights were extreme and skewed therefore, this analysis was restricted to the outcome mortality to demonstrate the difficulty in estimating weights with a reasonable distribution and its impact on treatment effect estimation.

6.4.1 Propensity score model

The pooled logistic regression model was used to estimate PS conditional on treatment history and covariates measured in the previous year, assuming it adequately captures covariate history; therefore, lagged covariates were not considered. Estimated PS contributed to the denominator in the estimation of stabilised weights. Covariates that were statistically significantly associated with outcome (previously assessed using the univariable complementary log-log model) and associated with treatment (assessed using the univariable random intercept logistic model), were included in the PS model if p-value <0.05.

The process of selecting a PS model was described in Section 6.3.2, Figure 6.1. Inclusion of non-linear terms for continuous covariates via FP and interaction terms were included in the PS model if it reduced the standard deviation (SD) of weights. Any problematic covariates that cause large weight variability were removed from the PS model.

For the numerator of the stabilised weight, the PS was conditional on baseline covariates (sex, deprivation, baseline SU level) and treatment history i.e., cumulative allopurinol use.

Covariate balance was assessed for each covariate in each year of follow-up using SMD in the weighted study sample; intervals were then pooled together to obtain an overall SMD.

6.4.2 Estimating treatment effect

The treatment effect on mortality was evaluated using the weighted Cox PH regression model (Section 5.6.5). Due to the use of stabilised weights, the Cox model adjusted for baseline covariates (SU level, sex and deprivation) and cumulative allopurinol use. HRs with robust standard errors, for possible misspecification of the PS model and the Cox model, with 95% CI were presented.

6.4.3 Sensitivity analyses

In addition to the process of finding a suitable PS model, other methods were used to reduce extreme weights as described below.

6.4.3.1 Normalised weights

The stabilised weights were normalised such that the mean weight was 1 at each time point. Covariate imbalance and treatment effect were re-evaluated weighting the study sample using normalised weights.

6.4.3.2 Truncating weights

Weight truncation was performed on both the stabilised and normalised weights to remove extreme weights that may be influential on the estimated treatment effect. The weights were truncated at between 1% and 10% percentiles of its distribution. SMD and treatment effect were re-evaluated based on the truncated weights.

6.4.3.3 Truncated follow-up

Due to the long follow-up period that is observed in using EHR data, potentially differences in covariate distribution between treatment groups may increase over time (as assessed using SMD). Extreme weights may also be present due to cumulatively multiplying large probabilities of treatment over time. Follow-up was therefore truncated at the point prior to severe covariate imbalance when SMD >0.25. Treatment effect and SMD were re-evaluated.

6.4.3.4 Intention-to-treat analysis

So far, analyses considered treatment as an intermittent time-varying treatment with weights estimated in each year of follow-up, thus estimating the effect of actual treatment. Estimated treatment effect was compared with estimates derived under the intention-to-treat principle to assess whether the issues encountered in estimating actual treatment effect persisted when estimating the effect of initiating treatment.

6.5 Complex MSM

In this analysis, associations between covariates and treatment initiation and continuation were allowed to differ in PS estimation.

6.5.1 Revisiting the repeated measures dataset and descriptive statistics

Previously in Section 6.4, only one PS model (for the denominator of weight) was used to estimate the probability of treatment. Here, two PS models were fitted, one that estimated the probability of initiating treatment, and the second estimated the probability of continuing with treatment. Within the repeated measures dataset, to identify which intervals correspond to covariates that were associated with initiating and continuing treatment, intervals were stratified by allopurinol use in the previous interval. This approach had previously been used by Cook et al. (2012), Xiao et al. (2014), and Yang et al. (2015a).

Table 6.1 illustrates how this stratification would appear for four hypothetical patients who (1) never initiated allopurinol, (2) initiated allopurinol and remained on allopurinol until the end of follow-up, (3) initiated allopurinol and then discontinued allopurinol, (4) initiated and then discontinued allopurinol, and subsequently restarted and stopped allopurinol.

Table 6.1: Example: initiation and continuation of allopurinol use

Patient identifier	Follow-up time (years)	Allopurinol use	Allopurinol use in the previous year	Initiates allopurinol	Continues allopurinol use
1	1	0	0	0	N/A
1	2	0	0	0	N/A
1	3	0	0	0	N/A
2	1	0	0	0	N/A
2	2	1	0	1	N/A
2	3	1	1	N/A	1
2	4	1	1	N/A	1
3	1	0	0	0	N/A
3	2	0	0	0	N/A
3	3	1	0	1	N/A
3	4	1	1	N/A	1
3	5	1	1	N/A	1
3	6	0	1	N/A	0
3	7	0	0	0	N/A
4	1	0	0	0	N/A
4	2	1	0	1	N/A
4	3	1	1	N/A	1
4	4	1	1	N/A	1
4	5	1	1	N/A	1
4	6	0	1	N/A	0
4	7	0	0	0	N/A
4	8	1	0	1	N/A
4	9	1	1	N/A	1
4	10	0	1	N/A	0
4	11	1	0	1	N/A
4	12	0	1	N/A	0

0: No; 1: Yes; N/A: Not applicable

The example dataset contains patient identifier (first column), year of follow-up (second column), prescription of allopurinol (third column), and allopurinol use in the previous year (fourth column). The fifth column identifies when non-users initiated treatment. The sixth column identifies when on-going allopurinol users discontinued treatment.

Patient 1 is followed-up for three years and is never prescribed allopurinol, hence only contributes to column 5. Patient 2 is followed up for four years, initiates treatment in year 2 and remains on allopurinol in years 3 and 4; years 1 and 2 contribute to column 5 and years 3 and 4 contribute to column 6. Patient 3 is followed-up for 7 years, initiates allopurinol in year 3 and discontinues allopurinol in year 6 for the remainder of the follow-up; years 1-3 and 7 contribute to column 5 and years 4-6 contribute to column 6. Patient identifier 4 is followed-

up for 12 years, initiate allopurinol in years 2, 8 and 11 and discontinue allopurinol in years 6, 10 and 12.

Covariates were described at the time of allopurinol initiation compared with non-use, and allopurinol continuation compared with discontinuation alongside with evaluating SMD of covariates between treatment groups. Unadjusted associations of covariates with allopurinol initiation and continuation were estimated using the random intercept logistic model.

6.5.2 Propensity score model

The probability of observing treatment given covariates (denominator of weights) was estimated separately for those initiating allopurinol and those continuing with allopurinol.

The probability of initiating allopurinol was estimated in non-users up to and including the year they initiated treatment or until the end of follow-up if they were never treated. Some patients initiated treatment multiple times; for simplicity, it was assumed the differences in covariates between non-users and those initiating allopurinol for the first time were similar to the differences between non-users who may previously have had treatment and those restarting allopurinol accepting that some residual confounding may remain; otherwise separate PS models would have to be fitted for these scenarios too.

The probability of continuing with allopurinol (those prescribed allopurinol in the previous year), was estimated in allopurinol users up to and including the year they discontinued treatment or until the end of follow-up if they remained on treatment. Similarly, it was assumed differences in covariates between those continuing allopurinol and those discontinuing allopurinol for the first time were similar to the differences between patients restarting and remaining on allopurinol and discontinuing treatment again accepting that residual confounding may remain.

Once PS were estimated, stabilised weights were initially estimated across the whole study sample; if the distribution of weights was not skewed, mean weight did not deviate from one, and did not have large weights, basic weights were instead estimated.

The process of selecting a PS model was described in Section 6.3.2, Figure 6.1. Inclusion of non-linear terms for continuous covariates via FP and interaction terms were included in the PS model if it reduced the SD of weights. Any problematic covariates that caused large weight variability were removed from the PS model. The same covariates were included in both PS models in estimating probability of initiating and continuing treatment.

Once PS were estimated weights were derived, SMD for each covariate was evaluated in the weighted study sample at each time point and then overall.

6.5.3 Estimating treatment effect

Treatment effect was estimated using the weighted Cox regression model; HR with 95% CI based on robust standard errors were presented.

Covariate imbalance was assessed between treatment groups in each year of follow-up, and then overall. If overall SMD was greater than 0.10, that covariate was adjusted for in the Cox model.

6.5.4 Sensitivity analyses

In addition to the process of finding a suitable PS model, other methods were used to reduce extreme weights as described below.

6.5.4.1 Weight truncation

Weights were truncated by 0.5% centile of its distribution. Larger truncation e.g., 1% was not considered as 0.5% truncation was sufficient to remove extreme weights. Treatment effects were re-estimated using the truncated weights.

6.5.4.2 Unmeasured confounding

More recently, a different method in measuring the impact of unmeasured confounding on treatment effect estimation was introduced, known as E-values (Mathur et al., 2018, VanderWeele and Ding, 2017). E-values are attractive to use as they can be applied to various situations where treatment estimate can be the HR, odds ratio, rate ratio etc. derived from various GLM, and software to estimate E-values is widely available.

Therefore, E-values were computed for complex MSM. E-values are measured on the risk ratio (RR) scale to assess how strong the association between an unmeasured confounding variable with treatment and outcome needs to be in order to explain away the treatment effect conditional on the observed covariates; therefore, an E-value of 1 indicates there are no unmeasured covariates. E-values were computed against the statistically significant treatment effects estimated from the main analysis without adjustment for imbalanced covariates. To assess the likelihood of such an unmeasured confounding variable to exist, unadjusted RRs between observed covariates with outcome, allopurinol initiation and allopurinol continuation were estimated using generalised linear models with Poisson distribution and log link. If the E-value was outside the range of observed RRs then there is some evidence unmeasured confounding is unlikely to explain away the treatment effect. Otherwise, if the E-value was within the range of observed RRs this suggests there is some evidence unmeasured confounding may be present and potentially explain away the treatment effect.

For rare outcomes where the prevalence of outcome is <15% at the end of follow-up, when the estimated HR is greater than 1, the formula to estimate E-value is given in Equation 6.2. When the HR is less than 1, the formula is given in Equation 6.3.

$$E\ value = HR + \sqrt{HR(HR-1)} \qquad 6.2$$

$$HR^* = 1/HR$$
$$E\ value = HR^* + \sqrt{HR^*(HR^*-1)} \qquad 6.3$$

For more common outcomes where the prevalence of outcome is between 15-85% at the end of follow-up, the estimated HR is first transformed according to Equation 6.4 before calculating E-values in Equations 6.2 and 6.3.

$$HR^* = \frac{1-0.5^{\sqrt{HR}}}{1-0.5^{\sqrt{\frac{1}{HR}}}} \qquad 6.4$$

6.6 Summary

This chapter described the statistical analysis plan that addressed the clinical objectives of this thesis, and how statistical methods were applied to CPRD data to address confounding by indication.

All analyses were performed in Stata v15 except for the weighted Cox model that was performed in R v3.6.2.

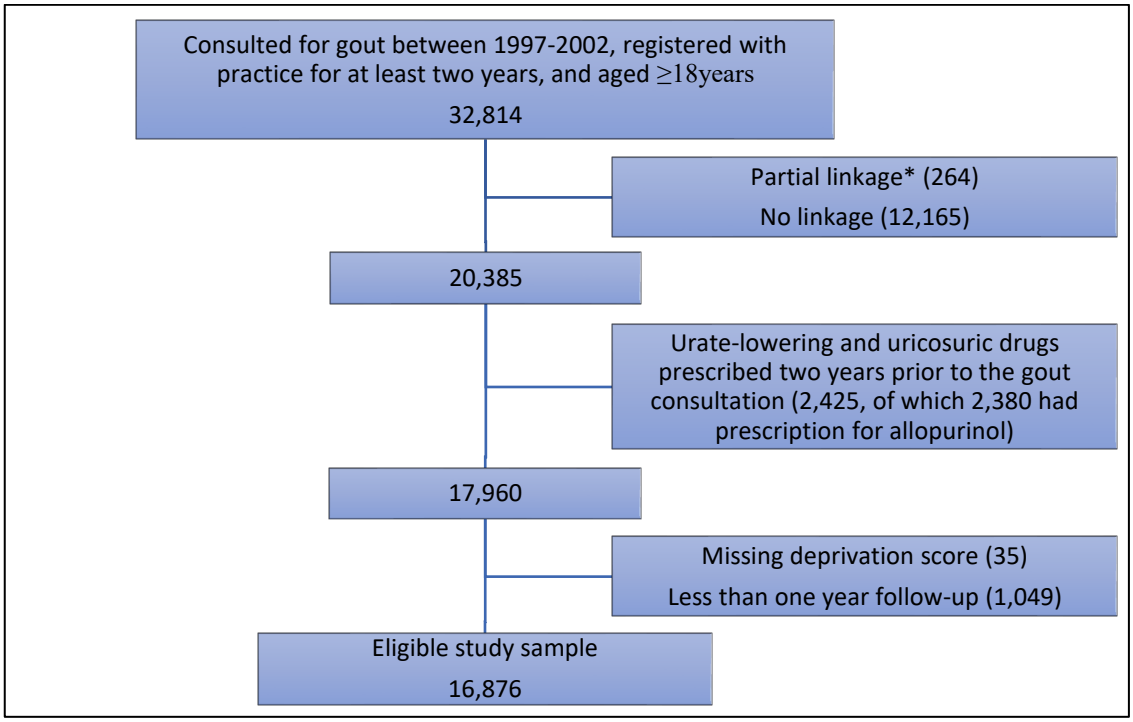
7 Effect of allopurinol: time-invariant PS subclassification

This chapter described the study sample that was obtained from Clinical Practice Research Database (CPRD). Allopurinol use and covariates were considered time-invariant, thus propensity score (PS) was estimated at baseline.

7.1 Study sample

Medical records were extracted from CPRD in January 2015. The extract identified 32,814 patients consulting for gout between the 1st January 1997 and the 31st December 2002, registered for at least two years with their practice, and aged at least 18 years. From this initial cohort 16,876 patients were eligible. Figure 7.1 shows the number of patients excluded and reasons for exclusion.

Figure 7.1: Study sample flow chart

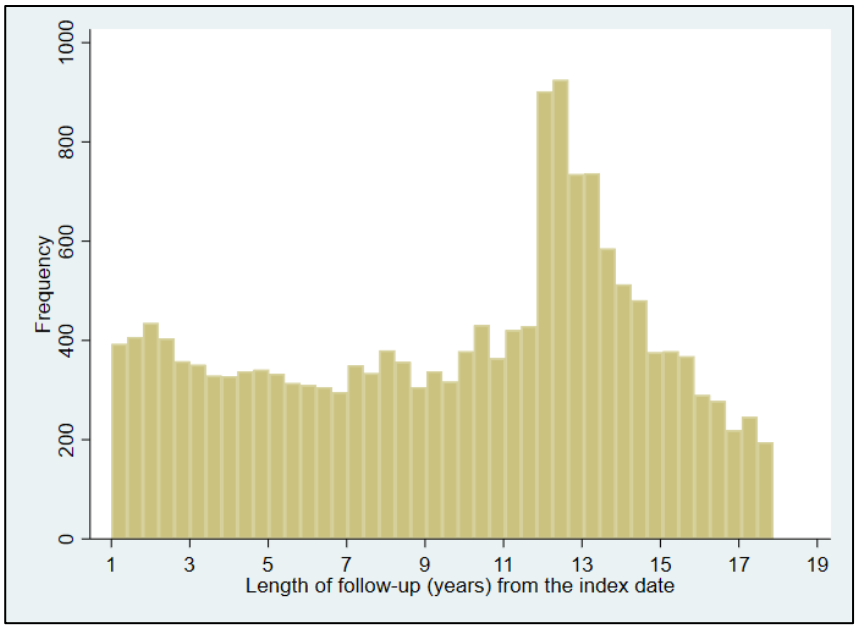


*Had linkage to either one or two secondary databases (Index of Multiple Deprivation, Hospital Episodes Statistics, or the Office of National Statistics mortality data)

Median follow-up was 10.7 years (interquartile range (IQR) 5.7, 13.3 years) contributing 163,607 person-years of follow-up time. The distribution of follow-up duration is given in Figure 7.2. Specific reasons for patient end of follow-up were as follows: last date of practice data collection (48%, n=8,141); death (29%, n=4,976); transferred out of practice (21%, n=3,583); prescribed either sulfinpyrazone, probenecid, or febuxostat (1%, n=176); none were prescribed benzbromarone.

As seen in Figure 7.2, there were a large number of patients (n=3,763) whose follow-up ended between 13 and 16 post index date due to a large number of practices leaving CPRD following a change in software systems from Vision to EMIS.

Figure 7.2: Distribution of follow-up duration



Patients were registered with 275 practices with a median (IQR) 49 (21, 99) patients (with gout) per practice. Over half of the patients were registered with practices located in the North West (16%, n=2,612), South West (12%, n=2,147), and East (13%, n=2,313) of England, and the West Midlands (15%, n=2,652). The least number of patients were registered with practices located in the North East of England (3%, n=571) and the East Midlands (4%, n=777). The baseline covariates for the study sample are described in Table 7.1. The majority were

male (77%) and mean (standard deviation (SD)) age was 62.1 (14.7) years. The three most prevalent comorbidities were hypertension (19%), coronary heart disease (13%), and osteoarthritis (7%). A small proportion of patients (6%) had an acceptable level of serum urate (SU) (360≤µmol/L). Patients were often prescribed pain relief (non-steroidal anti-inflammatory drugs (NSAIDS) (48%) and analgesics (33%)) and diuretics (36%). At least 30% of patients had missing data in lifestyle factors.

Table 7.1: Baseline covariates of study sample (N=16,876)

Demographics	N (%)
Age (Mean (SD), range 18, 101)	62.1 (14.7)
Sex	
Male	12,995 (77)
Female	3,881 (23)
Deprivation (Mean (SD), range 1, 20)	9.1 (5.5)
Comorbidities	
Anxiety	672 (4)
Depression	842 (5)
Cerebrovascular disease	407 (2)
Coronary heart disease	2,167 (13)
Diabetes	1,047 (6)
Hyperlipidaemia	783 (5)
Hypertension	3,137 (19)
Osteoarthritis	1,106 (7)
Peripheral vascular disease	257 (2)
Renal disease	217 (1)
Lifestyle factors	
Alcohol consumption	
Ever drinker	9,488 (56)
Never drinker	856 (5)
Missing	6,532 (39)
Body mass index	
Normal weight	2,517 (15)
Overweight	4,933 (29)
Obese	3,219 (19)
Missing	6,207 (37)
Smoking status	
Ever smoker	6,436 (38)
Never smoker	4,847 (29)
Missing	5,593 (33)
SU level	
≤360µmol/L	951 (6)
>360µmol/L	6,062 (36)
Missing	9,863 (58)
Medication use	
Analgesics	5,578 (33)
Colchicine	389 (2)
Diuretics	6,142 (36)
NSAIDS	8,024 (48)

NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Appendix H presents comparison of baseline covariates of the whole study sample with each subset of that sample which contains patients who were eligible for each outcome analysis; each outcome analysis was restricted to patients who did not have incident outcome during the first year of follow-up. Generally, covariates were similar across the subsets with a few exceptions. Patients eligible for the analysis of target SU level consisted of 10% (N=1,742) of the study sample; this subset of patients was younger and were prescribed analgesics and NSAIDS more than the whole study sample.

The median time to reaching target SU level was 3.5 years. Median time to occurrence of outcome is listed in Table 7.2.

Table 7.2: Median time (years) until occurrence of first outcome

Outcome	Median (Interquartile range)
Target SU level	3.5 (1.4, 7.1)
Mortality	5.1 (2.3, 8.4)
Gout consultation	1.9 (0.7, 4.1)
Gout hospitalisation	6.8 (3.4, 9.7)
Joint replacement	5.3 (2.6, 8.2)
Cerebrovascular disease	4.5 (2.0, 8.2)
Coronary heart disease	3.3 (1.4, 6.5)
Peripheral vascular disease	4.3 (1.8, 7.7)
Renal disease	5.5 (3.7, 7.6)

SU: Serum urate

7.2 Patient characteristics by allopurinol treatment

Large differences between the treatment groups were observed where standardised mean difference (SMD) was >0.10 (Table 7.3). Allopurinol users had higher prevalence of coronary heart disease (SMD=0.11), renal disease (0.12), and were prescribed analgesics (0.19), colchicine (0.13), diuretics (0.25), and NSAIDS (0.30) more than non-users. The prevalence of missing data between treatment groups were similar for alcohol consumption, body mass index (BMI), and smoking status.

The largest difference between treatment groups was observed for baseline SU level. Allopurinol users had a higher prevalence of having a baseline SU level above target (>360µmol/L) (0.39) than non-users however, had lower prevalence of missing data (-0.27).

Table 7.3: Baseline covariates by allopurinol treatment: whole study sample (N=16,876)

	No allopurinol N=12,919	Allopurinol N=3,957	SMD
Demographics			
Age (Mean (SD))	61.8 (14.8)	62.9 (14.5)	0.08
Sex: Female	2,919 (23)	962 (24)	0.04
Deprivation (Mean (SD))	9.0 (5.5)	9.6 (5.7)	0.10
Comorbidities			
Anxiety	529 (4)	143 (4)	-0.03
Depression	654 (5)	188 (5)	-0.01
Cerebrovascular disease	294 (2)	113 (3)	0.04
Coronary heart disease	1,544 (12)	623 (16)	0.11
Diabetes	754 (6)	293 (7)	0.06
Hyperlipidaemia	599 (5)	184 (5)	<0.01
Hypertension	2,362 (18)	775 (20)	0.03
Osteoarthritis	772 (6)	334 (8)	0.10
Peripheral vascular disease	179 (1)	78 (2)	0.05
Renal disease	121 (1)	96 (2)	0.12
Lifestyle factors			
Alcohol consumption			
Ever drinker	7,328 (57)	2,160 (55)	-0.04
Never drinker	663 (5)	193 (5)	-0.01
Missing	4,928 (38)	1,604 (41)	0.05
Body mass index			
Normal	2,021 (16)	496 (13)	-0.09
Overweight	3,829 (30)	1,104 (28)	-0.04
Obese	2,358 (18)	861 (22)	0.09
Missing	4,711 (36)	1,496 (38)	0.03
Smoking status			
Ever smoker	4,945 (38)	1,491 (38)	-0.01
Never smoker	3,776 (29)	1,071 (27)	-0.05
Missing	4,198 (32)	1,395 (35)	0.06
SU level			
≤360µmol/L	897 (7)	54 (1)	-0.28
>360µmol/L	4,064 (31)	1,998 (50)	0.39
Missing	7,958 (62)	1,905 (48)	-0.27
Medication use			
Analgesics	3,992 (31)	1,586 (40)	0.19
Colchicine	232 (2)	157 (4)	0.13
Diuretics	4,339 (34)	1,803 (46)	0.25
NSAIDS	5,689 (44)	2,335 (59)	0.30

N (%) were presented unless otherwise stated; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SMD: Standardised mean difference; SU: Serum urate

7.3 Propensity score model

As previously stated in Section 6.2.2, choice of covariates to be included in the PS model was based on clinical justification and a statistically significant association between outcome and covariate. Covariates used to estimate PS for each outcome analysis are shown in Table 7.4.

Age, sex, deprivation, renal disease, colchicine use, NSAIDS, diuretics, and baseline SU level were included in the PS model regardless of statistical significance a priori.

Anxiety was not associated with any of the outcomes and therefore was not included in estimation of PS. Depression was only associated with repeated gout consultations. The majority of demographics, comorbidities, lifestyle factors and medication usage were associated with most outcomes. The least number of covariates were associated with target SU level which may be due to a lower sample size than analysis of other outcomes. Alcohol consumption was not associated with gout hospitalisation or joint replacement but was included in the PS model regardless due to its effect on SU level (Section 2.3.3).

Table 7.4: Covariates entered into the PS model for each outcome

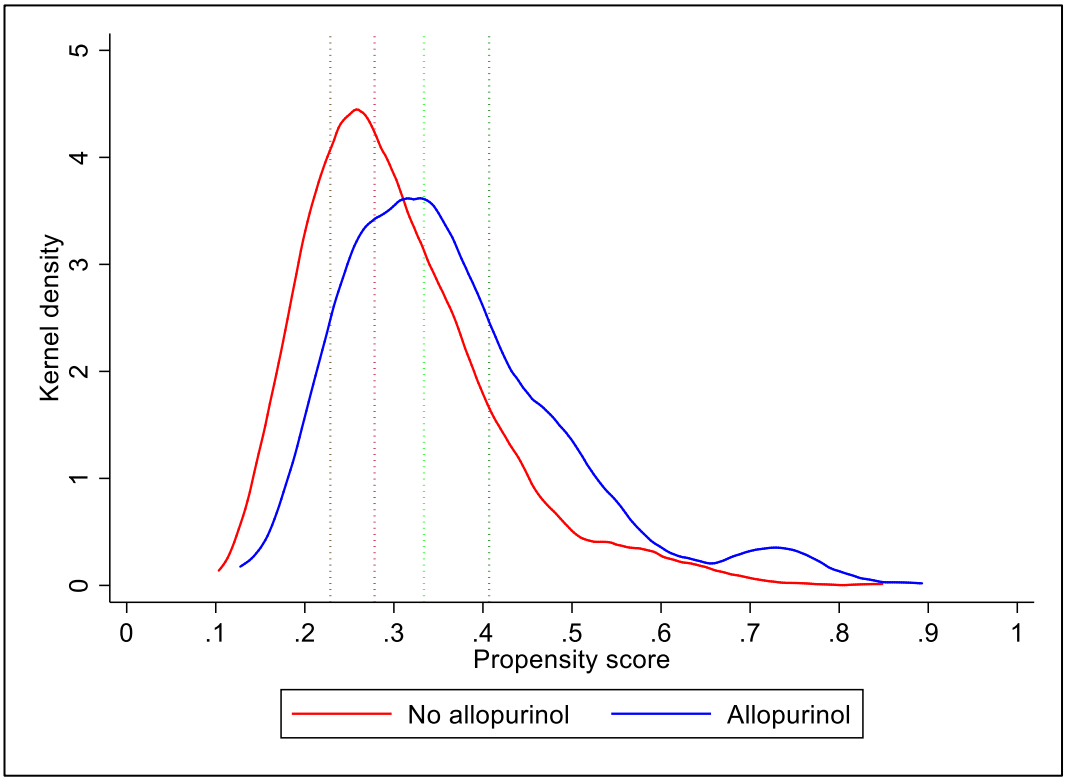
Baseline covariate	Target SU level N=1,742	Mortality N=16,876	Repeated gout consultations N=16,876	Gout hospitalisation N=14,087	Joint replacement N=16,644	Cerebrovascular disease N=16,253	Coronary heart disease N= 14,063	Peripheral vascular disease N=16,519	Renal disease N=16,508
Demographics									
Age	X	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X	X
Deprivation	X	X	X	X	X	X	X	X	X
Comorbidities									
Anxiety									
Depression			X						
Cerebrovascular disease		X	X	X	X		X	X	X
Coronary heart disease	X	X	X	X	X	X		X	X
Diabetes		X	X	X		X	X	X	X
Hyperlipidaemia	X		X			X	X	X	X
Hypertension		X		X	X	X	X	X	X
Osteoarthritis		X		X	X	X	X	X	X
Peripheral vascular disease		X	X			X	X		X
Renal disease	X	X	X	X	X	X	X	X	
Lifestyle factors									
Alcohol consumption	X	X	X	X	X	X	X	X	X
Body mass index	X	X	X	X	X	X	X	X	X
Smoking status	X	X	X	X		X	X	X	X
SU level	X	X	X	X	X	X	X	X	X
Medication use									
Analgesics	X	X	X	X	X	X	X	X	X
Colchicine	X	X	X	X	X	X	X	X	X
Diuretics	X	X	X	X	X	X	X	X	X
NSAIDS	X	X	X	X	X	X	X	X	X

X: Covariate entered into the PS model; Green cell: Covariate was associated with outcome ($p<0.05$); Red cell: Covariate was not associated with outcome ($p\geq0.05$); Black cell: Not applicable; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

7.4 Propensity score distribution

Across all outcome analyses, there was considerable overlap of PS between the treatment groups. For example, in the analysis of target SU level, PS for non-users ranged from 0.08 to 0.80 and for allopurinol users 0.14 to 0.88; the distribution of PS is illustrated in Figure 7.3. Four patients were outside the region of common support therefore the non-overlap of PS between treatment groups was considered minimal and these patients were not removed. Distribution of PS by treatment for the secondary outcomes is given in (Appendix I).

Figure 7.3: Distribution of PS by treatment status in the analysis of target SU level



Dotted lines indicate at which value of the PS subclasses were created; PS: Propensity score

Within all outcome analyses, five PS subclasses were created and within each subclass there were sufficient number of patients in both treatment groups. Furthermore, within each treatment group, in each subclass, outcome had occurred (Table 7.5).

Within all outcome analyses, although five subclasses were deemed sufficient to achieve overall complete covariate balance between treatment groups across subclasses with SMD <0.10 (Table 7.6), imbalance in some covariates remained within subclasses thus these were

subsequently adjusted for in treatment effect estimation. For example, in the analysis of target SU level, imbalance remained for the following covariates: age (subclass 1, 2, 3), deprivation (subclass 4, 5), coronary heart disease (subclass 1, 3, 4, 5), hyperlipidaemia (subclass 1, 2), renal disease (subclass 2, 3, 4, 5), alcohol consumption (subclass 3, 4), smoking status (subclass 1, 5), BMI (subclass 2, 3, 4), SU level (subclass 1, 2, 5), analgesics (subclass 3), colchicine (subclass 1, 2, 3, 4, 5), diuretics (subclass 3, 4, 5).

For the secondary outcome analyses, subclass 1 had the most number of imbalanced covariates (up to a maximum of five covariates including sex, alcohol consumption, BMI, NSAIDS, and baseline SU level) whereas balance was achieved on all covariates in subclasses 4 and 5 (Appendix I).

Increasing the number of subclasses or including interaction terms between imbalanced covariates did not improve balance within subclasses (data not shown).

Table 7.5: Distribution of patients across PS subclasses by treatment and outcome

Outcome	No allopurinol N (%)			Allopurinol N (%)		
Target SU level	SU target not met N=789	SU target met N=386	Total N=1,175	SU target not met N=246	SU target met N=321	Total N=567
Subclass 1	193 (24)	99 (26)	292 (25)	23 (9)	34 (11)	57 (10)
Subclass 2	181 (23)	80 (21)	261 (22)	31 (13)	56 (17)	87 (15)
Subclass 3	165 (21)	74 (19)	239 (20)	56 (23)	54 (17)	110 (19)
Subclass 4	130 (16)	75 (19)	205 (17)	61 (25)	82 (26)	143 (25)
Subclass 5	120 (15)	58 (15)	178 (15)	75 (30)	95 (30)	170 (30)
Mortality	Alive N=9,260	Died N=3,659	Total N=12,919	Alive N=2,640	Died N=1,317	Total N=3,957
Subclass 1	2,413 (26)	654 (18)	3,067 (24)	236 (9)	73 (6)	309 (8)
Subclass 2	1,951 (21)	842 (23)	2,793 (22)	404 (15)	178 (14)	582 (15)
Subclass 3	2,012 (22)	643 (18)	2,655 (21)	540 (20)	180 (14)	720 (18)
Subclass 4	1,528 (17)	833 (23)	2,361 (18)	642 (24)	372 (28)	1,014 (26)
Subclass 5	1,356 (15)	687 (19)	2,043 (16)	818 (31)	514 (39)	1,332 (34)
Repeated gout consultations	Never consulted N=5,787	Consulted ≥1 times N=7,132	Total N=12,919	Never consulted N=2,193	Consulted ≥1 times N=1,764	Total N=3,957
Subclass 1	1,744 (30)	1,319 (18)	3,063 (24)	173 (8)	140 (8)	140 (8)
Subclass 2	1,247 (22)	1,552 (22)	2,799 (22)	320 (15)	256 (15)	256 (15)
Subclass 3	1,038 (18)	1,610 (23)	2,648 (21)	371 (17)	356 (20)	356 (20)
Subclass 4	968 (17)	1,416 (20)	2,384 (18)	571 (26)	420 (24)	420 (24)
Subclass 5	790 (14)	1,235 (17)	2,025 (16)	758 (35)	592 (34)	592 (34)
Gout Hospitalisation	No N=9,546	Yes N=1,283	Total N=10,829	No N=2,569	Yes N=689	Total N=3,258
Subclass 1	2,370 (25)	198 (15)	2,568 (24)	199 (8)	51 (7)	250 (8)
Subclass 2	2,065 (22)	267 (21)	2,332 (22)	383 (15)	102 (15)	485 (15)
Subclass 3	1,928 (20)	284 (22)	2,212 (20)	484 (19)	122 (18)	606 (19)
Subclass 4	1,726 (18)	269 (20)	1,995 (18)	647 (25)	175 (25)	822 (25)
Subclass 5	1,457 (15)	265 (21)	1,722 (16)	856 (33)	239 (35)	1,095 (34)

PS: Propensity score; SU: Serum urate

Table 7.5 continued:

Outcome	No allopurinol			Allopurinol		
Joint Replacement	No: N=12,008	Yes: N=744	Total: N=12,752	No: N=3,621	Yes: N=271	Total: N=3,892
Subclass 1	2,876 (24)	151 (20)	3,027 (24)	290 (8)	12 (4)	302 (8)
Subclass 2	2,607 (22)	152 (20)	2,759 (22)	534 (15)	36 (13)	570 (15)
Subclass 3	2,462 (21)	150 (20)	2,612 (20)	666 (18)	51 (19)	717 (18)
Subclass 4	2,185 (18)	151 (20)	2,336 (18)	920 (25)	73 (27)	993 (26)
Subclass 5	1,878 (16)	140 (19)	2,018 (16)	1,211 (33)	99 (37)	1,310 (34)
Cerebrovascular disease	No: N=11,273	Yes: N=1,195	Total: N=12,468	No: N=3,428	Yes: N=357	Total: N=3,785
Subclass 1	2,681 (24)	277 (23)	2,958 (24)	268 (8)	25 (7)	293 (8)
Subclass 2	2,413 (21)	278 (23)	2,691 (22)	501 (15)	59 (17)	560 (15)
Subclass 3	2,345 (21)	209 (17)	2,554 (20)	652 (19)	44 (12)	696 (18)
Subclass 4	2,048 (18)	249 (21)	2,297 (18)	861 (25)	93 (26)	954 (25)
Subclass 5	1,786 (16)	182 (15)	1,968 (16)	1,146 (33)	136 (38)	1,282 (34)
Coronary heart disease	No: N=8,675	Yes: N=2,228	Total: N=10,903	No: N=2,438	Yes: N=722	Total: N=3,160
Subclass 1	2,086 (24)	475 (21)	2,561 (23)	196 (8)	56 (8)	252 (8)
Subclass 2	1,864 (21)	501 (22)	2,365 (22)	351 (14)	97 (13)	448 (14)
Subclass 3	1,778 (21)	436 (20)	2,214 (20)	485 (20)	113 (16)	598 (19)
Subclass 4	1,555 (18)	459 (21)	2,014 (18)	604 (25)	195 (27)	799 (25)
Subclass 5	1,392 (16)	357 (16)	1,749 (16)	802 (33)	261 (36)	1,063 (34)
Peripheral vascular disease	No: N=12,186	Yes: N=480	Total: N=12,666	No: N=3,698	Yes: N=155	Total: N=3,853
Subclass 1	2,910 (24)	92 (19)	3,002 (24)	289 (8)	14 (9)	303 (8)
Subclass 2	2,636 (22)	102 (21)	2,738 (22)	548 (15)	17 (11)	565 (15)
Subclass 3	2,494 (20)	92 (19)	2,586 (20)	695 (19)	23 (15)	718 (19)
Subclass 4	2,238 (18)	102 (21)	2,340 (18)	926 (25)	38 (25)	964 (25)
Subclass 5	1,908 (16)	92 (19)	2,000 (16)	1,240 (34)	63 (41)	1,303 (34)
Renal disease	No: N=9,867	Yes: N=2,864	Total: N=12,731	No: N=2,716	Yes: N=1,061	Total: N=3,777
Subclass 1	2,490 (25)	516 (18)	3,006 (24)	232 (9)	64 (6)	296 (8)
Subclass 2	2,151 (22)	595 (21)	2,746 (22)	432 (16)	124 (12)	556 (15)
Subclass 3	2,050 (21)	544 (19)	2,594 (20)	544 (20)	163 (15)	707 (19)
Subclass 4	1,727 (18)	622 (22)	2,349 (18)	660 (24)	293 (28)	953 (25)
Subclass 5	1,449 (15)	587 (21)	2,036 (16)	848 (31)	417 (39)	1,265 (33)

Table 7.6: Overall SMD across PS subclasses

Outcome	Target SU level	Mortality	Repeated gout consultations	Gout hospitalisation	Joint replacement	Cerebrovascular disease	Coronary heart disease	Peripheral vascular disease	Renal disease
Demographics									
Age	0.02	0.01	0.01	0.01	<0.01	0.01	0.02	0.01	0.01
Sex: Female	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01
Deprivation	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	0.01	<0.01	<0.01
Comorbidities									
Anxiety	-	-	-	-	-	-	-	-	-
Depression	-	-	<0.01	-	-	-	-	-	-
Cerebrovascular disease	-	<0.01	<0.01	<0.01	<0.01	-	0.01	<0.01	<0.01
Coronary heart disease	0.02	0.02	0.01	0.01	0.02	0.02	-	0.02	0.02
Diabetes	-	0.01	<0.01	0.01	-	0.01	0.01	0.01	0.01
Hyperlipidaemia	0.01	-	<0.01	-	-	<0.01	<0.01	<0.01	<0.01
Hypertension	-	<0.01	-	<0.01	<0.01	<0.01	0.01	<0.01	<0.01
Osteoarthritis	-	0.01	-	0.01	<0.01	0.01	0.01	0.01	0.01
Peripheral vascular disease	-	0.01	0.01	-	-	0.01	<0.01	-	<0.01
Renal disease	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.03	-
Lifestyle factors									
Alcohol consumption[a]	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Body mass index[a]	0.01	0.01	0.01	0.01	0.01	<0.01	0.01	0.01	0.01
Smoking status[a]	-0.01	<0.01	<0.01	<0.01	-	-0.01	<0.01	<0.01	<0.01
SU level[a]	0.08[b]	-0.09	-0.09	0.02	-0.08	<0.01	-0.08	-0.08	-0.08
Medication use									
Analgesics	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Colchicine	0.03	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04
Diuretics	0.03	0.02	<0.01	0.02	0.02	0.02	0.03	0.02	0.02
NSAIDS	-0.01	0.01	<0.01	0.01	0.01	<0.01	0.01	0.01	0.01

[a]For categorical variables, the largest SMD was presented; [b]SU level was a continuous covariate; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SMD: Standardised mean difference; SU: Serum urate

7.5 Treatment effect analysis

7.5.1 Landmark one-year analysis

For each outcome analysis, subclass-specific treatment effect estimates were pooled and are presented in Table 7.7 alongside with unadjusted treatment effect (hazard ratio (HR) with 95% confidence interval (CI)) and distribution of outcome across treatment groups.

For the majority of outcomes, the subclass-specific treatment effect estimates were homogeneous across subclasses as judged by the likelihood ratio test, which showed that including an interaction term between allopurinol and PS subclasses in addition to their main effects did not improve model fit (p-value >0.05). Allopurinol use was associated with higher chance of reaching target SU level (HR 2.32 (95% CI: 1.97, 2.74)), and increased risk of premature mortality (1.10 (1.03, 1.17)), coronary heart disease (1.11 (1.02, 1.21)) and renal disease (1.19 (1.10, 1.28)) (Table 7.7).

The subclass-specific treatment effect estimates for repeated gout consultations and gout hospitalisation were not homogeneous across subclasses ($p<0.001$ and $p=0.007$ respectively). The pooled estimates were therefore weighted by the inverse of the sample size and presented in Table 7.7, and subclass-specific treatment estimates are presented in Table 7.8 and Table 7.9 respectively alongside with description of covariates.

The pooled treatment effect showed that allopurinol was associated with fewer gout consultations (0.70 (0.65, 0.75)). Within subclass 5 (highest propensity for allopurinol), allopurinol was found to have the strongest protective effect against repeated gout consultations (0.60 (0.53, 0.67)); patients in that subclass were older, had higher prevalence of females, coronary heart disease, hypertension, obesity, and were also more likely to be prescribed analgesics, diuretics and NSAIDS than subclass 2 in which, allopurinol was found to

have the weakest protective effect (0.81 (0.68, 0.97)). Most patients in subclass 5 had a recorded baseline SU level, which was above 360µmol/L (Table 7.8).

The pooled treatment effect also showed that allopurinol was associated with increased risk of gout hospitalisation (1.82 (1.64, 2.02)). In Table 7.9, the highest risk of gout hospitalisation was observed in subclass 1 (2.46 (1.78, 3.40)). Patients in subclass 1 were slightly younger and resided in less deprived areas, had lower prevalence of coronary heart disease, osteoarthritis, analgesics, diuretics and NSAIDS use, and were more likely to be an ever drinker, never smoker and have normal-overweight BMI value, compared with patients in subclass 5 in which allopurinol users had the lowest risk of gout hospitalisation (1.46 (1.23, 1.74)).

Test for proportional hazards (PH) failed for allopurinol in the outcome analyses of target SU level, repeated gout consultations and renal disease in unadjusted Cox models. On graphical inspection of the Schoenfeld residuals plotted over time (presented in Appendix J), the log-HR was constant i.e., had a zero slope, until towards the end of follow-up where its direction changed from being constant to either increasing or decreasing. The change in HR may indicate that the lessening effect of treatment was due to a small number of patients with the longest follow-up times. Given the PH was satisfied for the majority of follow-up, and the Kaplan-Meier plots had shown no crossover of survival functions between the two treatment groups (graphs not shown), it was assumed overall, that the PH assumption was satisfied.

After PS subclassification, the PH assumption for allopurinol was met in the majority of subclasses for all outcome analyses with the following exceptions. For mortality, the PH assumption failed in one subclass; for repeated gout consultations the PH assumption failed in three subclasses; for renal disease the PH assumption failed in two subclasses. As above, on further inspection of the Schoenfeld residuals the decrease in log-HR towards the end of follow-up was due to a small number of patients with the longest follow-up times (Appendix

J). Kaplan-Meier plots had shown no crossover of survival functions between the two treatment groups (graphs not shown). Therefore, it was assumed the PH assumption was satisfied.

Table 7.7: Estimated treatment effect of allopurinol on outcome (1 year landmark analysis)

Outcome	No allopurinol N (%)	Allopurinol N (%)	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error
SU level				
Target level not met	789 (67)	246 (43)	2.27 (1.96, 2.64)	2.32 (1.97, 2.74)
Target level met	386 (33)	321 (57)	0.17	0.19
Mortality				
Alive	9,260 (72)	2,640 (67)	1.23 (1.15, 1.31)	1.10 (1.03, 1.17)
Died	3,659 (28)	1,317 (33)	0.04	0.04
Repeated gout consultations				
Never consulted	5,787 (45)	2,193 (55)	0.75 (0.70, 0.80)	0.70 (0.65, 0.75)*
Consulted at least once	7,132 (55)	1,764 (45)	0.03	0.03
Gout hospitalisation				
No	9,546 (88)	2,569 (79)	1.97 (1.80, 2.17)	1.82 (1.64, 2.02)*
Yes	1,283 (12)	689 (21)	0.09	0.10
Joint replacement				
No	12,008 (94)	3,621 (93)	1.26 (1.10, 1.45)	1.15 (0.99, 1.32)
Yes	744 (6)	271 (7)	0.09	0.08
Cerebrovascular disease				
No	11,273 (90)	3,428 (91)	1.03 (0.91, 1.16)	0.98 (0.87, 1.11)
Yes	1,195 (10)	357 (9)	0.06	0.06
Coronary heart disease				
Yes	8,675 (80)	2,438 (77)	1.16 (1.07, 1.26)	1.11 (1.02, 1.21)
No	2,228 (20)	722 (23)	0.05	0.05
Peripheral vascular disease				
No	12,186 (96)	3,698 (96)	1.10 (0.92, 1.32)	1.01 (0.84, 1.22)
Yes	480 (4)	155 (4)	0.10	0.10
Renal disease				
No	9,867 (78)	2,716 (72)	1.36 (1.27, 1.46)	1.19 (1.10, 1.28)
Yes	2,864 (23)	1,061 (28)	0.05	0.04

*Subclass-specific treatment effect estimates were not homogenous; CI: Confidence interval; SU: Serum urate

Table 7.8: Estimated treatment effect of allopurinol on repeated gout consultations and distribution of covariates within each PS subclass

	Subclass 1 N=3,376	Subclass 2 N=3,375	Subclass 3 N=3,375	Subclass 4 N=3,375	Subclass 5 N=3,375
HR (95% CI)	0.75 (0.62, 0.92)	0.81 (0.68, 0.97)	0.76 (0.65, 0.88)	0.62 (0.54, 0.71)	0.60 (0.53, 0.67)
Standard error	0.08	0.07	0.06	0.04	0.03
Demographics					
Age (Mean (SD))	59.8 (14.0)	62.0 (15.2)	59.6 (14.4)	64.8 (14.9)	64.1 (14.4)
Sex: Female	788 (23)	674 (20)	545 (16)	910 (27)	964 (29)
Deprivation (Mean (SD))	7.6 (5.1)	8.8 (5.5)	9.2 (5.4)	9.3 (5.5)	10.7 (5.6)
Comorbidities					
Anxiety	136 (4)	155 (5)	113 (3)	125 (4)	143 (4)
Depression	199 (6)	177 (5)	162 (5)	154 (5)	150 (4)
Cerebrovascular disease	47 (1)	66 (2)	68 (2)	92 (3)	134 (4)
Coronary heart disease	213 (6)	306 (9)	381 (11)	495 (15)	772 (23)
Diabetes	143 (4)	155 (5)	176 (5)	211 (6)	362 (11)
Hyperlipidaemia	155 (5)	171 (5)	122 (4)	167 (5)	168 (5)
Hypertension	387 (11)	578 (17)	489 (14)	812 (24)	871 (26)
Osteoarthritis	127 (4)	139 (4)	175 (5)	276 (8)	389 (12)
Peripheral vascular disease	23 (1)	25 (1)	40 (1)	73 (2)	96 (3)
Renal disease	NA (<1)	NA (0)	9 (<1)	42 (1)	162 (5)
Lifestyle factors					
Alcohol consumption					
Ever drinker	2,089 (62)	1,997 (59)	1,838 (54)	1,863 (55)	1,701 (50)
Never drinker	194 (6)	192 (6)	123 (4)	203 (6)	144 (4)
Missing	1,093 (32)	1,186 (35)	1,414 (42)	1,309 (39)	1,530 (45)
Body mass index					
Normal	772 (23)	659 (20)	382 (11)	433 (13)	271 (8)
Overweight	1,107 (33)	1,089 (32)	934 (28)	1,013 (30)	790 (23)
Obese	447 (13)	457 (14)	665 (20)	684 (20)	966 (29)
Missing	1,050 (31)	1,170 (35)	1,394 (41)	1,245 (37)	1,348 (40)
Smoking status					
Ever smoker	1,423 (42)	1,263 (37)	1,158 (34)	1,312 (39)	1,280 (38)
Never smoker	1,119 (33)	1,097 (33)	964 (29)	934 (28)	733 (22)
Missing	834 (25)	1,015 (30)	1,253 (37)	1,129 (33)	1,362 (40)
SU level					
≤360µmol/L	944 (28)	7 (<1)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	258 (8)	1,073 (32)	1,778 (53)	2,953 (87)
Missing	2,432 (72)	3,110 (92)	2,302 (68)	1,597 (47)	422 (13)
Medication use					
Analgesics	526 (16)	730 (22)	1,011 (30)	1,420 (42)	1,891 (56)
Colchicine	8 (<1)	12 (<1)	26 (1)	78 (2)	265 (8)
Diuretics	389 (12)	1,080 (32)	743 (22)	1,952 (58)	2,075 (61)
NSAIDS	292 (9)	942 (28)	1,765 (52)	2,066 (61)	2,862 (85)

N (%) presented unless otherwise stated; CI: Confidence interval; NA: Cannot report cell counts with less than five events; HR: Hazard ratio; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

Table 7.9: Estimated treatment effect of allopurinol on gout hospitalisation and distribution of covariates within each PS subclass

	Subclass 1 N=2,818	Subclass 2 N=2,817	Subclass 3 N=2,818	Subclass 4 N=2,817	Subclass 5 N=2,817
HR (95% CI)	2.46 (1.78, 3.40)	1.98 (1.57, 2.49)	1.60 (1.29, 1.98)	1.75 (1.44, 2.11)	1.46 (1.23, 1.74)
Standard error	0.40	0.23	0.17	0.17	0.13
Demographics					
Age (Mean (SD))	62.5 (13.6)	63.9 (14.8)	62.0 (14.0)	65.3 (14.7)	64.1 (14.2)
Sex: Female	722 (26)	591 (21)	548 (19)	785 (28)	829 (29)
Deprivation (Mean (SD))	7.8 (5.1)	9.0 (5.4)	9.3 (5.4)	9.6 (5.6)	10.5 (5.6)
Comorbidities					
Anxiety	108 (4)	117 (4)	123 (4)	103 (4)	135 (5)
Depression	140 (5)	133 (5)	157 (6)	148 (5)	171 (6)
Cerebrovascular disease	31 (1)	58 (2)	66 (2)	88 (3)	126 (4)
Coronary heart disease	213 (8)	301 (11)	354 (13)	441 (16)	627 (22)
Diabetes	158 (6)	165 (6)	149 (5)	213 (8)	262 (9)
Hyperlipidaemia	122 (4)	131 (5)	126 (4)	150 (5)	158 (6)
Hypertension	478 (17)	578 (21)	463 (16)	640 (23)	599 (21)
Osteoarthritis	107 (4)	99 (4)	178 (6)	246 (9)	381 (14)
Peripheral vascular disease	31 (1)	44 (2)	50 (2)	54 (2)	61 (2)
Renal disease	NA (<1)	NA (0)	18 (1)	39 (1)	127 (5)
Lifestyle factors					
Alcohol consumption					
Ever drinker	1,782 (63)	1,615 (57)	1,640 (58)	1,516 (54)	1,440 (51)
Never drinker	190 (7)	175 (6)	140 (5)	159 (6)	113 (4)
Missing	846 (30)	1,027 (36)	1,038 (37)	1,142 (41)	1,264 (45)
Body mass index					
Normal	678 (24)	504 (18)	412 (15)	324 (12)	253 (9)
Overweight	927 (33)	884 (31)	858 (30)	821 (29)	695 (25)
Obese	381 (14)	423 (15)	519 (18)	618 (22)	803 (29)
Missing	832 (30)	1,006 (36)	1,029 (37)	1,054 (37)	1,066 (38)
Smoking status					
Ever smoker	1,236 (44)	1,047 (37)	1,074 (38)	1,060 (38)	1,115 (40)
Never smoker	961 (34)	904 (32)	844 (30)	760 (27)	569 (20)
Missing	621 (22)	866 (31)	900 (32)	997 (35)	1,133 (40)
SU level					
≤360µmol/L	811 (29)	NA (<1)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	98 (3)	779 (28)	1,591 (56)	2,555 (91)
Missing	2,007 (71)	2,716 (96)	2,039 (72)	1,226 (44)	262 (9)
Medication use					
Analgesics	544 (19)	727 (26)	972 (34)	1,229 (44)	1,545 (55)
Colchicine	8 (<1)	8 (<1)	27 (1)	65 (2)	235 (8)
Diuretics	339 (12)	991 (35)	787 (28)	1,648 (59)	1,714 (61)
NSAIDS	359 (13)	901 (32)	1,635 (58)	1,564 (56)	2,365 (84)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity Score; SD: Standard deviation; SU: Serum urate

7.5.2 Treatment effect stratified by presence of renal disease

Of the study sample (n=16,876), 217 patients had renal disease at baseline. In patients with no renal disease, allopurinol users had higher prevalence of SU being recorded, the SU level being ≥360µmol/L, and were more likely to be prescribed analgesics, colchicine, diuretics and NSAIDS than non-users (SMD >0.10). In addition to the imbalanced covariates identified in patients with no renal disease, in patients with renal disease more differences were observed between the treatment groups with SMD >0.10 for BMI distribution, never smokers, and cerebrovascular disease, and peripheral vascular disease (Table 7.10).

PS subclassification was performed in patients with no renal disease. Mean and range of PS was similar between treatment groups indicating common support. Five subclasses were sufficient to achieve overall covariate balance across subclasses although some imbalance remained within subclasses. As in the main analysis, subclass 1 observed the most number of imbalanced covariates (Table 7.11).

PS subclassification could not be performed in patients with renal disease due to insufficient sample size and low frequency of occurrence of outcomes. For example, for mortality, five subclasses were sufficient to achieve overall balance in covariates across subclasses. However, this meant there were approximately 43 patients in each of the five subclasses, there was substantial imbalance in the majority of covariates within a subclass, and outcome did not occur within treatment groups in certain subclasses. Adjustment for all imbalanced covariates in estimating subclass-specific treatment effect was not possible due to small number of deaths in that subclass. Therefore, PS subclassification was not used. Instead, multivariable Cox regression adjusted for covariates that were originally included in the PS model. Given the low frequency of occurrence of outcome for gout hospitalisation, cerebrovascular disease, and coronary heart disease, the Cox model only adjusted for age, sex and BMI. Due to an even

lower frequency of occurrence of outcome for target SU level, joint replacement, and peripheral vascular disease, unadjusted HRs were presented.

Table 7.10: Baseline covariates stratified by renal disease

	No renal disease N=16,508		Renal disease N=217	
Baseline covariates	No allopurinol N=12,731	Allopurinol N=3,777	No allopurinol N=121	Allopurinol N=96
Demographics				
Age (Mean (SD))	61.7 (14.8)	62.5 (14.5)	70.8 (13.4)	72 (12.4)
Sex: Female	2,860 (22)	892 (24)	40 (33)	36 (38)
Deprivation (Mean (SD))	9.0 (5.5)	9.5 (5.7)	10.5 (5.3)	10.4 (5.3)
Comorbidities				
Anxiety	522 (4)	140 (4)	NA (3)	NA (2)
Depression	645 (5)	180 (5)	NA (2)	NA (4)
Cerebrovascular disease	291 (2)	105 (3)	NA (1)	5 (5)
Coronary heart disease	1,491 (12)	573 (15)	33 (27)	24 (25)
Diabetes	715 (6)	252 (7)	25 (21)	20 (21)
Hyperlipidaemia	580 (5)	164 (4)	9 (7)	9 (9)
Hypertension	2,301 (18)	726 (19)	42 (35)	32 (33)
Osteoarthritis	755 (6)	313 (8)	11 (9)	10 (10)
Peripheral vascular disease	170 (1)	63 (2)	5 (4)	12 (13)
Lifestyle factors				
Alcohol consumption				
Ever drinker	7,216 (57)	2,068 (55)	69 (57)	51 (53)
Never drinker	650 (5)	179 (5)	8 (7)	6 (6)
Missing	4,865 (38)	1,530 (41)	44 (36)	39 (41)
Body mass index				
Normal	1,986 (16)	469 (12)	22 (18)	20 (21)
Overweight	3,773 (30)	1,052 (28)	38 (31)	27 (28)
Obese	2,321 (18)	808 (21)	22 (18)	23 (24)
Missing	4,651 (37)	1,448 (38)	39 (32)	26 (27)
Smoking status				
Ever smoker	4,863 (38)	1,418 (38)	53 (44)	39 (41)
Never smoker	3,717 (29)	1,012 (27)	32 (26)	33 (34)
Missing	4,151 (33)	1,347 (36)	36 (30)	24 (25)
SU level				
≤360µmol/L	893 (7)	53 (1)	NA (2)	NA (1)
>360µmol/L	3,990 (31)	1,904 (50)	51 (42)	52 (54)
Missing	7,848 (62)	1,820 (48)	68 (56)	43 (45)
Medication use				
Analgesic	3,891 (31)	1,480 (39)	62 (51)	61 (64)
Colchicine	224 (2)	143 (4)	8 (7)	8 (8)
Diuretic	4,207 (33)	1,648 (44)	85 (70)	85 (89)
NSAIDS	5,613 (44)	2,241 (59)	45 (37)	45 (47)

N (%) presented unless otherwise stated; Cells highlighted in yellow indicate standardised mean difference >0.10; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Table 7.11: Distribution of PS subclasses in patients with no renal disease

Outcome	PS mean (range) Allopurinol No allopurinol	Number of subclasses[a]	Smallest cell[b]	Imbalanced covariates (SMD >0.10)
Target SU level	0.30 (0.08, 0.81) 0.36 (0.13, 0.83)	5	51/506 (10%)	S1: Age, sex, coronary heart disease, hyperlipidaemia, alcohol consumption, smoking status, SU level, analgesics S2: Sex, hyperlipidaemia, alcohol consumption, BMI, smoking status, SU level, diuretics S3: Age, coronary heart disease, alcohol consumption, BMI, smoking status, diuretics S4: Coronary heart disease, alcohol consumption, BMI, NSAIDS S5: Sex, BMI
Mortality	0.21 (0.02, 0.71) 0.28 (0.03, 0.72)	5	299/3,777 (8%)	S1: Sex, alcohol consumption, BMI, SU level, colchicine S2: Sex, deprivation, BMI, NSAIDS S4: Sex
Repeated gout consultations	0.21 (0.03, 0.71) 0.28 (0.03, 0.72)	5	304/3,777 (8%)	S1: Sex, alcohol consumption, BMI, SU level S2: Deprivation, NSAIDS
Gout hospitalisation	0.21 (0.02, 0.71) 0.27 (0.03, 0.72)	5	246/3,116 (8%)	S1: Sex, alcohol consumption, BMI, SU level, S2: Sex, BMI, NSAIDS S3: Alcohol consumption S4: Colchicine
Joint replacement	0.21 (0.03, 0.70) 0.28 (0.03, 0.69)	5	297/3,720 (8%)	S1: Sex, alcohol consumption, BMI, smoking status, SU level S2: NSAIDS
Cerebrovascular disease	0.21 (0.02, 0.70) 0.28 (0.03, 0.67)	5	287/3,621 (8%)	S1: Sex, alcohol consumption, BMI, SU level, colchicine S2: Deprivation S3: SU level S4: Sex
Coronary heart disease	0.21 (0.02, 0.70) 0.27 (0.04, 0.69)	5	245/3,050 (8%)	S1: Sex, cerebrovascular disease, BMI, SU level, colchicine S2: Deprivation, BMI S3: Alcohol consumption
Peripheral vascular disease	0.21 (0.02, 0.71) 0.28 (0.03, 0.72)	5	292/3,690 (8%)	S1: Sex, alcohol consumption, BMI, SU level, colchicine S2: BMI S3: SU level

[a]Number of subclasses needed to achieve overall covariate balance across subclasses; [b]Smallest cell defined as the number of allopurinol users within a subclass; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; S: Subclass; SMD: Standardised mean difference; SU: Serum urate

Table 7.12 provides the estimated treatment effect for each outcome stratified by presence of renal disease. In patients with no renal disease, allopurinol had a higher chance of reaching target SU level (2.26 (1.89, 2.71)), fewer gout consultations (0.71 (0.66, 0.76)), higher risk of gout hospitalisation (1.86 (1.65, 2.09)) and undergoing joint replacement (1.24 (1.06, 1.46)).

Subclass-specific treatment effects estimates are reported in Table 7.13 and Table 7.14 for repeated gout consultations and gout hospitalisation respectively, as the subclass-specific treatment effect estimates were not homogenous. As in the main analysis (Section 7.5.1), within subclass 5, allopurinol was found to have the strongest protective effect against repeated gout consultations (0.56 (0.50, 0.63)). Within subclass 1, allopurinol was found to have the highest risk of gout hospitalisation (2.96 (2.09, 4.20)).

In patients with renal disease, allopurinol use was not associated with any of the outcomes. Although not statistically significant, adjusted analyses had shown allopurinol may be protective against premature mortality (0.83 (0.57, 1.21)) and repeated gout consultations (0.58 (0.34, 0.97)), but may have higher risk of gout hospitalisation (1.48 (0.79, 2.80)), cerebrovascular disease (1.20 (0.57, 2.53)), and coronary heart disease (1.16 (0.55, 2.43)).

Table 7.12: Estimated treatment effect of allopurinol stratified by renal disease

	No renal disease				Renal disease			
	No allopurinol	Allopurinol	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error	No allopurinol	Allopurinol	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error
SU level								
Target SU level not met	743 (69)	233 (46)	2.27 (1.92, 2.67)[a]	2.26 (1.89, 2.71)[a]	9 (64)	10 (71)	0.81 (0.22, 2.94)	N/A
Target SU level met	333 (31)	273 (54)	0.19	0.21	5 (36)	NA (28)	0.53	
Mortality								
Alive	10,194 (80)	3,000 (79)	1.10 (1.02, 1.19)	1.02 (0.94, 1.11)	41 (34)	27 (28)	1.18 (0.86, 1.63)	0.83 (0.57, 1.21)
Dead	2,537 (20)	777 (21)	0.05	0.04	80 (66)	69 (72)	0.19	0.16
Repeated gout consultations								
Never consulted	6,022 (47)	2,193 (58)	0.75 (0.70, 0.80)[a]	0.71 (0.66, 0.76)[ab]	60 (50)	66 (69)	0.65 (0.40, 1.04)[a]	0.58 (0.34, 0.97)[a]
Consulted at least once	6,709 (53)	1,584 (42)	0.03	0.03	61 (50)	30 (31)	0.16	0.15
Gout hospitalisation								
No	9,720 (91)	2,642 (85)	1.95 (1.75, 2.18)	1.86 (1.65, 2.09)[ab]	95 (83)	59 (78)	1.53 (0.80, 2.94)	1.48 (0.79, 2.80)[c]
Yes	934 (9)	474 (15)	0.11	0.11	19 (17)	17 (22)	0.51	0.48
Joint replacement								
No	11,938 (95)	3,493 (94)	1.33 (1.14, 1.54)	1.24 (1.06, 1.46)	115 (97)	88 (97)	1.04 (0.24, 4.49)	N/A
Yes	628 (5)	227 (6)	0.10	0.10	NA (3)	NA (3)	0.78	
Cerebrovascular disease								
No	11,334 (92)	3,357 (93)	1.00 (0.87, 1.14)	0.96 (0.83, 1.11)	103 (879)	71 (83)	1.34 (0.66, 2.73)	1.20 (0.57, 2.53)[c]
Yes	955 (8)	264 (7)	0.07	0.07	15 (13)	15 (17)	0.42	0.46
Coronary heart disease								
No	8,788 (82)	2,447 (80)	1.13 (1.03, 1.24)	1.10 (1.00, 1.20)	64 (77)	46 (77)	1.17 (0.58, 2.34)	1.16 (0.55, 2.43)[c]
Yes	1,992 (18)	603 (20)	0.05	0.05	19 (23)	14 (23)	0.41	0.44
Peripheral vascular disease								
No	12,099 (97)	3,582 (97)	0.99 (0.80, 1.23)	0.94 (0.75, 1.17)	107 (93)	76 (93)	1.11 (0.39, 3.15)	N/A
Yes	391 (3)	108 (3)	0.11	0.11	8 (7)	6 (7)	0.59	

[a]Statistical test for the proportional hazards assumption failed, however change in treatment effect was represented by a small number of patients with the longest follow-up times, and there was no cross-over of survival functions from treatment groups. Therefore, the proportional hazards assumption was satisfied. [b]Subclass-specific treatment effects were homogenous; [c]Adjusted for age, sex, BMI; N/A: Adjusted Cox model not fitted due to small number of outcomes; CI: Confidence interval; NA: Cannot report cell counts with less than five events; SU: Serum urate

Table 7.13: Estimated PS subclass-specific treatment effect of allopurinol on repeated gout consultations in those with no renal disease

	Subclass 1 N=3,302	Subclass 2 N=3,302	Subclass 3 N=3,301	Subclass 4 N=3,302	Subclass 5 N=3,301
HR (95% CI)	0.80 (0.65, 0.98)	0.83 (0.69, 1.00)	0.69 (0.60, 0.81)	0.68 (0.58, 0.79)	0.56 (0.50, 0.63)
Standard error	0.08	0.08	0.05	0.05	0.03
Demographics					
Age (Mean (SD))	59.6 (13.8)	62.9 (15.4)	59.0 (14.1)	64.9 (15.0)	62.9 (14.4)
Sex: Female	751 (23)	711 (22)	498 (15)	914 (28)	878 (27)
Deprivation (Mean (SD))	7.3 (5.0)	9.0 (5.7)	9.0 (5.3)	9.6 (5.6)	10.5 (5.6)
Comorbidities					
Anxiety	132 (4)	146 (4)	108 (3)	140 (4)	136 (4)
Depression	190 (6)	191 (6)	139 (4)	160 (5)	145 (4)
Cerebrovascular disease	43 (1)	74 (2)	64 (2)	84 (3)	131 (4)
Coronary heart disease	189 (6)	335 (10)	331 (10)	501 (15)	708 (21)
Diabetes	142 (4)	170 (5)	153 (5)	210 (6)	292 (9)
Hyperlipidaemia	163 (5)	172 (5)	107 (3)	168 (5)	134 (4)
Hypertension	356 (11)	631 (19)	411 (12)	803 (24)	826 (25)
Osteoarthritis	121 (4)	137 (4)	179 (5)	271 (8)	360 (11)
Peripheral vascular disease	26 (1)	31 (1)	35 (1)	67 (2)	74 (2)
Lifestyle factors					
Alcohol consumption					
Ever drinker	1,790 (54)	1,854 (56)	1,861 (56)	1,862 (56)	1,917 (58)
Never drinker	156 (5)	173 (5)	128 (4)	203 (6)	169 (5)
Missing	1,356 (41)	1,275 (39)	1,312 (40)	1,237 (37)	1,215 (37)
Body mass index					
Normal	638 (19)	568 (17)	456 (14)	434 (13)	359 (11)
Overweight	944 (29)	994 (30)	959 (29)	1,026 (31)	902 (27)
Obese	443 (13)	499 (15)	585 (18)	689 (21)	913 (28)
Missing	1,277 (39)	1,241 (38)	1,301 (39)	1,153 (35)	1,127 (34)
Smoking status					
Ever smoker	1,181 (36)	1,181 (36)	1,175 (36)	1,283 (39)	1,461 (44)
Never smoker	988 (30)	1,031 (31)	977 (30)	960 (29)	773 (23)
Missing	1,133 (34)	1,090 (33)	1,149 (35)	1,059 (32)	1,067 (32)
SU level					
≤360µmol/L	940 (28)	6 (<1)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	139 (4)	1,095 (33)	1,666 (50)	2,994 (91)
Missing	2,362 (72)	3,157 (96)	2,206 (67)	1,636 (50)	307 (9)
Medication use					
Analgesics	450 (14)	823 (25)	847 (26)	1,487 (45)	1,764 (53)
Colchicine	7 (<1)	12 (<1)	23 (1)	71 (2)	254 (8)
Diuretics	255 (8)	1,214 (37)	583 (18)	1,967 (60)	1,836 (56)
NSAIDS	387 (12)	821 (25)	1,795 (54)	1,970 (60)	2,881 (87)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

Table 7.14: Estimated PS subclass-specific treatment effect of allopurinol on gout hospitalisation in those with no renal disease

	Subclass 1 N=2,755	Subclass 2 N=2,753	Subclass 3 N=2,754	Subclass 4 N=2,754	Subclass 5 N=2,754
HR (95% CI)	2.96 (2.09, 4.20)	1.82 (1.38, 2.40)	1.64 (1.28, 2.10)	1.69 (1.34, 2.13)	1.47 (1.19, 1.82)
Standard error	0.53	0.26	0.21	0.20	0.16
Demographics					
Age (Mean (SD))	62.7 (13.6)	63.6 (14.9)	62.4 (13.8)	64.7 (14.8)	63.6 (14.1)
Sex: Female	722 (26)	579 (21)	539 (20)	757 (27)	765 (28)
Deprivation (Mean (SD))	7.8 (5.2)	9.1 (5.4)	9.1 (5.4)	9.8 (5.6)	10.4 (5.6)
Comorbidities					
Anxiety	106 (4)	116 (4)	113 (4)	111 (4)	130 (5)
Depression	139 (5)	129 (5)	149 (5)	148 (5)	169 (6)
Cerebrovascular disease	30 (1)	64 (2)	65 (2)	83 (3)	118 (4)
Coronary heart disease	202 (7)	298 (11)	327 (12)	431 (16)	592 (21)
Diabetes	158 (6)	166 (6)	140 (5)	190 (7)	226 (8)
Hyperlipidaemia	119 (4)	128 (5)	120 (4)	144 (5)	144 (5)
Hypertension	470 (17)	555 (20)	461 (17)	605 (22)	575 (21)
Osteoarthritis	103 (4)	100 (4)	182 (7)	231 (8)	359 (13)
Peripheral vascular disease	31 (1)	39 (1)	48 (2)	51 (2)	50 (2)
Lifestyle factors					
Alcohol consumption					
Ever drinker	1,741 (63)	1,553 (56)	1,636 (59)	1,457 (53)	1,428 (52)
Never drinker	196 (7)	168 (6)	134 (5)	151 (5)	105 (4)
Missing	818 (30)	1,032 (37)	984 (36)	1,146 (42)	1,221 (44)
Body mass index					
Normal	678 (25)	483 (18)	405 (15)	305 (11)	246 (9)
Overweight	912 (33)	835 (30)	868 (32)	795 (29)	681 (25)
Obese	378 (14)	426 (15)	508 (18)	595 (22)	759 (28)
Missing	787 (29)	1,009 (37)	973 (35)	1,059 (38)	1,068 (39)
Smoking status					
Ever smoker	1,226 (45)	1,005 (37)	1,067 (39)	1,013 (37)	1,086 (39)
Never smoker	949 (34)	868 (32)	849 (31)	722 (26)	548 (20)
Missing	580 (21)	880 (32)	838 (30)	1,019 (37)	1,120 (41)
SU level					
≤360µmol/L	805 (29)	NA (0)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	75 (3)	719 (26)	1,544 (56)	2,541 (92)
Missing	1,950 (71)	2,674 (97)	2,035 (74)	1,210 (44)	213 (8)
Medication use					
Analgesics	528 (19)	722 (26)	927 (34)	1,202 (44)	1,459 (53)
Colchicine	7 (<1)	8 (0)	28 (1)	64 (2)	220 (8)
Diuretics	355 (13)	994 (36)	744 (27)	1,569 (57)	1,571 (57)
NSAIDS	352 (13)	799 (29)	1,684 (61)	1,486 (54)	2,359 (86)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

7.5.3 Treatment effect stratified by severity of hyperuricaemia

Of the study sample (n=16,876), 3,160 patients had non-severe hyperuricaemia and 2,902 had severe hyperuricaemia. Among patients with non-severe hyperuricaemia, allopurinol users were older and more likely to be female, reside in a more deprived area, have osteoarthritis, and be prescribed analgesics, diuretics and NSAIDS than non-users (SMD >0.10). In patients with severe hyperuricaemia, allopurinol users had higher prevalence of osteoarthritis and were more likely to be prescribed analgesics, diuretics and NSAIDS than non-users (SMD >0.10) (Table 7.15).

PS subclassification for each outcome in non-severe and severe hyperuricaemic groups was performed. Table 7.16 and Table 7.17 shows the mean and range of PS were similar between treatment groups indicating common support between treatment groups in non-severe and severe hyperuricaemic groups respectively. In both stratified analyses, five subclasses were sufficient to achieve overall balance for the majority of covariates however, imbalance remained within subclasses. In patients with non-severe hyperuricaemia, there were a greater number of imbalanced covariates within subclasses compared with patients with severe hyperuricaemia (Table 7.16 and Table 7.17 respectively). In the analysis of peripheral vascular disease, in patients with severe hyperuricaemia, use of four or five subclasses led to no occurrence of outcome in subclass 1 (the lowest PS), thus this subclass could not be used in outcome analysis as treatment effect could not be estimated. Therefore, three PS subclasses were created which still achieved overall covariate balance across subclasses.

Table 7.15: Baseline characteristics stratified by severity of hyperuricaemia

	Non-severe hyperuricaemia (360-480µmol/L) N=3,160		Severe hyperuricaemia (>480µmol/L) N=2,902	
Baseline covariates	No allopurinol N=2,432	Allopurinol use N=728	No allopurinol N=1,632	Allopurinol N=1,270
Demographics				
Age (Mean (SD))	61.3 (13.7)	63.0 (13.0)	61.5 (16.0)	63.1 (15.3)
Sex: Female	443 (18)	195 (27)	385 (24)	322 (25)
Deprivation (Mean (SD))	8.6 (5.4)	9.9 (5.8)	9.1 (5.5)	9.6 (5.7)
Comorbidities				
Anxiety	92 (4)	26 (4)	64 (4)	52 (4)
Depression	113 (5)	36 (5)	72 (4)	63 (5)
Cerebrovascular disease	54 (2)	10 (1)	38 (2)	44 (3)
Coronary heart disease	280 (12)	96 (13)	255 (16)	237 (19)
Diabetes	118 (5)	37 (5)	116 (7)	122 (10)
Hyperlipidaemia	107 (4)	29 (4)	85 (5)	73 (6)
Hypertension	479 (20)	155 (21)	354 (22)	293 (23)
Osteoarthritis	146 (6)	69 (9)	95 (6)	115 (9)
Peripheral vascular disease	29 (1)	16 (2)	25 (2)	25 (2)
Renal disease	14 (1)	10 (1)	37 (2)	42 (3)
Lifestyle factors				
Alcohol consumption				
Ever drinker	1,401 (58)	400 (55)	930 (57)	727 (57)
Never drinker	82 (3)	31 (4)	80 (5)	54 (4)
Missing	949 (39)	297 (41)	622 (38)	489 (39)
Body mass index				
Normal	316 (13)	98 (13)	202 (12)	147 (12)
Overweight	777 (32)	194 (27)	477 (29)	370 (29)
Obese	474 (19)	164 (23)	377 (23)	302 (24)
Missing	865 (36)	272 (37)	576 (35)	451 (36)
Smoking status				
Ever smoker	1,025 (42)	301 (41)	739 (45)	557 (44)
Never smoker	599 (25)	171 (23)	380 (23)	281 (22)
Missing	808 (33)	256 (35)	513 (31)	432 (34)
Medication use				
Analgesic	684 (28)	278 (38)	539 (33)	550 (43)
Colchicine	47 (2)	26 (4)	48 (3)	53 (4)
Diuretic	789 (32)	275 (38)	783 (48)	682 (54)
NSAIDS	1,099 (45)	422 (58)	781 (48)	761 (60)

N (%) presented unless otherwise stated; Cells highlighted in yellow indicate SMD >0.10; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Table 7.16: Distribution of PS subclasses in patients with non-severe hyperuricaemia

	PS Mean (range) Allopurinol No allopurinol	Number of PS subclasses[a]	Smallest cell[b]	Imbalanced covariates (SMD>0.10)
Target SU level	0.23 (0.05, 0.57) 0.27 (0.09, 0.58)	5	19/195 (10%)	S1: Age, deprivation, coronary heart disease, renal disease, alcohol consumption, smoking status, analgesics, colchicine, diuretics S2: Age, coronary heart disease, hyperlipidaemia, renal disease, alcohol consumption, BMI, SU level, analgesics, diuretics S3: Age, sex, coronary heart disease, hyperlipidaemia, alcohol consumption, BMI, SU level, diuretics, S4: Age, sex, coronary heart disease, alcohol consumption, BMI, smoking status, SU level, analgesics, diuretics S5: Age, deprivation, hyperlipidaemia, BMI,
Mortality	0.22 (0.07, 0.64) 0.26 (0.08, 0.62)	5	72/728 (10%)	S1: Sex, diabetes, hypertension, osteoarthritis, alcohol consumption, smoking status, analgesics, NSAIDS S2: Deprivation, coronary heart disease, hypertension, renal disease, alcohol consumption, BMI, analgesics, NSAIDS S3: age, osteoarthritis, peripheral vascular disease, BMI, smoking status, analgesics, NSAIDS S4: cerebrovascular disease, hypertension, alcohol consumption, BMI, smoking status, colchicine S5: Hypertension
Repeated gout consultations	0.22 (0.07, 0.64) 0.26 (0.08, 0.63)	5	69/728 (9%)	S1: Sex, diabetes, alcohol consumption, smoking status, analgesics S2: Deprivation, depression, coronary heart disease, hyperlipidaemia, renal disease, alcohol consumption, analgesics, colchicine S3: Peripheral vascular disease, smoking status, analgesics S4: Depression, renal disease, BMI, smoking status
Gout hospitalisation	0.23 (0.07, 0.55) 0.26 (0.09, 0.64)	5	62/626 (9.90%)	S1: Sex, diabetes, hypertension, osteoarthritis, alcohol consumption, smoking status, analgesics, NSAIDS S2: Age, deprivation, cerebrovascular disease, coronary heart disease, osteoarthritis, alcohol consumption, BMI, analgesics, NSAIDS S3: BMI, smoking status, analgesics S4: Age, deprivation, cerebrovascular disease, renal disease, alcohol consumption S5: Deprivation, cerebrovascular disease, alcohol consumption, analgesics

Table 7.16 continued:

Joint replacement	0.22 (0.07, 0.59) 0.25 (0.08, 0.61)	5	74/715 (10%)	S1: Sex, osteoarthritis, alcohol consumption, smoking status, analgesics, diuretics S2: Deprivation, coronary heart disease, hypertension, osteoarthritis, renal disease, alcohol consumption, BMI, analgesics, NSAIDS S3: osteoarthritis, smoking status, analgesics, NSAIDS S4: BMI S5: Hypertension
Cerebrovascular disease	0.22 (0.10, 0.60) 0.26 (0.11, 0.59)	5	70/709 (9%)	S1: Sex, hypertension, alcohol consumption, smoking status, analgesics, diuretics, NSAIDS S2: Coronary heart disease, alcohol consumption, smoking status, analgesics S3: osteoarthritis, alcohol consumption, BMI, smoking status, analgesics, colchicine S4: Deprivation, diabetes, renal disease, alcohol consumption, BMI, smoking status S5: Hypertension, BMI, analgesics
Coronary heart disease	0.22 (0.07, 0.60) 0.25 (0.10, 0.51)	5	58/595 (10%)	S1: Sex, deprivation, diabetes, hyperlipidaemia, hypertension, peripheral vascular disease, alcohol consumption, BMI, smoking status, analgesics, diuretics S2: Alcohol consumption, BMI, analgesics S3: Sex, cerebrovascular disease, osteoarthritis, renal disease, alcohol consumption, BMI, smoking status, analgesics S4: Deprivation, alcohol consumption, BMI, smoking status, NSAIDS S5: Hypertension
Peripheral vascular disease	0.22 (0.06, 0.60) 0.25 (0.07, 0.63)	5	67/706 (9%)	S1: Sex, deprivation, diabetes, hyperlipidaemia, osteoarthritis, alcohol consumption, smoking status, analgesics, NSAIDS S2: Sex, coronary heart disease, hypertension, alcohol consumption, smoking status, analgesics S3: Coronary heart disease, diabetes, osteoarthritis, alcohol consumption, smoking status, analgesics S4: Deprivation, hypertension, smoking status S5: Hypertension, analgesics
Renal disease	0.22 (0.05, 0.62) 0.25 (0.08, 0.55)	5	70/713 (10%)	S1: Sex, deprivation, diabetes, hypertension, osteoarthritis, alcohol consumption, smoking status, analgesics S2: Sex, deprivation, coronary heart disease, analgesics, NSAIDS S3: Age, analgesics, deprivation, smoking status S4: Deprivation, smoking status S5: Hypertension, smoking status

[a]Number of subclasses needed to achieve overall balance across subclasses; [b]Smallest cell defined as the number of allopurinol users within a subclass; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; S: Subclass; SMD: Standardised mean difference; SU: Serum urate

Table 7.17: Distribution of PS subclasses in patients with severe hyperuricaemia

	PS Mean (range) Allopurinol No allopurinol	Number of PS subclasses[a]	Smallest cell[b]	Imbalanced covariates (SMD>0.10)
Target SU level	0.39 (0.17, 0.82) 0.43 (0.23, 0.89)	5	51/372 (14%)	S1: Age, sex, alcohol consumption, BMI, smoking status, SU level, analgesics, colchicine, diuretics, S2: Sex, deprivation, smoking status, analgesics, colchicine, diuretics S3: Age, hyperlipidaemia, alcohol consumption, smoking status, colchicine S4: Deprivation, renal disease, alcohol consumption, BMI, smoking status S5: Hyperlipidaemia, renal disease, alcohol consumption, BMI, smoking status
Mortality	0.42 (0.24, 0.74) 0.46 (0.25, 0.76)	5	171/1,270 (13%)	S1: osteoarthritis, alcohol consumption, smoking status S2: Diabetes, BMI, diuretics, NSAIDS S3: Peripheral vascular disease, diuretics S4: Smoking status S5: BMI, smoking status
Repeated gout consultations	0.42 (0.24, 0.72) 0.46 (0.24, 0.77)	5	173/1,270 (14%)	S1: Sex, smoking status, diuretics S2: Diabetes, BMI, diuretics S3: Depression, coronary heart disease S4: Smoking status, diuretics S5: Smoking status
Gout hospitalisation	0.42 (0.25, 0.71) 0.45 (0.24, 0.75)	5	145/1,036 (14$)	S1: osteoarthritis, BMI, smoking status, NSAIDS S2: Diabetes, renal disease, BMI, NSAIDS S3: Cerebrovascular disease, coronary heart disease S4: Age, cerebrovascular disease, alcohol consumption S5: Deprivation, renal disease
Joint replacement	0.43 (0.25, 0.71) 0.46 (0.24, 0.71)	5	164/1,254 (13%)	S1: Coronary heart disease, hypertension, alcohol consumption, smoking status, analgesics, NSAIDS S2: BMI S3: Sex, coronary heart disease S4: osteoarthritis, alcohol consumption, smoking status S5: Smoking status
Cerebrovascular disease	0.42 (0.23, 0.73) 0.45 (0.26, 0.69)	5	164/1,210 (14%)	S1: Sex, coronary heart disease, hyperlipidaemia, osteoarthritis, renal disease, smoking status, diuretics S2: BMI. Diuretics S3: BMI, analgesics S4: Alcohol consumption, BMI, Smoking status, diuretics S5: BMI, smoking status

Table 7.17 continued:

Coronary heart disease	0.41 (0.21, 0.75) 0.45 (0.22, 0.77)	5	126/972 (13%)	S1: Sex, diabetes, hyperlipidaemia, osteoarthritis, peripheral vascular disease, alcohol consumption, BMI, smoking status, colchicine S2: Renal disease, diuretics S3: Age, diabetes, peripheral vascular disease S4: Age, sex, cerebrovascular disease, diabetes, alcohol consumption, BMI, colchicine, diuretics, NSAIDS S5: Cerebrovascular disease, hyperlipidaemia
Peripheral vascular disease	0.42 (0.23, 0.74) 0.46 (0.23, 0.74)	3	305/1,231 (25%)	S1: Smoking status, diuretics
Renal disease	0.41 (0.24, 0.72) 0.45 (0.24, 0.71)	5	161/1,191 (14%)	S1: osteoarthritis, alcohol consumption, smoking status, diuretics S3: Coronary heart disease S4: Age, diabetes, BMI, smoking status, diuretics S5: NSAIDS

[a]Number of propensity score subclasses needed to achieve overall balance across subclasses; [b]Smallest cell defined as the number of allopurinol users within a subclass; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; S: Subclass; SMD: Standardised mean difference; SU: Serum urate

Table 7.18 gives the estimated treatment effect for each outcome stratified by severity of hyperuricaemia. Treatment effect estimates for target SU level was greater in patients with non-severe hyperuricaemia than those with severe hyperuricaemia (HR 2.83 vs. 2.20). In patients with non-severe hyperuricaemia, allopurinol had a higher risk of joint replacement (1.37 (1.10, 1.84)) and coronary heart disease (1.27 (1.04, 1.55)) than non-users. Pooled effect of allopurinol across subclasses had fewer gout consultations with effect sizes that were similar across the stratified groups. Risk of gout hospitalisation was higher amongst allopurinol users in patients with non-hyperuricaemia (1.61 (1.24, 2.08)) but no association was observed in patients with severe hyperuricaemia.

Table 7.19 and Table 7.20 show subclass specific treatment effect estimates for repeated gout consultations in patients with non-severe and severe hyperuricaemia respectively as the treatment effect estimates were not homogenous across subclasses. In non-severe hyperuricaemic patients, within subclass 2, allopurinol use was found to have the strongest protective effect against repeated gout consultations (0.48 (0.35, 0.66)); these patients resided in less deprived areas and had lower prevalence of females, coronary heart disease, hypertension, osteoarthritis and obesity, had higher prevalence of ever drinker, overweight BMI, ever smoker, and prescribed analgesics, colchicine, diuretics and NSAIDS less than subclass 5, in which, allopurinol was found to have the least protective treatment effect (0.75 (0.57, 0.98)). In contrast, among severe hyperuricaemic patients, within subclass 5, allopurinol use was found to have the strongest protective effect against repeated gout consultations (0.44 (0.34, 0.57)) compared to subclass 2, in which, allopurinol was found to have the least protective treatment effect (0.69 (0.51, 0.93)).

Table 7.21 gives subclass specific treatment effect estimates for gout hospitalisation in patients with non-severe hyperuricaemia only as the treatment effect estimates were not homogenous across subclasses. No association was observed between allopurinol use and

gout hospitalisation in subclasses 1, 2 and 4. In subclasses 3 and 5, allopurinol use doubled the risk of gout hospitalisation (2.08 (1.26, 3.44) and 2.90 (1.86, 4.53) respectively). In patients with severe hyperuricaemia, there was no association between allopurinol and gout hospitalisation (1.09 (0.91, 1.31)).

Table 7.18: Treatment effect of allopurinol stratified by severity of hyperuricaemia

	Non-severe hyperuricaemia				Severe hyperuricaemia			
	No allopurinol	Allopurinol	Unadjusted HR ratio (95% CI) Standard error	Adjusted HR Ratio (95% CI) Standard error	No allopurinol	Allopurinol	Unadjusted HR (95% CI) Standard error	Adjusted HR (95% CI) Standard error
SU level								
Not reached target	419 (67)	74 (38)	2.42 (1.92, 3.05)	2.83 (2.17, 3.68)	370 (68)	172 (46)	2.17 (1.77, 2.66)	2.20 (1.76, 2.75)
Reached target	209 (33)	121 (62)	0.28	0.38	177 (32)	200 (54)	0.23	0.25
Mortality								
Alive	1,880 (77)	543 (75)	1.10 (0.93, 1.29)	0.93 (0.79, 1.11)	1,125 (69)	828 (65)	1.13 (1.00, 1.29)	0.96 (0.84, 1.09)
Dead	552 (23)	185 (25)	0.09	0.08	507 (31)	442 (35)	0.07	0.06
Repeated gout consultations								
Never consulted	1,727 (52)	433 (55)	0.74 (0.65, 0.84)[a]	0.60 (0.53, 0.69)[ab]	503 (31)	694 (55)	0.57 (0.51, 0.64)[a]	0.57 (0.51, 0.64)[ab]
Consulted at least once	1,602 (48)	349 (45)	0.05	0.04	1,129 (69)	576 (45)	0.03	0.03
Gout hospitalisation								
No	1,784 (89)	505 (81)	1.79 (1.44, 2.24)	1.61 (1.24, 2.08)[b]	1,103 (81)	816 (79)	1.18 (0.99, 1.42)	1.09 (0.91, 1.31)
Yes	223 (11)	121 (19)	0.20	0.21	251 (19)	220 (21)	0.11	0.10
Joint replacement								
No	2,263 (94)	650 (91)	1.55 (1.15, 2.08)	1.37 (1.01, 1.84)	1,525 (95)	1,177 (94)	1.30 (0.95, 1.78)	1.08 (0.77, 1.51)
Yes	142 (6)	65 (9)	0.23	0.21	78 (5)	77 (6)	0.21	0.18
Cerebrovascular disease								
No	2,151 (91)	631 (89)	1.26 (0.97, 1.63)	1.26 (0.96, 1.65)	1,452 (93)	1,100 (91)	1.26 (0.97, 1.64)	1.04 (0.79, 1.37)[a]
Yes	206 (9)	78 (11)	0.17	0.17	115 (7)	110 (9)	0.17	0.15
Coronary heart disease								
No	1,663 (81)	455 (76)	1.24 (1.02, 1.50)	1.27 (1.04, 1.55)	1,016 (79)	753 (77)	1.08 (0.90, 1.29)	1.01 (0.83, 1.21)
Yes	396 (19)	140 (24)	0.12	0.13	274 (21)	219 (23)	0.10	0.10
Peripheral vascular disease								
No	2,298 (96)	682 (97)	0.86 (0.55, 1.35)	0.76 (0.48, 1.21)[a]	1,542 (97)	1,186 (96)	1.22 (0.82, 1.84)[a]	1.02 (0.67, 1.57)
Yes	92 (4)	24 (3)	0.20	0.18	48 (3)	45 (4)	0.25	0.22
Renal disease								
No	1,847 (77)	516 (72)	1.21 (1.03, 1.43)[a]	1.06 (0.90, 1.25)[a]	1,113 (70)	797 (67)	1.15 (1.01, 1.32)	1.05 (0.91, 1.20)[a]
Yes	561 (23)	197 (27)	0.10	0.09	469 (30)	394 (33)	0.08	0.07

[a]Statistical test for the proportional hazards assumption failed, however change in treatment effect was represented by a small number of patients with the longest follow-up times, and there was no cross-over of survival functions for treatment groups. Therefore, the proportional hazards assumption was satisfied. [b]Subclass-specific treatment effects were not homogenous; CI: Confidence intervals; HR: Hazard ratio; SU: Serum urate

Table 7.19: Estimated PS subclass-specific treatment effect of allopurinol on repeated gout consultations in patients with non-severe hyperuricaemia

	Subclass 1 N=632	Subclass 2 N=632	Subclass 3 N=632	Subclass 4 N=632	Subclass 5 N=632
HR (95% CI)	0.65 (0.48, 0.89)	0.48 (0.35, 0.66)	0.66 (0.49, 0.89)	0.52 (0.39, 0.68)	0.75 (0.57, 0.98)
Standard error	0.11	0.08	0.10	0.07	0.10
Demographics					
Age (Mean (SD))	56.5 (13.0)	59.8 (13.4)	60.7 (13.5)	63.7 (13.2)	67.8 (12.1)
Sex: Female	9 (1)	48 (8)	82 (13.0)	158 (25)	341 (54.0)
Deprivation (Mean (SD))	4.6 (3.3)	7.9 (4.6)	8.2 (5.5)	10.4 (5.1)	13.3 (4.9)
Comorbidities					
Anxiety	21 (3)	18 (3)	21 (3)	24 (4)	34 (5)
Depression	27 (4)	28 (4)	27 (4)	29 (5)	38 (6)
Cerebrovascular disease	37 (6)	12 (2)	7 (1)	8 (1)	0 (0)
Coronary heart disease	42 (7)	60 (9)	83 (13)	80 (13)	111 (18)
Diabetes	35 (6)	17 (3)	34 (5)	29 (5)	40 (6)
Hyperlipidaemia	27 (4)	39 (6)	24 (4)	27 (4)	19 (3)
Hypertension	105 (17)	103 (16)	129 (20)	119 (19)	178 (28)
Osteoarthritis	8 (1)	21 (3)	30 (5)	55 (9)	101 (16)
Peripheral vascular disease	0 (0)	NA (<1)	NA (<1)	7 (1)	33 (5)
Renal disease	0 (0)	NA (<1)	NA (<1)	NA (<1)	21 (3)
Lifestyle factors					
Alcohol consumption					
Ever drinker	434 (69)	376 (59)	343 (54)	328 (52)	320 (51)
Never drinker	9 (1)	11 (2)	23 (4)	23 (4)	47 (7)
Missing	189 (30)	245 (39)	266 (42)	281 (44)	265 (42)
Body mass index					
Normal	78 (12)	78 (12)	82 (13)	90 (14)	86 (14)
Overweight	325 (51)	241 (38)	174 (28)	121 (19)	110 (17)
Obese	78 (12)	100 (16)	116 (18)	139 (22)	205 (32)
Missing	151 (24)	213 (34)	260 (41)	282 (45)	231 (37)
Smoking status					
Ever smoker	276 (44)	284 (45)	270 (43)	253 (40)	243 (38)
Never smoker	199 (31)	152 (24)	132 (21)	134 (21)	153 (24)
Missing	157 (25)	196 (31)	230 (36)	245 (39)	236 (37)
Medication use					
Analgesics	32 (5)	85 (13)	163 (26)	257 (41)	425 (67)
Colchicine	0 (0)	NA (<1)	6 (1)	14 (2)	52 (8)
Diuretics	135 (21)	169 (27)	186 (29)	223 (35)	351 (56)
NSAIDS	19 (3)	146 (23)	357 (56)	449 (71)	550 (87)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation

Table 7.20: Estimated PS subclass-specific treatment effect of allopurinol on repeated gout consultations and distribution of covariates in patients with severe hyperuricaemia

	Subclass 1 N=581	Subclass 2 N=580	Subclass 3 N=581	Subclass 4 N=580	Subclass 5 N=580
HR (95% CI)	0.64 (0.49, 0.85)	0.69 (0.51, 0.93)	0.49 (0.38, 0.61)	0.65 (0.51, 0.83)	0.44 (0.34, 0.57)
Standard error	0.09	0.11	0.06	0.08	0.06
Demographics					
Age (Mean (SD))	55.8 (14.8)	62.2 (16.0)	60.6 (16.2)	64.2 (15.3)	69.2 (12.9)
Sex: Female	107 (18)	125 (22)	125 (22)	160 (28)	190 (33)
Deprivation (Mean (SD))	7.4 (5.1)	8.6 (5.6)	9.0 (5.4)	10.2 (5.5)	11.2 (5.6)
Comorbidities					
Anxiety	19 (3)	17 (3)	16 (3)	28 (5)	36 (6)
Depression	23 (4)	17 (3)	24 (4)	23 (4)	48 (8)
Cerebrovascular disease	0 (0)	NA (<1)	14 (2)	20 (3)	46 (8)
Coronary heart disease	36 (6)	84 (14)	80 (14)	116 (20)	176 (30)
Diabetes	9 (2)	30 (5)	29 (5)	52 (9)	118 (20)
Hyperlipidaemia	25 (4)	23 (4)	31 (5)	32 (6)	47 (8)
Hypertension	101 (17)	128 (22)	116 (20)	136 (23)	166 (29)
Osteoarthritis	11 (2)	18 (3)	41 (7)	55 (9)	85 (15)
Peripheral vascular disease	NA (<1)	10 (2)	6 (1)	10 (2)	22 (4)
Renal disease	NA (<1)	12 (2)	5 (1)	16 (3)	45 (8)
Lifestyle factors					
Alcohol consumption					
Ever drinker	327 (56)	318 (55)	364 (63)	313 (54)	335 (58)
Never drinker	41 (7)	37 (6)	16 (3)	24 (4)	16 (3)
Missing	213 (37)	225 (39)	201 (35)	243 (42)	229 (39)
Body mass index					
Normal	89 (15)	75 (13)	64 (11)	67 (12)	54 (9)
Overweight	175 (30)	176 (30)	171 (29)	159 (27)	166 (29)
Obese	126 (22)	125 (22)	146 (25)	116 (20)	166 (29)
Missing	191 (33)	204 (35)	200 (34)	238 (41)	194 (33)
Smoking status					
Ever smoker	285 (49)	277 (48)	255 (44)	248 (43)	231 (40)
Never smoker	159 (27)	138 (24)	146 (25)	106 (18)	112 (19)
Missing	137 (24)	165 (28)	180 (31)	226 (39)	237 (41)
Medication use					
Analgesics	11 (2)	134 (23)	174 (30)	267 (46)	503 (87)
Colchicine	NA (<1)	12 (2)	13 (2)	13 (2)	61 (11)
Diuretics	162 (28)	282 (49)	250 (43)	331 (57)	440 (76)
NSAIDS	NA (<1)	161 (28)	372 (64)	469 (81)	538 (93)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation

Table 7.21: Estimated PS subclass-specific treatment effect of allopurinol on gout hospitalisation in patients with non-severe hyperuricaemia

	Subclass 1 N=527	Subclass 2 N=527	Subclass 3 N=526	Subclass 4 N=527	Subclass 5 N=526
HR (95% CI)	1.69 (0.80, 3.58)	0.83 (0.43, 1.57)	2.08 (1.26, 3.44)	1.27 (0.80, 2.03)	2.90 (1.86, 4.53)
Standard error	0.65	0.27	0.53	0.30	0.66
Demographics					
Age (Mean (SD))	61.1 (12.6)	62.8 (12.9)	62.2 (13.3)	63.8 (12.9)	66.5 (12.9)
Sex: Female	20 (4)	66 (13)	85 (16)	138 (26)	266 (51)
Deprivation (Mean (SD))	5.2 (3.6)	7.9 (4.8)	8.6 (5.4)	10.2 (5.3)	13.6 (4.6)
Comorbidities					
Anxiety	15 (3)	20 (4)	20 (4)	18 (3)	30 (6)
Depression	18 (3)	20 (4)	32 (6)	20 (4)	38 (7)
Cerebrovascular disease	34 (6)	14 (3)	NA (1)	NA (1)	NA (0)
Coronary heart disease	46 (9)	61 (12)	67 (13)	81 (15)	92 (17)
Diabetes	37 (7)	26 (5)	35 (7)	21 (4)	26 (5)
Hyperlipidaemia	26 (5)	21 (4)	27 (5)	19 (4)	21 (4)
Hypertension	132 (25)	112 (21)	110 (21)	94 (18)	114 (22)
Osteoarthritis	5 (1)	13 (2)	27 (5)	41 (8)	110 (21)
Peripheral vascular disease	NA (<1)	6 (1)	14 (3)	6 (1)	13 (2)
Renal disease	0 (0)	0 (0)	NA (<1)	NA (1)	18 (3)
Lifestyle factors					
Alcohol consumption					
Ever drinker	378 (72)	320 (61)	300 (57)	276 (52)	230 (44)
Never drinker	9 (2)	12 (2)	17 (3)	17 (3)	45 (9)
Missing	140 (27)	195 (37)	209 (40)	234 (44)	251 (48)
Body mass index					
Normal	50 (9)	74 (14)	73 (14)	64 (12)	83 (16)
Overweight	273 (52)	194 (37)	149 (28)	112 (21)	88 (17)
Obese	93 (18)	91 (17)	103 (20)	118 (22)	143 (27)
Missing	111 (21)	168 (32)	201 (38)	233 (44)	212 (40)
Smoking status					
Ever smoker	275 (52)	239 (45)	230 (44)	213 (40)	178 (34)
Never smoker	156 (30)	138 (26)	120 (23)	107 (20)	104 (20)
Missing	96 (18)	150 (28)	176 (33)	207 (39)	244 (46)
Medication use					
Analgesics	41 (8)	118 (22)	161 (31)	228 (43)	334 (63)
Colchicine	NA (<1)	7 (1)	10 (2)	15 (3)	32 (6)
Diuretics	156 (30)	166 (31)	172 (33)	196 (37)	269 (51)
NSAIDS	9 (2)	119 (23)	286 (54)	396 (75)	487 (93)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

7.6 Sensitivity analyses

7.6.1 Treatment effect using two-year landmark period

For this analysis, 15,873 patients had a minimum of two years follow-up. During this period 28% of patients were prescribed allopurinol. Table 7.22 describes the distribution of baseline covariates between treatment groups. Similarly to the one-year landmark analysis (main analysis), allopurinol users resided in more deprived areas and had higher prevalence of recorded SU level and SU level above target (>360μmol/L), and were prescribed colchicine, diuretics, NSAIDS and analgesics more than non-users.

After PS estimation, the distribution of PS was similar between treatment groups (Table 7.23). Within most outcome analyses, five subclasses were sufficient to achieve overall balance for all covariates across subclasses although imbalance remained on some covariates within subclasses; the number of imbalanced covariates within subclasses were similar to what was found in the main analysis. In the analysis of cerebrovascular disease, six subclasses were required to achieve overall covariate balance compared with the main analysis that required five subclasses.

Table 7.22: Baseline covariates by treatment: two-year landmark period

	No allopurinol N=11,360	Allopurinol N=4,513	SMD
Demographics			
Age (Mean (SD))	61.4 (14.6)	62.2 (14.4)	0.05
Sex: Female	2,557 (23)	1,051 (23)	0.02
Deprivation (Mean (SD))	8.9 (5.5)	9.5 (5.6)	0.11
Comorbidities			
Anxiety	452 (4)	163 (4)	-0.02
Depression	554 (5)	225 (5)	0.001
Cerebrovascular disease	244 (2)	119 (3)	0.03
Coronary heart disease	1,295 (11)	678 (15)	0.10
Diabetes	641 (6)	307 (7)	0.05
Hyperlipidaemia	529 (5)	222 (5)	0.01
Hypertension	2,078 (18)	888 (20)	0.04
Osteoarthritis	678 (6)	361 (8)	0.08
Peripheral vascular disease	150 (1)	74 (2)	0.03
Renal disease	86 (1)	91 (2)	0.10
Lifestyle factors			
Alcohol consumption			
Ever drinker	6,434 (57)	2,495 (55)	-0.03
Never drinker	570 (5)	221 (5)	-0.01
Missing	4,356 (38)	1,797 (40)	0.03
Body mass index			
Normal	1,790 (16)	554 (12)	-0.10
Overweight	3,363 (30)	1,303 (29)	-0.02
Obese	2,081 (18)	955 (21)	0.07
Missing	4,126 (36)	1,701 (38)	0.03
Smoking status			
Ever smoker	4,330 (38)	1,708 (38)	-0.01
Never smoker	3,327 (29)	1,224 (27)	-0.05
Missing	3,703 (33)	1,581 (35)	0.05
SU level			
≤360µmol/L	846 (7)	64 (1)	-0.30
>360µmol/L	3,505 (31)	2,229 (49)	0.39
Missing	7,009 (62)	2,220 (49)	-0.25
Medication use			
Analgesics	3,378 (30)	1,761 (39)	0.20
Colchicine	177 (2)	179 (4)	0.15
Diuretics	3,638 (32)	1,963 (43)	0.24
NSAIDS	4,967 (44)	2,627 (58)	0.29

N (%) were presented unless otherwise stated; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SMD: Standardised mean difference; SU: Serum urate

Table 7.23: Distribution of PS subclasses for the two year landmark period

Outcome	PS Mean (range) Allopurinol No allopurinol	Number of subclasses[a]	Smallest cell[b]	Imbalanced covariates (SMD>0.10)
Target SU level	0.33 (0.10, 0.82) 0.42 (0.13, 0.91)	5	83/298 (28%)	S1: Age, sex, coronary heart disease, renal disease, alcohol consumption, BMI, smoking status, analgesics S2: SU level, colchicine, NSAIDS S3: Coronary heart disease, hyperlipidaemia, BMI, SU level, colchicine S4: Age, hyperlipidaemia, renal disease, NSAIDS S5: Sex, coronary heart disease, SU level, colchicine, NSAIDS
Mortality	0.26 (0.03, 0.79) 0.34 (0.04, 0.82)	5	376/4,513 (8%)	S1: Sex, SU level S2: NSAIDS S4: Sex
Repeated gout consultations	0.26 (0.03, 0.77) 0.34 (0.04, 0.81)	5	377/4,513 (8%)	S1: Sex, SU level S4: SU level
Gout hospitalisation	0.26 (0.03, 0.78) 0.33 (0.04, 0.80)	5	312/3,687 (8%)	S1: Sex, SU level S2: Sex, NSAIDS S3: Alcohol consumption S4: Sex
Joint replacement	0.26 (0.03, 0.79) 0.34 (0.04, 0.82)	5	366/4,417 (8%)	S1: Sex, SU level S4: SU level
Cerebrovascular disease	0.26 (0.03, 0.78) 0.33 (0.04, 0.79)	6	291/4,281 (7%)	S1: Sex, SU level, colchicine S2: BMI, NSAIDS S5: Sex
Coronary heart disease	0.25 (0.03, 0.76) 0.32 (0.04, 0.79)	5	308/3,464 (9%)	S1: Sex, SU level S3: Alcohol consumption S4: Sex, SU level
Peripheral vascular disease	0.27 (0.03, 0.79) 0.34 (0.04, 0.81)	5	369/4,380 (8%)	S1: Sex, SU level
Renal disease	0.26 (0.03, 0.78) 0.33 (0.04, 0.79)	5	363/4,273 (9%)	S1: Sex, SU level S2: NSAIDS S4: SU level

[a]Number of subclasses needed to achieve overall covariate balance across subclasses; [b]Smallest cell defined as the number of allopurinol users within a subclass; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; S: Subclass; SD: Standard deviation; SMD: Standardised mean difference; SU: Serum urate

The treatment effect estimate was obtained for each outcome as shown in Table 7.24. Allopurinol had greater chance of reaching target SU level (2.15 (1.79, 2.58)), and higher risk of premature mortality (1.11 (1.04, 1.18)), and renal disease (1.20 (1.11, 1.29)); these treatment effect estimates and standard errors were similar to what was estimated in the main analysis. Both this analysis (two-year landmark period) and the main analysis found allopurinol increased risk of joint replacement but the effect was greater in the two-year landmark period (HR 1.27 vs. 1.15).

The subclass-specific treatment effect estimates were not homogenous across subclasses for repeated gout consultations and gout hospitalisation; these estimates are presented in Table 7.25 and Table 7.26 respectively.

Similarly as the main analysis, all subclasses had shown allopurinol was protective against repeated gout consultations, with subclass 5 found to have the strongest protective effect. However, the subclass-specific treatment effects were closer to the null than the main analysis.

Within all subclasses, allopurinol increased the risk of gout hospitalisation with the strongest effect observed in subclass 1. Compared with the main analysis, the subclass-specific treatment effects were greater across all subclasses. Pooled HRs were larger in this analysis compared with the main analysis (2.05 vs. 1.82).

Table 7.24: Treatment effect of allopurinol using two-year landmark period

Outcome	No allopurinol	Allopurinol	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error
Target SU level				
Target level not met	655 (69)	265 (49)	2.11 (1.79, 2.49)	2.15 (1.79, 2.58)
Target level met	298 (31)	273 (51)	0.18	0.20
Mortality				
Alive	8,353 (74)	3,115 (69)	1.23 (1.15, 1.31)	1.11 (1.04, 1.18)
Died	3,007 (26)	1,398 (31)	0.040	0.04
Repeated gout consultations				
Never consulted	5,428 (48)	2,548 (56)	0.83 (0.79, 0.88)	0.77 (0.73, 0.82)[ab]
Consulted at least once	5,932 (52)	1,965 (44)	0.02	0.02
Gout hospitalisation				
No	8,444 (89)	2,894 (78)	2.23 (2.03, 2.45)	2.05 (1.86, 2.27)[ab]
Yes	1,031 (11)	793 (22)	0.11	0.11
Joint replacement				
No	10,531 (94)	4,111 (93)	1.34 (1.17, 1.53)	1.27 (1.10, 1.46)
Yes	615 (6)	306 (7)	0.09	0.09
Cerebrovascular disease				
No	9,904 (91)	3,905 (91)	1.03 (0.91, 1.16)	1.00 (0.87, 1.14)
Yes	972 (9)	376 (9)	0.06	0.06
Coronary heart disease				
Yes	7,622 (82)	2,778 (80)	1.12 (1.02, 1.22)	1.10 (1.00, 1.20)[a]
No	1,709 (18)	686 (20)	0.05	0.05
Peripheral vascular disease				
No	10,726 (97)	4,208 (96)	1.25 (1.04, 1.50)	1.16 (0.96, 1.40)
Yes	366 (3)	172 (4)	0.12	0.11
Renal disease				
No	8,598 (77)	3,060 (72)	1.34 (1.25, 1.44)[a]	1.20 (1.11, 1.29)[a]
Yes	2,564 (23)	1,213 (28)	0.05	0.04

[a]Statistical test for the proportional hazards assumption failed, however change in treatment effect was represented by a small number of patients with the longest follow-up times, and there was no cross-over of survival functions from treatment groups. Therefore, the proportional hazards assumption was satisfied; [b]Subclass specific treatment effects were not homogenous across subclasses; CI: Confidence interval; SU: Serum urate

Table 7.25: Estimated PS subclass-specific treatment effect of allopurinol on repeated gout consultations and distribution of covariates

	Subclass 1 N=3,175	Subclass 2 N=3,175	Subclass 3 N=3,174	Subclass 4 N=3,175	Subclass 5 N=3,174
HR (95% CI)	0.85 (0.72, 1.00)	0.83 (0.73, 0.94)	0.83 (0.73, 0.94)	0.69 (0.62, 0.78)	0.68 (0.62, 0.76)
Standard error	0.07	0.05	0.05	0.04	0.04
Demographics					
Age (Mean (SD))	60.3 (13.7)	61.6 (15.0)	59.4 (14.1)	63.7 (14.9)	63.2 (14.4)
Sex: Female	779 (25)	639 (20)	517 (16)	812 (26)	861 (27)
Deprivation (Mean (SD))	7.5 (5.1)	8.6 (5.4)	9.2 (5.4)	9.4 (5.6)	10.8 (5.6)
Comorbidities					
Anxiety	114 (4)	136 (4)	517 (16)	110 (3)	141 (4)
Depression	158 (5)	146 (5)	114 (4)	139 (4)	170 (5)
Cerebrovascular disease	46 (1)	54 (2)	166 (5)	86 (3)	115 (4)
Coronary heart disease	191 (6)	273 (9)	62 (2)	455 (14)	713 (22)
Diabetes	150 (5)	142 (4)	341 (11)	192 (6)	304 (10)
Hyperlipidaemia	138 (4)	147 (5)	160 (5)	153 (5)	181 (6)
Hypertension	361 (11)	559 (18)	132 (4)	742 (23)	844 (27)
Osteoarthritis	123 (4)	131 (4)	460 (14)	254 (8)	360 (11)
Peripheral vascular disease	27 (1)	29 (1)	171 (5)	63 (2)	68 (2)
Renal disease	NA (<1)	NA (<1)	37 (1)	24 (1)	143 (5)
Lifestyle factors					
Alcohol consumption					
Ever drinker	1,904 (60)	1,844 (58)	1,746 (55)	1,775 (56)	1,660 (52)
Never drinker	172 (5)	170 (5)	122 (4)	179 (6)	148 (5)
Missing	1,099 (35)	1,161 (37)	1,306 (41)	1,221 (38)	1,366 (43)
Body mass index					
Normal	756 (24)	616 (19)	362 (11)	369 (12)	241 (8)
Overweight	975 (31)	969 (31)	925 (29)	987 (31)	810 (26)
Obese	451 (14)	474 (15)	598 (19)	663 (21)	850 (27)
Missing	993 (31)	1,116 (35)	1,289 (41)	1,156 (36)	1,273 (40)
Smoking status					
Ever smoker	1,297 (41)	1,182 (37)	1,108 (35)	1,219 (38)	1,232 (39)
Never smoker	1,057 (33)	1,022 (32)	908 (29)	884 (28)	680 (21)
Missing	821 (26)	971 (31)	1,158 (36)	1,072 (34)	1,262 (40)
SU level					
≤360µmol/L	903 (28)	7 (<1)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	281 (9)	1,017 (32)	1,702 (54)	2,734 (86)
Missing	2,272 (72)	2,887 (91)	2,157 (68)	1,473 (46)	440 (14)
Medication use					
Analgesics	473 (15)	617 (19)	925 (29)	1,321 (42)	1,803 (57)
Colchicine	6 (<1)	7 (<1)	16 (1)	48 (2)	279 (9)
Diuretics	291 (9)	916 (29)	734 (23)	1,715 (54)	1,945 (61)
NSAIDS	380 (12)	913 (29)	1,640 (52)	1,988 (63)	2,673 (84)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

Table 7.26: Estimated PS subclass-specific treatment effect of allopurinol on gout hospitalisation and distribution of covariates

	Subclass 1 N=2,634	Subclass 2 N=2,631	Subclass 3 N=2,633	Subclass 4 N=2,632	Subclass 5 N=2,632
HR (95% CI)	2.82 (2.10, 3.79)	2.04 (1.63, 2.56)	1.75 (1.42, 2.16)	2.02 (1.67, 2.45)	1.79 (1.49, 2.16)
Standard error	0.42	0.24	0.19	0.20	0.17
Demographics					
Age (Mean (SD))	63.2 (13.1)	62.9 (14.8)	61.9 (13.7)	64.1 (14.7)	63.4 (14.1)
Sex: Female	726 (28)	545 (21)	524 (20)	700 (27)	710 (27)
Deprivation (Mean (SD))	7.8 (5.1)	8.9 (5.4)	9.3 (5.4)	9.6 (5.6)	10.6 (5.6)
Comorbidities					
Anxiety	94 (4)	111 (4)	101 (4)	104 (4)	119 (5)
Depression	128 (5)	123 (5)	135 (5)	136 (5)	163 (6)
Cerebrovascular disease	36 (1)	53 (2)	56 (2)	79 (3)	103 (4)
Coronary heart disease	194 (7)	242 (9)	323 (12)	393 (15)	611 (23)
Diabetes	159 (6)	150 (6)	154 (6)	176 (7)	216 (8)
Hyperlipidaemia	116 (4)	126 (5)	121 (5)	138 (5)	153 (6)
Hypertension	445 (17)	533 (20)	442 (17)	596 (23)	578 (22)
Osteoarthritis	114 (4)	119 (5)	159 (6)	235 (9)	318 (12)
Peripheral vascular disease	30 (1)	32 (1)	39 (1)	51 (2)	56 (2)
Renal disease	NA (<1)	NA (<1)	13 (<1)	26 (1)	106 (4)
Lifestyle factors					
Alcohol consumption					
Ever drinker	1,593 (60)	1,527 (58)	1,504 (57)	1,441 (55)	1,409 (54)
Never drinker	160 (6)	157 (6)	118 (4)	160 (6)	119 (5)
Missing	881 (33)	947 (36)	1,011 (38)	1,031 (39)	1,104 (42)
Body mass index					
Normal	680 (26)	477 (18)	371 (14)	272 (10)	706 (27)
Overweight	812 (31)	821 (31)	795 (30)	810 (31)	210 (8)
Obese	359 (14)	407 (15)	498 (19)	580 (22)	725 (28)
Missing	783 (30)	926 (35)	969 (37)	970 (37)	991 (38)
Smoking status					
Ever smoker	1,112 (42)	982 (37)	991 (38)	1,016 (39)	1,053 (40)
Never smoker	886 (34)	869 (33)	763 (29)	713 (27)	538 (20)
Missing	636 (24)	780 (30)	879 (33)	903 (34)	1,041 (40)
SU level					
≤360µmol/L	772 (29)	5 (<1)	0 (0)	0 (0)	0 (0)
>360µmol/L	0 (0)	141 (5)	714 (27)	1,506 (57)	2,352 (89)
Missing	1,862 (71)	2,485 (94)	1,919 (73)	1,126 (43)	280 (11)
Medication use					
Analgesics	509 (19)	619 (24)	903 (34)	1,111 (42)	1,440 (55)
Colchicine	5 (<1)	5 (<1)	13 (<1)	45 (2)	241 (9)
Diuretics	318 (12)	818 (31)	755 (29)	1,441 (55)	1,621 (62)
NSAIDS	356 (14)	857 (33)	1,512 (57)	1,515 (58)	2,172 (83)

N (%) presented unless otherwise stated; CI: Confidence interval; HR: Hazard ratio; NA: Cannot report cell counts with less than five events; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate

7.6.2 Impact of omission of confounding variables

The effect of omission of a single unmeasured binary confounding variable was evaluated for target SU level and mortality. The prevalence of the confounding variable amongst allopurinol users and non-users, and its strength of association with outcome, are unknown. To obtain prevalence and association estimates that can be seen as representing plausible scenarios which could be assumed for an unmeasured binary confounding variable, prevalence of each observed covariate amongst allopurinol users and non-users and its association (HR) with outcome were calculated.

In the analysis of target SU level, the estimated unadjusted HR between observed covariates and target SU level ranged between 0.84 and 1.63, with peripheral vascular disease having the largest association thus, it was assumed the association between an unmeasured confounding variable and target SU level could be represented by a HR of 1.63. Prevalence of covariates ranged from 2% to 66% in allopurinol users, and 1% to 60% in non-users. As seen in Table 7.27 (and previously in Table 7.7), the effect of allopurinol on target SU level assuming no unmeasured confounding was HR 2.32 (1.97, 2.74). In the presence of an unmeasured confounding variable, as the prevalence of the unmeasured confounding variable increased amongst allopurinol users but remained fixed in non-users, the estimated HR corrected for unmeasured confounding decreased but remained statistically significant; the smallest HR observed was 1.95 (1.35, 2.84).

In the analysis of mortality, the estimated unadjusted HRs between observed covariates and mortality ranged from 0.89 to 3.87, with renal disease having the largest association. The unmeasured confounding variable was assumed to have a more conservative HR of 1.45, the median HR of observed HRs. Prevalence of any covariate ranged from 1% to 59% in allopurinol users and 1% to 56% in non-users. As seen in Table 7.28, allopurinol use was associated with greater risk of premature mortality assuming no unmeasured confounding (HR (1.10 (1.03,

1.17)). As the prevalence of the unmeasured confounding variable increased amongst allopurinol users but remained fixed in non-users, the estimated HR corrected for unmeasured confounding decreased and lost statistical significance. However, as the difference in prevalence of the unmeasured confounding variable had become larger between treatment groups, potentially, allopurinol may be protective against premature mortality with the smallest HR of 0.83 (0.76, 0.89).

Table 7.27: Effect of omitted confounding on treatment effect of allopurinol on target SU level

	Prevalence of unmeasured cofounding variable among non-users			
Prevalence of unmeasured confounding variable among allopurinol users	0%	10%	30%	50%
0%	2.32 (1.97, 2.74)			
10%	2.26 (1.56, 3.28)	2.32 (1.97, 2.74)		
30%	2.15 (1.48, 3.12)	2.21 (1.52, 3.20)	2.32 (1.97, 2.74)	
50%	2.05 (1.41, 2.97)	2.11 (1.45, 3.06)	2.22 (1.53, 3.22)	2.32 (1.97, 2.74)
70%	1.95 (1.35, 2.84)	2.02 (1.39, 2.93)	2.13 (1.47, 3.09)	2.23 (1.54, 3.23)

Hazard ratios (95% confidence interval) are presented; SU: Serum urate

Table 7.28: Effect of omitted confounding on treatment effect of allopurinol on mortality

	Prevalence of unmeasured confounding variable among non-users			
Prevalence of unmeasured confounding variable among allopurinol users	0%	10%	30%	50%
0%	1.10 (1.03, 1.17)			
10%	1.06 (0.98, 1.14)	1.10 (1.03, 1.17)		
30%	0.97 (0.90, 1.05)	1.02 (0.94, 1.10)	1.10 (1.03, 1.17)	
50%	0.90 (0.83, 0.97)	0.94 (0.87, 1.02)	1.02 (0.95, 1.11)	1.10 (1.03, 1.17)
70%	0.83 (0.76, 0.89)	0.87 (0.80, 0.94)	0.95 (0.88, 1.03)	1.03 (0.95, 1.11)

Hazard ratios (95% confidence interval) are presented

7.6.3 Impact of missing data

The landmark 1-year analysis was repeated by performing complete case analysis for outcomes target SU level and mortality. For target SU level, 52% (N=909) had complete data and for mortality 21% (N=3,609) had complete data.

Table 7.29 describes the distribution of covariates by treatment. In patients eligible for target SU level outcome analysis, allopurinol users were older, had higher prevalence of coronary heart disease, renal disease, overweight BMI, prescriptions for analgesics and colchicine, different BMI distribution, and had higher mean SU level than non-users where SMD >0.10. Compared with the main analysis, this analysis found SMD >0.10 for coronary heart disease, renal disease, and BMI that were previously below 0.10.

For patients eligible for mortality outcome analysis, in addition to the differences identified above, allopurinol users resided in more deprived areas, had higher prevalence of osteoarthritis, obesity, and were prescribed diuretics and NSAIDS more than non-users. This analysis found SMD >0.10 for age, osteoarthritis, and BMI that were not previously seen in the main analysis.

Table 7.29: Distribution of covariates by treatment for target SU level and mortality

	Target SU		Mortality	
	No allopurinol N=622	Allopurinol N=287	No allopurinol N=2,583	Allopurinol N=1,026
Demographics				
Age (Mean (SD))	59.1 (13.7)	61.3 (13.9)	61.3 (14.1)	63.7 (13.9)
Sex: Female	134 (22)	67 (23)	678 (26)	287 (28)
Deprivation (Mean (SD))	9.1 (5.4)	9.3 (6.0)	9.1 (5.5)	9.8 (5.8)
Comorbidities				
Anxiety	26 (4)	11 (4)	122 (5)	43 (4)
Depression	34 (5)	20 (7)	131 (5)	66 (6)
Cerebrovascular disease	12 (2)	10 (3)	62 (2)	28 (3)
Coronary heart disease	97 (16)	61 (21)	418 (16)	228 (22)
Diabetes	50 (8)	20 (7)	179 (7)	97 (9)
Hyperlipidaemia	40 (6)	21 (7)	168 (7)	65 (6)
Hypertension	168 (27)	75 (26)	625 (24)	275 (27)
Osteoarthritis	41 (7)	30 (10)	167 (6)	103 (10)
Peripheral vascular disease	8 (1)	7 (2)	35 (1)	25 (2)
Renal disease	10 (2)	11 (4)	29 (1)	30 (3)
Lifestyle factors				
Alcohol consumption				
Ever drinker	586 (94)	274 (95)	2399 (93)	960 (94)
Never drinker	36 (6)	13 (5)	184 (7)	66 (6)
Body mass index				
Normal	99 (16)	55 (19)	610 (24)	211 (21)
Overweight	313 (50)	128 (45)	1229 (48)	464 (45)
Obese	210 (34)	104 (36)	744 (29)	351 (34)
Smoking status				
Ever smoker	385 (62)	183 (64)	1619 (63)	654 (64)
Never smoker	237 (38)	104 (36)	964 (37)	372 (36)
SU level	486.0 (13.7)*	526.6 (83.5)*		
≤360µmol/L	-	-	485 (19)	17 (2)
>360µmol/L	-	-	2,098 (81)	1,009 (98)
Medication use				
Analgesics	219 (35)	126 (44)	829 (32)	465 (45)
Colchicine	21 (3)	17 (6)	52 (2)	36 (4)
Diuretics	244 (39)	123 (43)	1,002 (39)	519 (51)
NSAIDS	377 (61)	178 (62)	1,158 (45)	612 (60)

N (%) presented unless otherwise stated; Cells highlighted in yellow indicated standardised mean difference >0.10; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

After PS estimation, the mean and range of PS were similar between treatment groups indicating adequate common support. Five and six subclasses were required to achieve overall balance for all covariates across subclasses for the analyses of target SU level and mortality respectively; previously in the main analysis five subclasses were sufficient to achieve overall covariate balance for both outcomes.

Compared with the main analysis, more covariates were imbalanced within subclasses (Table 7.30). For example, for mortality, in the main analysis covariate imbalance was observed only in subclasses 1 and 2 whereas in this analysis, covariate imbalance was observed across all subclasses.

Table 7.30: Distribution of PS subclasses

Outcome	PS Mean (range) Allopurinol No allopurinol	Number of PS subclasses[a]	Smallest cell[b]	Imbalanced covariates
Target SU level	0.29 (0.09, 0.85) 0.37 (0.12, 0.90)	5	52/224 (23%)	S1: Age, deprivation, alcohol consumption, BMI, smoking status, SU level, analgesics, diuretics, NSAIDS S2: Renal disease, alcohol consumption, BMI, smoking status, colchicine, diuretics S3: Age, hyperlipidaemia, BMI, smoking status, SU level, analgesics, colchicine, diuretics S4: Coronary heart disease, renal disease, BMI, smoking status, diuretics S5: Deprivation, coronary heart disease, hyperlipidaemia, renal disease, smoking status, SU level, diuretics, NSAIDS
Mortality	0.35 (0.02, 0.75) 0.26 (0.01, 0.69)	6	30/1,026 (3%)	S1: Sex, deprivation, cerebrovascular disease, coronary heart disease, hypertension, peripheral vascular disease, alcohol consumption, BMI, smoking status, SU level, analgesics, colchicine, diuretics S2: Smoking status, analgesics S3: Smoking status, NSAIDS S4: Peripheral vascular disease, colchicine, NSAIDS S5: Deprivation, coronary heart disease S6: Smoking status

[a]Number of propensity score subclasses needed to achieve overall balance across subclasses; [b]Smallest cell defined as the number of allopurinol users within a subclass; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; S: Subclass; SU: Serum urate

Treatment effect estimated for target SU level and mortality is shown in Table 7.31. For target SU level, compared with the main analysis, the estimated treatment effect was lower (2.02 vs. 2.32), although the HR remained over two and statistical significance persisted. As expected, the standard error had increased due to decreased sample size (0.24 vs. 0.19).

Treatment effect estimated for mortality was also lower (1.07 vs. 1.10) and lost statistical significance which may be due to increased standard error from a reduced sample size (0.08 vs. 0.04)

Table 7.31: Treatment effect of allopurinol on outcome (1-year landmark analysis): complete case analysis

Outcome	No allopurinol N (%)	Allopurinol N (%)	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error
SU level				
Target level not met	398 (64)	126 (44)	2.15 (1.75, 2.64) 0.23	2.02 (1.60, 2.53) 0.24
Target level met	224 (36)	161 (56)		
Mortality				
Alive	1,953 (76)	698 (68)	1.36 (1.19, 1.55) 0.09	1.07 (0.93, 1.23) 0.08
Died	630 (24)	328 (32)		

CI: Confidence interval; SU: Serum urate

7.7 Summary

This chapter aimed to estimate the long-term effect of allopurinol, measured at baseline, on various outcomes, whilst adjusting for baseline covariates via PS subclassification. Allopurinol use had higher chance of reaching target SU level and fewer gout consultations in primary care, and had higher risk of premature mortality, gout hospitalisation, coronary heart disease and renal disease. The magnitude and direction of treatment effects persisted when extending the landmark period to two years. Treatment effect estimate for target SU level was robust to unmeasured confounding, and was in the same direction when complete case analysis was performed. In contrast, for mortality, treatment effect may potentially be protective of premature mortality in the presence of unmeasured confounding, and no association was observed in complete case analysis.

7.7.1 Comparison with published studies

Target SU level

As expected, allopurinol users were more likely to reach target SU level than non-users. This finding persisted within patients with no renal disease and regardless of severity of hyperuricaemia. It was found that SU levels were not frequently measured and patients who had SU level measured at baseline and during follow-up differed from the whole study sample.

No studies using large EHR databases have evaluated the effect of allopurinol on reaching target SU level however, few small non-EHR suboptimal studies had done so; general consensus across these studies were that allopurinol use (vs. no use) and increasing allopurinol dose were associated with reaching target (Dalbeth et al., 2006, Dalbeth et al., 2012, Pandya et al., 2011).

A randomised controlled trial (RCT) compared the efficacy of nurse-led care with usual general practitioner (GP)-led care on reaching primary outcome, target SU level ≤360µmol/L, with urate-lowering therapy (ULT). In nurse-led care, the majority of participants were treated with allopurinol and over 90% achieved target compared with only 30% in usual GP-led care (Doherty et al., 2018). Findings from this PhD found 57% of allopurinol users (comparable to the GP led-care) reached target which was higher than the 30% observed in that RCT. This may be due to differing covariate distributions of the study sample, for example, participants recruited to the RCT tended to be older, and had a higher prevalence of males, obesity, and comorbidities thus may have had more severe gout than the study sample from this PhD.

Mortality

Allopurinol use was found to be weakly associated with premature mortality and statistical significance was lost when results were stratified by presence of renal disease and severity of hyperuricaemia.

Kuo et al. (2015a) CPRD study had used the one-year landmark period, the same as this PhD. They found allopurinol use was not associated with mortality. Key differences between Kuo et al. (2015a) and this PhD are adjustment for more covariates within PS estimation, such the Charlson comorbidity index (Charlson et al., 1987), and prescriptions for aspirin and lipid lowering drugs; covariates being measured over a longer period of five years prior to the index date; the study sample was restricted to patients with a new diagnosis of gout; the definition of allopurinol use required a six-month prescription rather than three months; use of PS matching to create comparable treatment groups resulting in analysing a smaller study sample and larger standard error in the treatment effect; a lower prevalence of missing data in lifestyle factors. The results from this PhD also differed to a recent systematic review of four cohort EHR studies in gout that found no association between allopurinol use and mortality (Hay et al., 2020).

As stated above, there is the possibility that weak unmeasured confounding may nullify the treatment effect and thus, results would be similar to the treatment effects obtained from Kuo et al. (2015a) CPRD study and the systematic review.

Repeated gout consultations

Allopurinol users were less likely to consult for gout in primary care compared to non-users. Few case-cross over studies have shown allopurinol use was associated with less flares (Neogi et al., 2014, Zhang et al., 2012, Zhang et al., 2014). As allopurinol users were also more successful in lowering their SU levels to target, this had likely resulted in the long-term fall in the number of gout flares thus leading to fewer consultations.

A limitation of this analysis was that consultations occurring in the landmark period were ignored which may be problematic for patients initiating allopurinol who may have had more

consultations in the short term in monitoring SU level and titrating allopurinol dose which may have led to a stronger protective HR.

Doherty et al. (2018) RCT found within nurse-led care, the number of participants reporting two or more flares had reduced from 80% at baseline to 8% two years later, whilst in GP-led care it had reduced from 80% to 24%. This is the same conclusion obtained in this PhD, that the number of primary care gout consultations had reduced over time amongst allopurinol users.

Gout hospitalisation

Allopurinol was found to be associated with increased risk of gout hospitalisation. One possible reason could be that when patients initiate allopurinol or dose increases, they are at an increased short-term risk of flares due to partial crystal dissolution that may require hospital treatment. Over the long term, risk of flares (and hospitalisations) should decrease as allopurinol users reach target SU level and crystal dissolution is complete preventing occurrence of flares. Indeed, one small case-control study using hospital records in New Zealand found patients who were hospitalised were less likely to be on allopurinol and those who were on allopurinol had lower doses. Furthermore, colchicine prophylaxis (to prevent flares) was less likely to be used in hospitalised patients compared to non-hospitalised patients (Hutton et al., 2009). Similarly, a Swedish study found a small proportion of patients, between 19%-27%, received ULT in the six months prior to hospitalisation (Dehlin and Jacobsson, 2018).

Possible reasons why allopurinol users were at increased risk of hospitalisation may be due to suboptimal management of ULT and when allopurinol treatment was measured. In the UK, prophylactic treatment is often not prescribed and allopurinol dose is inadequate (due to failure in titrating allopurinol dose) to lower SU level to target (Roddy et al., 2007b, Cottrell et

al., 2013). As suggested by Hutton et al. (2009), these are factors for greater risk of hospitalisation. Patients with a prescription for allopurinol was only captured in the one-year landmark period. The analysis did not account for patients whose dose may have increased (potentially after the landmark period) thus increasing the risk of flares; patients who were prescribed allopurinol after the landmark period potentially attributing protective effects of allopurinol to non-users; nor accounted for prophylactic treatment.

Joint replacement

No association between allopurinol use and joint replacement at the hip or knee was found. Similar findings of no association between cumulative daily dose of ULT with joint replacement were observed within CPRD and the Taiwan National Health Insurance Database using a nested case-control study amongst incident gout (Kuo et al., 2018).

Gout is an independent risk factor for joint replacement as irreversible joint damage may occur from chronic gouty arthritis, tophi, and bone erosion. Osteoarthritis is the most common reason for patients undergoing joint replacement. It has been shown osteoarthritic joints are more susceptible to flares (Roddy et al., 2007a) and frequency of osteoarthritis is higher amongst patients diagnosed with gout than those without gout (Kuo et al., 2016b). Therefore, it was expected allopurinol use would lower SU levels to target and prevent continuing joint damage. However, the lack of association may be due to inadequate dose of allopurinol, failing to lower SU level to target (Roddy et al., 2007b, Cottrell et al., 2013), and non-adherence to allopurinol (Scheepers et al., 2018) which results in continuing flares and joint damage from crystal deposition.

Cerebrovascular disease

No association between allopurinol use and cerebrovascular disease was found in this thesis. The same observation was found in a matched cohort study using insurance claims from

Taiwan with a higher estimated HR of 1.18 (0.95, 1.47) (Kok et al., 2014). That study contained patients with higher prevalence of comorbidities such as hypertension, diabetes, and hyperlipidaemia, and used different methods to control for confounding by matching allopurinol users with non-users, and had considered other covariates such as atrial fibrillation, uraemia, and gastric ulcer.

Coronary heart disease

Allopurinol use was found to be increase the risk of coronary heart disease. Many studies have evaluated the risk of allopurinol on cardiovascular disease but definitions of outcome varied. An incident user study using American insurance claims records found compared to previous allopurinol users, current users were less likely to be hospitalized due to myocardial infarction or stroke in patients with both gout and diabetes (Singh et al., 2017). A matched cohort study using insurance claims from Taiwan found allopurinol users were at an increased risk of coronary heart disease (HR 1.41 (1.10, 1.79)) (Kok et al., 2014). Kok et al. (2014) estimated higher HR than this thesis which may be due to using a different data source and control for confounding variables. In contrast, a large population-based case-control study using a Spanish primary care database, and a cohort study using American insurance claims data had found allopurinol use, and longer duration of treatment (over 180 days) were protective of myocardial infarction (de Abajo et al., 2015, Singh and Yu, 2016).

Large RCTs have provided conflicting evidence of the effects of febuxostat vs. allopurinol on cardiovascular safety. The CARES trial had shown all-cause and cardiovascular mortality was higher amongst febuxostat users (White et al., 2018), whilst the FAST trial had shown there was no association (Mackenzie et al., 2020). The differences in study findings could be attributed to recruiting from different populations, febuxostat dose, and attrition bias. It is unclear why febuxostat increases cardiovascular risk and the observed effect in the CARES trial

may be due to allopurinol use having a protective effect (Choi et al., 2018). However, this PhD project found allopurinol use (vs. non-use) increased risk of mortality and coronary heart disease but did not examine effects of febuxostat. A limitation is that unmeasured confounding may be present. An ongoing ALL-HEART RCT will conclusively decide if allopurinol vs. non-use increases cardiovascular risk (Mackenzie et al., 2016).

Peripheral vascular disease

This thesis had shown there is no association between allopurinol use and peripheral vascular disease with the estimated HR close to null. A cohort study using American insurance claims found allopurinol use was protective against peripheral arterial disease (HR 0.85 (0.79, 0.93)) and those who were on allopurinol for >2 years the HR decreased to 0.75 (0.63, 0.89) compared to non-users (Singh and Cleveland, 2018a). Reasons for differing results could be due to using data collected in the USA; that study adjusted for confounding variables not considered in this thesis (race, Charlson-Romano score, beta blockers, ACE inhibitors, statins) but this thesis adjusted for more confounding variables overall.

Renal disease

This thesis found allopurinol use increased the risk of renal disease. Few studies have evaluated the effect of allopurinol on renal disease. Roughley et al. (2018) employed the one-year landmark method using data from CPRD. That study found no association between 6-month use of ULT (99% allopurinol prescription) with chronic kidney disease stage ≥3 (HR 1.09 (95% CI 0.99, 1.18)). That HR was smaller than the HR estimated in this thesis; this may be due to the definition of outcome as that study looked at a more severe renal disease than this thesis plus that study adjusted for more confounding variables (myocardial infarction, systemic lupus erythematosus, rheumatoid arthritis, heart failure, and hospitalisations). Furthermore, that study had higher prevalence of hypertension as they had used prescription

data, and higher prevalence of renal disease due to use of estimated glomerular filtration rate as opposed to using Read codes to define those comorbidities in this PhD; this suggest using Read codes were underused.

Vargas-Santos et al. (2018) used a time-stratified PS matched cohort study using data from THIN. They found allopurinol use of 300mg or more per day was protective against developing chronic renal disease stage ≥3 with HR of 0.87 (0.77, 0.97). Sensitivity analysis restricting patients not changing treatment status over time yielded lower HR (0.83 (0.72, 0.95). Strong assumptions on missing data were made as those with no serum creatinine or no Read code for chronic kidney disease stage 2 were considered to have normal kidney function. However, after using multiple imputation the protective effect of allopurinol was borderline statistically significant (0.92 (0.84, 1.00)). The difference in findings is because that study based allopurinol use on higher dose of 300mg which would expect to have more of an effect. In this thesis dose was not considered however given that a previous UK primary care study found 58% of patients remained on their starting dose of 100mg (Cottrell et al., 2013), it suggests a protective effect was not found due to suboptimal dosing.

7.7.2 Strengths and limitations

Simple strategies were utilised to allow estimation of the effect of allopurinol. Use of the landmark method was advantageous in that all patients were allowed a fixed period of time to be prescribed allopurinol that was not dependent upon length of follow-up time. PS subclassification is a straightforward and intuitive approach that divides patients into homogenous subclasses with patients having similar distribution of covariates and allows one to estimate subclass-class specific treatment effects. Despite the advantages of these approaches, some limitations were encountered.

Propensity score subclassification

PS subclassification performed fairly well ensuring overall balance of covariates between treatment groups across subclasses was satisfied. However, within subclasses, balance was not achieved on all covariates. Intuitively a greater number of subclasses would create more homogenous subclasses however this was not found in this analysis with imbalance remaining when up to 10 subclasses were used. Furthermore, one would expect more imbalanced covariates to be observed in subclasses containing a wider range of PS, however the opposite was found in this analysis.

PS subclassification was restricted by small sample size as seen when stratifying analyses by presence of renal disease. Only a small proportion of patients had renal disease and consequently outcome was infrequent. Dividing patients into PS subclasses was not possible as outcome had not occurred in some subclasses thus treatment effect could not be estimated. Where estimation was possible, corresponding standard errors were large due to small sample sizes. Treatment effect estimates were not comparable in patients with and without renal disease due to differing baseline covariates and sample size.

Despite these issues, overall PS subclassification estimated treatment effects that were comparable with previous studies.

Misclassification of treatment status

Determining landmark period is a balance between retaining sample size by not excluding too many patients with short follow-up and capturing patients prescribed allopurinol. A landmark period of one year was chosen for the main analysis with a sensitivity analysis extending the landmark period to two years. The one- and two- year landmark periods identified 3,957 and 4,513 allopurinol users, respectively. However, misclassification of treatment was present as a total of 7,767 patients were prescribed allopurinol at any time during follow-up. Potentially,

treatment effect estimates may be biased although treatment effect estimates from the one- and two- year landmark periods yielded similar HRs.

Infrequent recording of SU level

BSR guidelines state SU level should be monitored monthly during allopurinol titration and thereafter yearly once SU target has been obtained (Jordan et al., 2007b). Recording of SU level was low with nearly 60% of the study sample not having a measurement at baseline, although these patients had yet to start allopurinol which may be reason why SU was not measured. However, the number of patients having a follow-up SU measurement was low (29%). Recording of SU in practice is variable.

Defining gout hospitalisation

Gout hospitalisation was defined using inpatient Hospital Episodes Statistics data. Each hospitalisation may have up to 20 diagnoses recorded on the same date however, the primary reason for hospitalisation was not recorded. Although all patients had a diagnosis of gout within primary care, it cannot be differentiated if hospitalisation after diagnosis was attributed to gout or was recorded as a comorbidity within the reason for hospitalisation.

Residual confounding

This PhD had found allopurinol was associated with increased risk of many poor outcomes when it was expected that allopurinol may be protective or to observe no association. This may be due to residual confounding from unobserved covariates such as diet and severity of gout, but also from incomplete adjustment for covariates with missing data, and misclassification of comorbidities, for example basing hypertension solely on Read codes rather than prescription data for antihypertensives.

7.7.3 Conclusions

PS subclassification with the landmark analysis was a useful method to estimate the long-term effect of allopurinol on outcome. The greatest limitation was ignoring changes in treatment status after the landmark period. Therefore, advanced modelling strategies was required to model changes in allopurinol use over time. The next chapter accommodates this by estimating PS scores over time for time-varying allopurinol use and using PS subclassification to create homogenous subclasses prior to treatment effect estimation.

8 Effect of allopurinol: time-varying PS subclassification

8.1 Study sample

As shown in Chapter 7, the eligible study sample contained 16,876 patients. After dividing follow-up time into one-year intervals (or one-year follow-up periods), resulting in a maximum of 18 intervals per patient. Between all patients, there were a total of 155,331 intervals (16,876 baseline intervals and 138,544 follow-up intervals). Median (interquartile range (IQR)) number of follow-up intervals per patient was 11 (6, 14).

Table 8.1 describes how the distribution of covariates changed from baseline to the last available follow-up interval. Prevalence of comorbidities increased from baseline with the largest increases observed in hypertension (+31%), renal disease (+19%), and hyperlipidaemia (+18%). Lifestyle factors were more likely to be recorded over time with increased prevalence of ever drinkers (+24%), ever smokers (+28%), and across all body mass index (BMI) categories (6-10%). There was a small increase in prevalence of prescription for colchicine and analgesics but prescription for non-steroidal anti-inflammatory drugs (NSAIDS) had a small decrease. Appendix K describes the distribution of covariates over time in each year of follow-up.

Table 8.1: Covariates at baseline and end of follow-up (N=16,876)

Demographics	Baseline	End of follow-up
Age (Mean (SD))	62.1 (14.7)	70.3 (13.9)
Sex		
Male	12,995 (77)	12,995 (77)
Female	3,881 (23)	3,881 (23)
Deprivation (Mean (SD))	9.1 (5.5)	9.1 (5.5)
Comorbidities		
Anxiety	672 (4)	2,202 (13)
Depression	842 (5)	2,838 (17)
Cerebrovascular disease	407 (2)	1,636 (10)
Coronary heart disease	2,167 (13)	5,132 (30)
Diabetes	1,047 (6)	3,391 (20)
Hyperlipidaemia	783 (5)	3,923 (23)
Hypertension	3,137 (19)	8,370 (50)
Osteoarthritis	1,106 (7)	4,204 (25)
Peripheral vascular disease	257 (2)	826 (5)
Renal disease	217 (1)	3,449 (20)
Lifestyle factors		
Alcohol consumption		
Ever drinker	9,488 (56)	13,472 (80)
Never drinker	856 (5)	838 (5)
Missing	6,532 (39)	2,566 (15)
Body mass index		
Normal weight	2,517 (15)	3,483 (21)
Overweight	4,933 (29)	6,011 (36)
Obese	3,219 (19)	4,973 (29)
Missing	6,207 (37)	2,409 (14)
Smoking status		
Ever smoker	6,436 (38)	11,070 (66)
Never smoker	4,847 (29)	4,619 (27)
Missing	5,593 (33)	1,187 (7)
SU level		
≤360µmol/L	951 (6)	951 (6)
>360µmol/L	6,062 (36)	6,062 (36)
Missing	9,863 (58)	9,863 (58)
Medication use		
Analgesics	5,578 (33)	6,644 (39)
Colchicine	389 (2)	1,030 (6)
Diuretics	6,142 (36)	6,189 (37)
NSAIDS	8,024 (48)	5,881 (35)

Number (%) presented unless otherwise stated; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

8.2 Patterns of allopurinol use over time

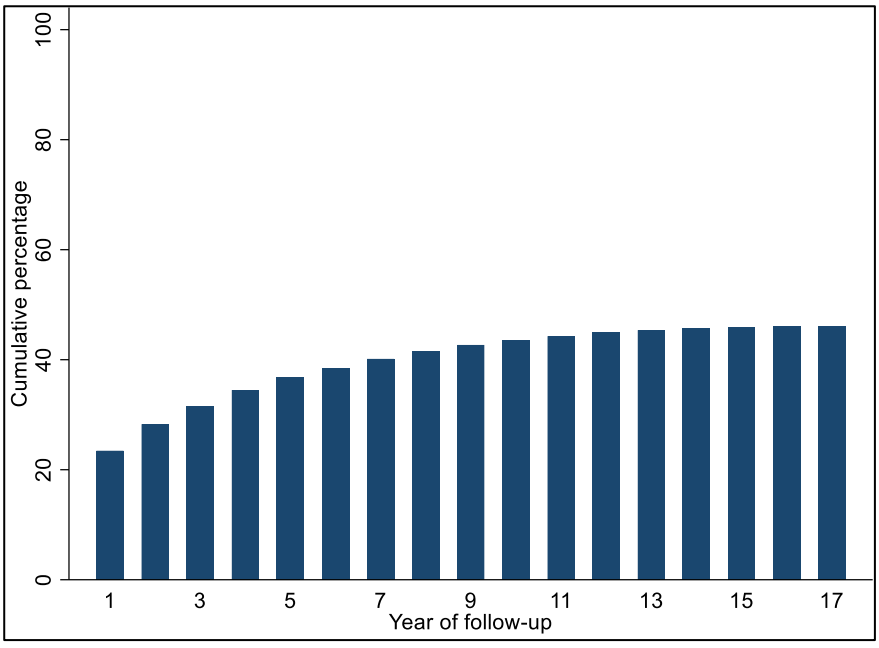
Table 8.2 shows the proportion of patients prescribed a total of three months prescription for allopurinol in each year of follow-up. The percentage of patients prescribed allopurinol was no higher than 44% in any year.

Table 8.2: Number of patients prescribed allopurinol over time

Follow-up year	No allopurinol N (%)	Allopurinol N (%)	Total
1	12,919 (77)	3,957 (23)	16,876
2	12,124 (77)	3,749 (24)	15,873
3	11,175 (75)	3,713 (25)	14,888
4	10,303 (73)	3,730 (27)	14,033
5	9,515 (72)	3,692 (28)	13,207
6	8,783 (71)	3,608 (29)	12,391
7	8,082 (69)	3,575 (31)	11,657
8	7,339 (68)	3,438 (32)	10,777
9	6,644 (67)	3,289 (33)	9,933
10	6,059 (66)	3,069 (34)	9,128
11	5,312 (65)	2,814 (35)	8,126
12	4,437 (64)	2,481 (36)	6,918
13	2,990 (63)	1,774 (37)	4,764
14	1,930 (61)	1,229 (39)	3,159
15	1,180 (58)	839 (42)	2,019
16	619 (562)	488 (44)	1,107
17	273 (57)	202 (43)	475

Overall, 7,767 (46%) patients initiated allopurinol and 9,109 (54%) were never prescribed allopurinol. Figure 8.1 illustrates the cumulative percentage of patients initiating treatment over time. 3,957 (23%) patients initiated allopurinol in the first year of follow-up. This cumulatively increased to 4,773 (28%) and 5,328 (32%) by the second and third year, respectively.

Figure 8.1: Cumulative percentage of patients prescribed allopurinol over time



Next it was explored whether patients were continuously prescribed allopurinol over time. Of all allopurinol users, 60% (n=4,696) remained on treatment until the end of follow-up and 40% (n=3,071) discontinued treatment. Of patients who discontinued treatment, 43% (n=1,341) resumed treatment. The majority of allopurinol users initiated treatment once (82%, n=6,426); and those stopping treatment had only stopped once (82%, n=2,507); a small proportion of patients repeatedly initiated and discontinued treatment (Table 8.3).

Table 8.3: Number of times patients repeatedly initiated and discontinued treatment

Number of times	Initiated allopurinol N=7,767 N (%)	Discontinued allopurinol N=3,071 N (%)
1	6,426 (83)	2,507 (82)
2	1,078 (14)	452 (15)
3	218 (3)	98 (3)
4	38 (<1)	11 (<1)
5	NA (<1)	NA (<1)
6	NA (<1)	NA (<1)

NA: Cannot report cell counts with less than five events

Table 8.4: Frequency and timing of allopurinol initiation

	Number of times allopurinol was initiated, N (%)					
Follow-up year	1: N=7,767	2: N=1,341	3: N=263	4: N=45	5: N=7	6: N=NA
1	3,957 (51)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2	816 (11)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
3	555 (7)	139 (10)	0 (0)	0 (0)	0 (0)	0 (0)
4	491 (6)	157 (12)	0 (0)	0 (0)	0 (0)	0 (0)
5	393 (5)	141 (11)	8 (3)	0 (0)	0 (0)	0 (0)
6	275 (4)	171 (13)	16 (6)	0 (0)	0 (0)	0 (0)
7	288 (4)	131 (10)	28 (11)	0 (0)	0 (0)	0 (0)
8	226 (3)	117 (9)	35 (13)	NA (2)	0 (0)	0 (0)
9	202 (3)	114 (9)	35 (13)	NA (4)	0 (0)	0 (0)
10	136 (2)	84 (6)	26 (9)	9 (20)	0 (0)	0 (0)
11	133 (2)	90 (7)	27 (10)	8 (18)	NA (29)	0 (0)
12	119 (2)	62 (5)	35 (13)	9 (20)	NA (14)	0 (0)
13	66 (1)	57 (4)	20 (8)	7 (16)	NA (14)	0 (0)
14	53 (1)	38 (3)	13 (5)	5 (11)	NA (43)	NA (33)
15	37 (<1)	20 (1)	13 (5)	NA (4)	0 (0)	0 (0)
16	17 (<1)	18 (1)	6 (2)	NA (2)	0 (0)	NA (33)
17	NA (<1)	NA (<1)	NA (<1)	NA (2)	0 (0)	NA (33)

NA: Cannot report cell counts with less than five events

As shown in Table 8.4, 51% (n=3,957) of allopurinol users were prescribed treatment for the first time in the first follow-up interval; the proportion of patients initiating allopurinol

dwindled over time. 54% (n=739) of allopurinol users who initiated treatment for the second time (after discontinuation) were between three and seven years of follow-up.

From Table 8.5, 25% (n=764) of allopurinol users discontinued treatment for the first time in the second year of follow-up. Between 5 and 11 years of follow-up, approximately 10% of patients discontinued allopurinol for the second time in each year of follow-up.

Table 8.5: Frequency and timing of allopurinol discontinuation

	Number of times allopurinol was discontinued, N (%)					
Follow-up year	1 N=3,071	2 N=564	3 N=112	4 N=14	5 N=NA	6 N=NA
2	764 (25)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
3	455 (15)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
4	351 (11)	31 (6)	0 (0)	0 (0)	0 (0)	0 (0)
5	285 (9)	62 (11)	0 (0)	0 (0)	0 (0)	0 (0)
6	238 (8)	79 (14)	NA (1)	0 (0)	0 (0)	0 (0)
7	195 (6)	69 (12)	10 (9)	0 (0)	0 (0)	0 (0)
8	183 (6)	57 (10)	10 (9)	0 (0)	0 (0)	0 (0)
9	160 (5)	56 (10)	18 (16)	0 (0)	0 (0)	0 (0)
10	132 (4)	60 (11)	17 (15)	NA (21)	0 (0)	0 (0)
11	97 (3)	59 (10)	18 (16)	NA (21)	0 (0)	0 (0)
12	88 (3)	36 (6)	11 (10)	NA (14)	NA (33)	0 (0)
13	54 (2)	23 (4)	13 (12)	NA (21)	0 (0)	0 (0)
14	30 (1)	16 (3)	5 (4)	NA (7)	0 (0)	0 (0)
15	23 (1)	8 (1)	8 (7)	NA (7)	NA (33)	0 (0)
16	11 (<1)	NA (1)	NA (1)	NA (7)	NA (33)	NA (100)
17	5 (<1)	5 (1)	0 (0)	0 (0)	0 (0)	0 (0)

NA: Cannot report cell counts with less than five events

The median (IQR) number of consecutive years patients were prescribed allopurinol for was five (2, 9) years.

8.3 Patient characteristics by allopurinol treatment

Table 8.6 describes and compares the distribution of covariates between intervals where patients were prescribed and not prescribed allopurinol.

Table 8.6: Distribution of covariates by treatment (N=16,876)

Covariates	Non-allopurinol intervals N=109,684	Allopurinol intervals N=45,647	SMD
Demographics			
Age (Mean (SD))	64.5 (14.0)	65.54 (13.29)	0.08
Sex: Female	23,908 (22)	8,346 (18)	-0.09
Deprivation (Mean (SD))	8.9 (5.5)	9.24 (5.61)	0.06
Comorbidities			
Anxiety	10,965 (10)	4,495 (10)	-0.01
Depression	13,440 (12)	5,448 (12)	-0.01
Cerebrovascular disease	5,861 (5)	2,868 (6)	0.04
Coronary heart disease	23,046 (21)	12,302 (27)	0.14
Diabetes	14,337 (13)	7,557 (17)	0.09
Gout consultation	13,832 (13)	9,338 (20)	0.21
Hyperlipidaemia	18,126 (17)	9,426 (21)	0.10
Hypertension	41,426 (38)	21,323 (47)	0.18
Osteoarthritis	18,931 (17)	9,592 (21)	0.09
Peripheral vascular disease	3,057 (3)	1,456 (3)	0.02
Renal disease	8,774 (8)	6,807 (15)	0.22
Lifestyle factors			
Alcohol consumption			
Ever drinker	81,194 (74)	35,485 (78)	0.09
Never drinker	5,044 (5)	1,911 (4)	-0.02
Missing	23,446 (21)	8,251 (18)	-0.08
Body mass index			
Normal	20,361 (19)	6,241 (14)	-0.13
Overweight	38,511 (35)	16,400 (36)	0.02
Obese	28,438 (26)	15,283 (33)	0.17
Missing	22,374 (20)	7,723 (17)	-0.09
Smoking status			
Ever smoker	63,245 (58)	28,277 (62)	0.09
Never smoker	31,793 (29)	12,477 (27)	-0.04
Missing	14,646 (13)	4,893 (11)	-0.08
SU level			
≤360µmol/L	8,468 (8)	637 (1)	-0.31
>360µmol/L	34,192 (31)	20,804 (46)	0.30
Missing	67,024 (61)	24,206 (53)	-0.16
Medication use			
Analgesics	32,443 (30)	16,550 (36)	0.14
Colchicine	4,398 (4)	3,681 (8)	0.17
Diuretics	29,088 (27)	16,073 (35)	0.19
NSAIDS	46,698 (43)	20,095 (44)	0.03
Cumulative allopurinol use, years (Mean (SD))	0.3 (1.2)	4.0 (3.5)	1.41

Number (%) presented unless otherwise stated; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SMD: Standardised mean difference; SU: Serum urate

Within intervals, large differences between the treatment groups were observed where standardised mean difference (SMD) was >0.10 (Table 8.6). Patients within allopurinol intervals had higher prevalence of coronary heart disease (SMD=0.14), gout consultation (0.21), hypertension (0.18), renal disease (0.22), obese (0.17), serum urate (SU) level above target (0.30), and had more prescriptions for analgesics (0.14), colchicine (0.17), and diuretics (0.19) than non-users within intervals. Allopurinol users were also previously on allopurinol for a longer period of time than non-users (4 years vs. 0.34 years).

8.4 Propensity score model and distribution

Covariates used to estimate propensity score (PS) for each outcome analysis are shown in Table 8.7. Covariates were included in PS estimation regardless of statistical significance: age, sex, deprivation, renal disease, colchicine, NSAIDS, diuretics, SU level, gout consultation, cumulative allopurinol use and follow-up time.

Covariates that were significantly associated with outcome, determined via the univariable complementary log-log regression model, were also included in the PS model. Gout consultation was only associated with outcomes gout hospitalisation, coronary heart disease, and renal disease. Generally, the majority of demographics, comorbidities, lifestyle factors and medication usage were associated with most outcomes. Cumulative allopurinol use and follow-up time was not associated with peripheral vascular disease.

Table 8.7: Covariates entered into the PS model for each outcome analysis

Outcomes	Target SU level N=1,742	Mortality N=16,876	Gout hospitalisation N=14,087	Joint replacement N=16,644	Cerebrovascular disease N=16,253	Coronary heart disease N=14,063	Peripheral vascular disease N=16,519	Renal disease N=16,508
Demographics								
Age	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Deprivation	X	X	X	X	X	X	X	X
Comorbidities								
Anxiety			X			X		
Depression		X				X		X
Cerebrovascular disease		X	X	X		X	X	X
Coronary heart disease	X	X	X	X	X		X	X
Diabetes	X	X	X		X	X	X	X
Gout consultation	X	X	X	X	X	X	X	X
Hyperlipidaemia	X		X		X	X	X	X
Hypertension		X	X	X	X	X	X	X
Osteoarthritis		X	X	X	X	X	X	X
Peripheral vascular disease		X	X		X	X		X
Renal disease	X	X	X	X	X	X	X	
Lifestyle factors								
Alcohol consumption	X	X	X	X	X	X	X	X
Body mass index	X	X	X	X	X	X	X	X
Smoking status	X	X	X	X	X	X	X	X
SU level	X	X	X	X	X	X	X	X
Medication use								
Analgesics	X	X	X	X	X	X	X	X
Colchicine	X	X	X	X	X	X	X	X
Diuretics	X	X	X	X	X	X	X	X
NSAIDS	X	X	X	X	X	X	X	X
Cumulative allopurinol use	X	X	X	X	X	X	X	X
Follow-up time	X	X	X	X	X	X	X	X

X: Covariate was entered into the propensity score model; Green cell: Covariate was associated with outcome ($p<0.05$); Red cell: Covariate was not associated with outcome ($p\geq0.05$); Black cell: Not applicable; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

The initial main effects PS model (specification 1) with linear terms for continuous covariates was fitted to estimate PS for each outcome analysis. For the analysis of target SU level, distribution of PS was skewed and differed between treatment groups (Figure 8.2); median (IQR) PS was 0.01 (0.01, 0.10) and 0.87 (0.55, 0.97) for non-allopurinol and allopurinol intervals, respectively. 53% (n=3,721) of non-allopurinol intervals and 12% (n=361) of allopurinol intervals (total 41%, n=4,082) were outside the PS region of common support (Table 8.8). Common support improved by considering fractional polynomial terms of dimension 2 (FP2) terms for follow-up time and FP1 terms for age, baseline SU level and cumulative allopurinol use (specification 4). Backwards selection identified baseline SU level as a potential problematic covariate; omitting this covariate reduced lack of common support from 21% to 15% (Table 8.9) however, this covariate was retained as adding two interactions to the PS model (colchicine*smoking status and BMI*gout consultation) improved common support by approximately the same amount (specification 6) (Table 8.8). Specification 6 was chosen as the final PS model as adding another interaction term to the PS model did not further improve common support. The distribution of the final estimated PS is shown in Figure 8.2. Despite the improvement in common support, the distribution of PS by treatment status was not similar between treatment groups.

Figure 8.2: Distribution of PS by treatment

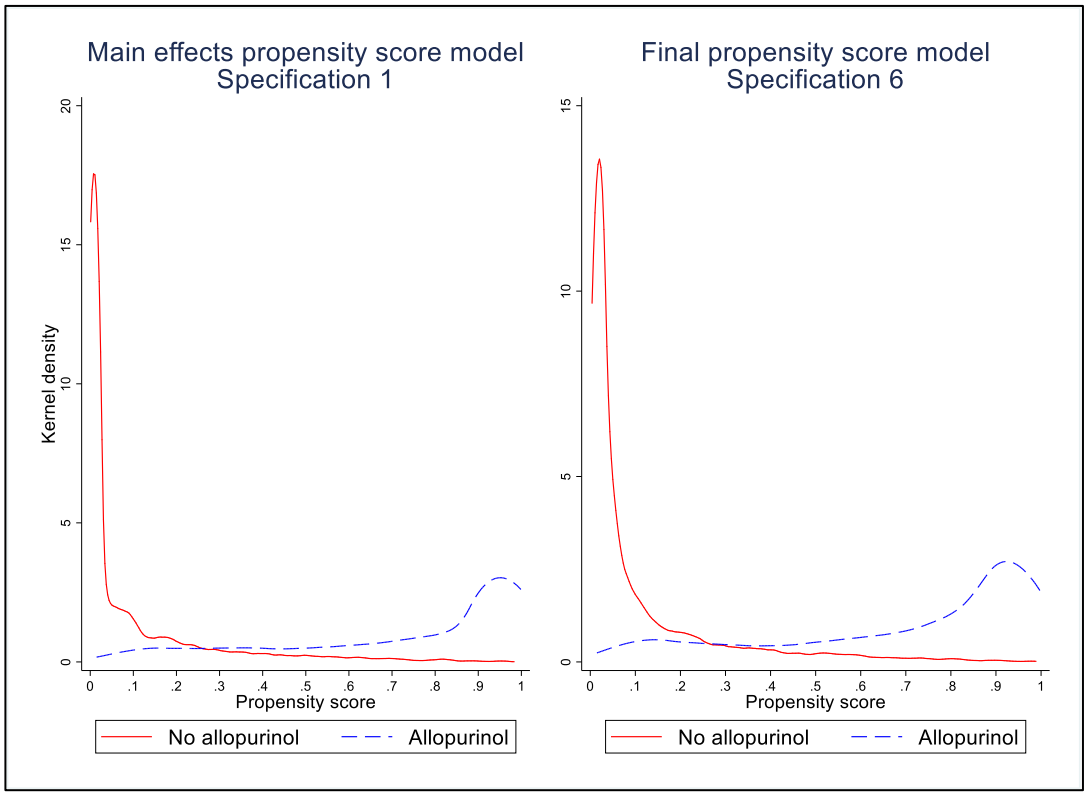


Table 8.8: PS model specification and degree of common support

	Outcome: Target SU level Propensity score model specification	Median propensity score (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=7,052	Allopurinol N=2,975	Overall N=10,027
1	Main effects model + linear terms for age, deprivation, SU level, cumulative allopurinol use, follow-up time	0.01 (0.01, 0.10) (<0.01, 0.98)	0.87 (0.55, 0.97) (0.02, 0.99)	3,721 (53%)	361 (12%)	4,082 (41%)
2	Main effects model + FP1 terms for age$^{(-2)}$, deprivation$^{(3)}$, SU level$^{(2)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(3)}$	0.02 (0.01, 0.11) (0.02, 0.98)	0.86 (0.55, 0.96) (0.02, 0.99)	3,341 (47%)	300 (10%)	3,641 (36%)
3	Main effects model + FP2 terms for age$^{(0.5,\ 3)}$, deprivation $^{(1,\ 2)}$, SU level$^{(-2,\ -2)}$, cumulative allopurinol use$^{(0.5,\ 3)}$, follow-up time$^{(-0.5,\ 3)}$	0.04 (0.02, 0.12) (<0.01, 0.95)	0.83 (0.52, 0.94) (0.02, 0.99)	1,754 (25%)	646 (22%)	2,400 (24%)
4	Main effects model + FP2 terms follow-up time$^{(-0.5,\ 3)}$ + FP1 terms for age$^{(-2)}$, SU level$^{(2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms deprivation	0.04 (0.02, 0.12) (<0.01, 0.99)	0.82 (0.52, 0.94) (0.02, 0.99)	2,062 (29%)	82 (3%)	2,144 (22%)
5	Main effects model + FP2 terms follow-up time$^{(-0.5,\ 3)}$ + FP1 terms for age$^{(-2)}$, SU level$^{(2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms deprivation + colchicine*smoking status	0.04 (0.02, 0.12) (<0.01, 0.99)	0.82 (0.52, 0.94) (0.02, 0.99)	1,803 (26%)	85 (3%)	1,888 (18%)
6	Main effects model + FP2 terms follow-up time$^{(-0.5,\ 3)}$ + FP1 terms for age$^{(-2)}$, SU level$^{(2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms deprivation + colchicine*smoking status + BMI*gout consultation	0.03 (0.02, 0.12) (<0.01, 0.99)	0.83 (0.51, 0.94) (0.01, 0.99)	1,405 (20%)	60 (2%)	1,465 (15%)

Table 8.8 continued:

7	Main effects model + FP2 terms follow-up time$^{(-0.5, 3)}$ + FP1 terms for age$^{(-2)}$, SU level$^{(2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms deprivation + colchicine*smoking status + BMI*gout consultation + colchicine use*gout consultation	0.04 (0.02, 0.12) (<0.01, 0.99)	0.83 (0.52, 0.94) (0.01, 0.99)	1,398 (22%)	55 (3%)	1,453 (16%)

Specification highlighted in green was the chosen propensity score model; Values in brackets (column 2) indicate which fractional polynomial terms were used; BMI: Body mass index; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score; SU: Serum urate

Similar findings were observed for secondary outcomes with relevant results tables presented in Appendix L. For all secondary outcome analyses, the initial main effects PS model (specification 1) highlighted the distribution of PS differed between treatment groups (illustrated in Appendix L). Approximately 50% of intervals were outside the PS region of common support. Common support improved considerably when FP terms were considered for continuous covariates: for the analysis of mortality, lack of common support reduced from 49% (based on specification 1) to 19% (based on specification 4); gout hospitalisation (from 54% to 19%); joint replacement (from 50% to 15%); peripheral vascular disease (from 50% to 15%); and renal disease (from 54% to 20%). There was a minor improvement in common support in the analysis of coronary heart disease when including FP terms for continuous covariates in the PS model however, no improvement was observed for cerebrovascular disease. Despite these improvements in common support, the distribution of PS remained skewed and were different between treatment groups.

Backwards selection did not identify a single covariate that considerably worsened common support. From Table 8.9, in the analysis for mortality, removing SU level from the PS model improved lack of common support from 19% to 16%. The largest improvement in common support was seen in the analysis for joint replacement, where lack of common support reduced from 15% to 6% when SU level was omitted from the PS model. At this stage, no covariates were omitted from the PS model as they were deemed strong confounding variables based on clinical grounds.

Inclusion of two-way interaction terms that were associated with outcome and improved common support, were included in the PS score model (Appendix L). For example, in the analysis for mortality, including four interaction terms (baseline SU level*NSAIDS, baseline SU level*hypertension, diuretics*sex, and alcohol consumption*sex) improved lack of common support from 19% to 7%. For the other secondary outcomes, up to three interaction terms

were included in the PS model with the exception of in the analyses of cerebrovascular disease and coronary heart disease, where no interaction terms were found to improve common support.

Although three-way interaction terms were considered in the PS model, none considerably improved common support more than using two-way interaction terms, thus they were no longer considered.

Table 8.10 shows which interaction terms and FP terms for continuous covariates were included in the final PS model for each outcome analysis. Table 8.11 summarises and compares the degree of common support from the main effects model (specification 1) compared with the final PS model, which shows despite improvements in lack of common support, the distribution of PS remained skewed and differed between treatment groups across all outcomes (also shown graphically in Appendix L).

Table 8.9: Backwards selection to identify problematic covariates

Outcome	Number of intervals outside the region of common support (PS model specification 4)	Number of intervals outside the region of common support after omitting problematic covariate from the PS model	Omitted covariate
SU level	2,144 (22%)	1,546 (15%)	Baseline SU level
Mortality	29,137 (19%)	24,565 (16%)	Baseline SU level
Gout hospitalisation	23,075 (19%)	20,459 (17%)	Colchicine
Joint replacement	22,340 (15%)	9,858 (7%)	Baseline SU level
Cerebrovascular disease	72,178 (50%)	72,199 (50%)	Follow-up time
Coronary heart disease	58,791 (50%)	58,490 (50%)	Follow-up time
Peripheral vascular disease	21,827 (15%)	24,574 (16%)	Sex
Renal disease	26,735 (20%)	16,096 (12%)	NSAIDS

PS: Propensity score; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

Table 8.10: Addition of fractional polynomial terms and interactions to the main effects PS model

Outcome	Fractional polynomial terms	Interaction terms
Target SU level	age$^{(-2)}$ Baseline SU level$^{(2)}$ cumulative allopurinol use$^{(0.5)}$ follow-up time$^{(-0.5, 3)}$	colchicine*smoking status colchicine*gout consultation
Mortality	age$^{(3, 3)}$ deprivation$^{(-2, -1)}$ cumulative allopurinol use$^{(0.5)}$	SU level*NSAIDS SU level*hypertension diuretics*sex alcohol consumption*sex
Gout hospitalisation	age$^{(-2\ 3)}$ deprivation $^{(3, 3)}$ cumulative allopurinol use$^{(-0.5, 0)}$ follow-up time$^{(0.5)}$	SU level*NSAIDS renal disease*gout consultation diuretics*sex
Joint replacement	age$^{(2, 3)}$ cumulative allopurinol use$^{(0.5)}$ follow-up time$^{(-2)}$	colchicine*coronary heart disease BMI*osteoarthritis alcohol consumption*gout consultation
Cerebrovascular disease	N/A	N/A
Coronary heart disease	age$^{(2, 2)}$ deprivation $^{(-2, -2)}$ follow-up time$^{(-2, -0.5)}$	N/A
Peripheral vascular disease	follow-up time$^{(-0.5, 3)}$ age$^{(-2)}$ cumulative allopurinol use$^{(0.5)}$	smoking status*diuretics
Renal disease	cumulative allopurinol use$^{(0.5)}$ follow-up time$^{(-2)}$	SU level*colchicine

BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

Table 8.11: Comparison of degree of common support between PS models

Outcome	Number of intervals outside the region of common support: Main effects PS model (Specification 1)			Number of intervals outside the region of common support: Final PS model			Distribution of propensity scores from the final PS model Median (IQR) (Range)	
	No allopurinol	Allopurinol	Overall	No allopurinol	Allopurinol	Overall	No allopurinol	Allopurinol
Target SU level	3,721 (53%)	361 (12%)	4,082 (41%)	1,405 (20%)	60 (2%)	1,465 (15%)	0.03 (0.02, 0.12) (0.003, 0.99)	0.83 (0.51, 0.94) (0.01, 0.99)
Mortality	76,177 (69%)	702 (2%)	76,879 (49%)	10,584 (10%)	683 (1%)	11,267 (7%)	0.01 (0.01, 0.06) ($2*10^{-4}$, 0.99)	0.92 (0.68, 0.98) (0.003, 0.99)
Gout hospitalisation	63,658 (72%)	2,026 (6%)	65,684 (54%)	7,048 (8%)	2,409 (7%)	9,457 (8%)	0.01 (0.01, 0.04) ($2*10^{-4}$, 0.99)	0.92 (0.69, 0.98) (0.003, 0.99)
Joint replacement	72,043 (67%)	2,542 (6%)	74,585 (50%)	15,979 (15%)	771 (2%)	16,750 (11%)	0.01 (0.007, 0.061) ($5*10^{-4}$, 0.99)	0.91 (0.69, 0.97) (0.005, 0.99)
Cerebrovascular disease	71,857 (70%)	1,069 (3%)	72,926 (50%)	70,628 (69)	756 (2%)	71,384 (49%)	0.003 (0.002, 0.05) ($1*10^{-5}$, 0.99)	0.93 (0.66, 0.99) (0.01, 0.99)
Coronary heart disease	58,716 (70%)	364 (1%)	59,080 (51%)	58,257 (69%)	534 (2%)	58,791 (50%)	0.003 (0.002, 0.05) ($7*10^{-5}$, 0.99)	0.92 (0.64, 0.98) (0.009, 0.99)
Peripheral vascular disease	74,245 (70%)	1,079 (2%)	75,324 (50%)	21,456 (20%)	371 (1%)	21,827 (15%)	0.02 (0.01, 0.07) ($6*10^{-4}$, 0.99)	0.91 (0.67, 0.97) (0.01, 0.99)
Renal disease	71,622 (73%)	1,871 (5%)	73,493 (54%)	21,211 (22%)	766 (2%)	21,977 (16%)	0.01 (0.01, 0.05) ($4*10^{-4}$, 0.99)	0.91 (0.68, 0.97) (0.01, 0.99)

IQR: Interquartile range; PS: Propensity score; SU: Serum urate

8.5 Number of propensity score subclasses

In the analysis of target SU level, four PS subclasses achieved overall covariate balance across subclasses, whilst increasing the number the subclasses (≥5) resulted in imbalance for cumulative allopurinol use and baseline SU level. As the PS distribution was skewed and differed between treatment groups, in the first subclass no outcome was observed within allopurinol intervals (Table 8.12 and Table 8.13). However, four PS subclasses were retained for treatment effect estimation as 75% of intervals from 99% of patients would be utilised, compared with say five subclasses, where a lower number of intervals (60%) from a lower number of patients (77%) would be utilised. Within subclasses, there were large differences in covariates between treatment groups where SMD >0.10; subclasses 2 and 4 had at least 11 imbalanced covariates whereas subclass 3 only had two (Table 8.15).

Within all secondary outcome analyses, use of four PS subclasses resulted in no occurrence of outcome within allopurinol intervals in subclass 1. Increasing the number of subclasses up to seven, generally resulted in fewer intervals and patients that would be utilised in treatment effect estimation (Appendix M). However, for some outcomes, a larger number of subclasses were required as overall covariate balance across subclasses was not achieved when using four PS subclasses. Choice of number of PS subclasses was based on achieving overall covariate balance across subclasses and maximising the number of intervals and patients that would be used in treatment effect estimation.

Four subclasses were sufficient to achieve overall covariate balance across subclasses for the analyses of mortality and joint replacement. For the analysis of renal disease, although overall covariate balance was achieved using five subclasses, six subclasses were chosen instead as a greater number of intervals and patients would be utilised in treatment effect estimation; similarly for all other outcome analyses, five subclasses achieved overall covariate balance and

would utilise a greater number intervals and patients in treatment effect estimation than using four subclasses, or that use of four subclasses did not achieve overall covariate balance. In the analysis of cerebrovascular disease, five PS subclasses were created however, outcome did not occur in subclasses 1 and 2 within allopurinol intervals. The distribution of outcome and treatment across PS subclasses is shown in Table 8.13. Table 8.14 shows the overall SMD for each covariate across subclasses, and that overall covariate balance was achieved.

Covariate imbalance within each subclass were identified and listed in Table 8.15. Generally, across all outcome analyses, the subclass with the lowest PS observed the most number of imbalanced covariates. The most frequently observed imbalanced covariates within subclasses across all outcomes were gout consultation (n=18), age (n=17), diuretics (n=15), colchicine (n=14), NSAIDS (n=12), and cumulative allopurinol use (n=11).

Table 8.12: Occurrence of outcome within PS subclasses

Propensity score range	Reached target SU level in non-allopurinol intervals N (%) N=247	Reached target SU level in allopurinol Intervals N (%) N=460	Number to be analysed in treatment effect analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates (SMD>0.10)[b]
4 subclasses				
1: <0.01, 0.02	81 (33)	0 (0)		
2: 0.02, 0.09	108 (44)	16 (3)	7,520 (75)	Overall balance
3: 0.09, 0.53	43 (17)	104 (23)	1,729 (99)	achieved
4: 0.53, 0.99	15 (6)	340 (74)		
5 subclasses				
1: <0.01, 0.02	58 (23)	0 (0)		
2: 0.02, 0.05	88 (36)	0 (0)	6,016 (60)	
3: 0.05, 0.18	66 (27)	53 (12)	1,339 (77)	Cumulative
4: 0.18, 0.69	26 (11)	142 (31)		allopurinol use
5: 0.69, 0.99	9 (4)	265 (58)		
6 subclasses				
1: <0.01, 0.02	49 (20)	0 (0)		
2: 0.02, 0.03	73 (30)	0 (0)		
3: 0.03, 0.09	67 (27)	16 (3)	6,684 (67)	SU level;
4: 0.09, 0.29	35 (14)	55 (12)	1,595 (92)	Cumulative
5: 0.29, 0.79	20 (8)	169 (37)		allopurinol use
6: 0.79, 0.99	3 (1)	220 (48)		
7 subclasses				
1: <0.01, 0.01	43 (17)	0 (0)		
2: 0.01, 0.02	53 (21)	0 (0)		
3: 0.02, 0.05	60 (24)	0 (0)	5,729 (57)	SU level;
4: 0.05, 0.15	51 (21)	44 (10)	1,236 (71)	Cumulative
5: 0.15, 0.42	23 (9)	42 (9)		allopurinol use
6: 0.42, 0.85	14 (6)	183 (40)		
7: 0.85, 0.99	3 (1)	191 (42)		

Propensity score subclassification was performed on 10,027 intervals from 1,742 patients. Results highlighted in green indicated the number of subclasses used for treatment effect estimation. [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD was assessed in subclasses with outcome occurring in both non-allopurinol and allopurinol intervals; PS: Propensity score; SMD: Standardised mean difference; SU: Serum urate

Table 8.13: Distribution of outcome and treatment across PS subclasses

Outcome	Non-allopurinol intervals N(%)			Allopurinol intervals N (%)		
Target SU level	SU target not met: N=6,805	SU target met: N=247	Total: N=7,052	SU target not met: N=2,515	SU target met: N=460	Total: N=2,975
Subclass 1	2,424 (36)	81 (33)	2,505 (36)	2 (0)	0 (0)	2 (0)
Subclass 2	2,305 (34)	108 (44)	2,413 (34)	78 (3)	16 (3)	94 (3)
Subclass 3	1,791 (26)	43 (17)	1,834 (26)	569 (23)	104 (23)	673 (23)
Subclass 4	285 (4)	15 (6)	300 (4)	1,866 (74)	340 (74)	2,206 (74)
Mortality	Alive: N=106,338	Died: N=3,346	Total: N=109,684	Alive: N=44,017	Died: N=1,630	Total: N=45,647
Subclass 1	37,799 (36)	1,029 (31)	38,828 (35)	5 (0)	0 (0)	5 (0)
Subclass 2	36,748 (35)	1,311 (39)	38,059 (35)	761 (2)	13 (1)	774 (2)
Subclass 3	29,094 (27)	857 (26)	29,951 (27)	8,537 (19)	345 (21)	8,882 (19)
Subclass 4	2697 (3)	149 (4)	2,846 (3)	34,714 (79)	1,272 (78)	35,986 (79)
Gout hospitalisation	No: N=87,694	Yes: N=923	Total: N=88,617	No: N=32,402	Yes: N=1,049	Total: N=33,451
Subclass 1	24,268 (23)	143 (4)	24,411 (22)	3 (0)	0 (0)	3 (0)
Subclass 2	24,062 (23)	231 (7)	24,293 (22)	120 (0)	1 (0)	121 (0)
Subclass 3	22,441 (21)	341 (10)	22,782 (21)	1,589 (4)	42 (3)	1,631 (4)
Subclass 4	16,009 (15)	169 (5)	16,178 (15)	7,999 (18)	237 (15)	8,236 (18)
Subclass 5	914 (1)	39 (1)	953 (1)	22,691 (52)	769 (47)	23,460 (51)
Joint replacement	No: N=104,135	Yes: N=692	Total: N=104,827	No: N=42,731	Yes: N=323	Total: N=43,054
Subclass 1	36,770 (35)	194 (28)	36,964 (35)	7 (<1)	0 (0)	7 (<1)
Subclass 2	36,030 (35)	311 (45)	36,341 (35)	628 (1)	1 (<1)	629 (1)
Subclass 3	28,456 (27)	161 (23)	28,617 (27)	8,319 (19)	34 (11)	8,353 (19)
Subclass 4	2,879 (3)	26 (4)	2,905 (3)	33,777 (79)	288 (89)	34,065 (79)

Table 8.13 continued:

	Non-allopurinol intervals N (%)			Allopurinol intervals N (%)		
Cerebrovascular disease	No: N=101,655	Yes: N=1,114	Total: N=102,769	No: N=41,901	Yes: N=438	Total: N=42,339
Subclass 1	28,918 (28)	104 (9)	29,022 (28)	0 (0)	0 (0)	0 (0)
Subclass 2	28,689 (28)	333 (30)	29,022 (28)	0 (0)	0 (0)	0 (0)
Subclass 3	26,597 (26)	594 (53)	27,191 (26)	1,822 (4)	8 (2)	1,830 (4)
Subclass 4	16,356 (16)	69 (6)	16,425 (16)	12,494 (30)	103 (24)	12,597 (30)
Subclass 5	1,095 (1)	14 (1)	1,109 (1)	27,585 (66)	327 (75)	27,912 (66)
Coronary heart disease	No: N=82,268	Yes: N=2,080	Total: N=84,348	No: N=31,438	Yes: N=870	Total: N=32,308
Subclass 1	23,132 (28)	200 (10)	23,332 (28)	0 (0)	0 (0)	0 (0)
Subclass 2	22,801 (28)	530 (25)	23,331 (28)	0 (0)	0 (0)	0 (0)
Subclass 3	21,044 (26)	1,168 (56)	22,212 (26)	1,107 (4)	12 (1)	1,119 (3)
Subclass 4	14,086 (17)	143 (7)	14,229 (17)	8,916 (28)	186 (21)	9,102 (28)
Subclass 5	1,205 (1)	39 (2)	1,244 (1)	21,415 (68)	672 (77)	22,087 (68)
Peripheral vascular disease	No: N=105,741	Yes: N=432	Total: N=106,173	No: N=43,775	Yes: N=203	Total: N=43,978
Subclass 1	29,936 (28)	93 (22)	30,029 (28)	2 (0)	0 (0)	2 (0)
Subclass 2	29,674 (28)	141 (33)	29,815 (28)	214 (0)	1 (0)	215 (0)
Subclass 3	27,411 (26)	153 (35)	27,564 (26)	2,454 (6)	12 (6)	2,466 (6)
Subclass 4	17,503 (17)	40 (9)	17,543 (17)	12,437 (28)	50 (25)	12,487 (28)
Subclass 5	1,217 (1)	5 (1)	1,222 (1)	28,668 (65)	140 (69)	28,808 (66)
Renal disease	No: N=96,077	Yes: N=2,541	Total: N=98,618	No: N=35,986	Yes: N=1,384	Total: N=37,370
Subclass 1	22,210 (23)	454 (18)	22,664 (23)	1 (0)	0 (0)	1 (0)
Subclass 2	21,937 (23)	712 (28)	22,649 (23)	16 (0)	0 (0)	16 (0)
Subclass 3	21,530 (22)	800 (31)	22,330 (23)	328 (1)	6 (0)	334 (1)
Subclass 4	19,378 (20)	311 (12)	19,689 (20)	2,888 (8)	88 (6)	2,976 (8)
Subclass 5	10,417 (11)	227 (9)	10,644 (11)	11,676 (32)	345 (25)	12,021 (32)
Subclass 6	605 (1)	37 (1)	642 (1)	21,077 (59)	945 (68)	22,022 (59)

PS: Propensity score; SU: Serum urate

Table 8.14: Overall SMD after PS subclassification

	Target SU level	Mortality	Gout hospitalisation	Joint replacement	Cerebrovascular disease	Coronary heart disease	Peripheral vascular disease	Renal disease
Demographics								
Age	-0.04	0.01[a]	-0.01[a]	0.01[a]	0.01	0.03[a]	-0.03	0.06
Sex: Female	0.01	0.02	0.03	0.02	0.01	0.01	0.02	0.036
Deprivation	-0.02	-0.01[a]	<0.01	<0.01	-0.01	<0.01	<0.01	-0.01
Comorbidities								
Anxiety	-	-	0.01	-	-	0.02	-	-
Depression	-	0.01	-	-	-	0.02	-	0.01
Cerebrovascular disease	-	0.01	0.01	0.01	-	0.01	0.02	0.01
Coronary heart disease	0.01	<0.01	<0.01	0.01	<0.01	-	0.01	0.02
Diabetes	0.02	0.01	<0.01	-	0.01	0.02	0.01	0.02
Gout consultation	-0.03	0.01	0.01	0.02	0.02	0.01	-0.01	-0.03
Hyperlipidaemia	0.02	-	<0.01	-	0.01	0.01	0.01	0.02
Hypertension	-	0.01	-0.01	0.01	0.02	0.02	0.01	0.02
Osteoarthritis	-	<0.01	0.01	0.01	0.01	0.02	0.02	0.02
Peripheral vascular disease	-	<0.01	<0.01	-	0.01	0.01	-	0.01
Renal disease	<0.01	<0.01	<0.01	0.01	0.02	0.01	0.02	-
Lifestyle factors								
Alcohol consumption[a]	-0.02	0.02	0.01	0.02	-0.03	-0.03	-0.01	-0.02
Body mass index[a]	-0.03	0.01	0.02	-0.02	-0.03	-0.03	-0.02	-0.02
Smoking status[a]	0.04	<0.01	-0.01	-0.01	-0.03	-0.04	0.02	-0.02
SU level[a]	-0.07[b]	0.04	-0.02	0.04	-0.05	-0.04	0.04	0.04
Medication use								
Analgesics	-0.02	<0.01	0.02	-0.01	<0.01	0.01	0.03	0.02
Colchicine	-0.03	<0.01	0.01	-0.01	0.01	<0.01	<0.01	-0.01
Diuretics	0.01	0.01	0.01	-0.01	-0.02	<0.01	0.02	0.01
NSAIDS	-0.05	<0.01	-0.01	-0.02	-0.02	<0.01	-0.03	-0.04
Cumulative allopurinol use	-0.05	0.05	-0.04[a]	0.08	0.08	0.08	0.02	-0.05
Interaction terms	-0.06[a]	-0.03[a]	0.02	0.01	-	-	0.02	-0.01

[a]Largest standardised difference presented; [b]Continuous covariate; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SMD: Standardised mean difference; SU: Serum urate

Table 8.15: Imbalanced covariates within each PS subclass

Outcome	Imbalanced covariates (SMD >0.10)
Target SU level	S2: Age, coronary heart disease, hyperlipidaemia, alcohol consumption, BMI, smoking status, SU level, analgesics, diuretics, NSAIDS, cumulative allopurinol use, colchicine*gout consultation S3: Smoking status, cumulative allopurinol use S4: Age, gout consultation, BMI, SU level, analgesics, colchicine, diuretics, NSAIDS, cumulative allopurinol use, colchicine*smoking status, colchicine*gout consultation
Mortality	S2: Age, cerebrovascular disease, coronary heart disease, gout consultation, hypertension, analgesics, colchicine, diuretics, NSAIDS, cumulative allopurinol use, SU level*NSAIDS, diuretics*sex S3: Gout consultation S4: Age, gout consultation, colchicine, diuretics, NSAIDS, SU level*NSAIDS
Gout hospitalisation	S2: Age, deprivation, cerebrovascular disease, coronary heart disease, gout consultation, hypertension, peripheral vascular disease renal disease, analgesics, colchicine, diuretics, alcohol consumption, SU level, SU level*NSAIDS, diuretics*sex S3: Age, coronary heart disease, diuretics, cumulative allopurinol use S5: Age, diabetes, gout consultation, renal disease, colchicine, NSAIDS, SU level*NSAIDS, renal disease*gout consultation
Joint replacement	S2: Age, sex, cerebrovascular disease, coronary heart disease, gout consultation, osteoarthritis, analgesics, diuretics, NSAIDS, alcohol consumption, SU level S3: Gout consultation S4: Age, gout consultation, colchicine, NSAIDS
Cerebrovascular disease	S3: Age, coronary heart disease, diabetes, gout consultation, hyperlipidaemia, hypertension, osteoarthritis, renal disease, analgesics, colchicine, diuretics, alcohol consumption, smoking status, SU level, BMI S4: Gout consultation, smoking status, cumulative allopurinol use S5: Gout consultation, osteoarthritis, renal disease, colchicine, NSAIDS, cumulative allopurinol use, smoking status
Coronary heart disease	S3: Age, cerebrovascular disease, diabetes, gout consultation, hypertension, hyperlipidaemia, osteoarthritis, renal disease, analgesics, colchicine, diuretics, smoking status, SU level, BMI S4: Cumulative allopurinol use S5: Age, gout consultation, colchicine, diuretics, NSAIDS
Peripheral vascular disease	S3: Sex, deprivation, cerebrovascular disease, coronary heart disease, gout consultation, hyperlipidaemia, analgesics, colchicine, diuretics, NSAIDS, alcohol consumption, smoking status, SU level, BMI, smoking status*diuretics S4: Cumulative allopurinol use S5: Age, diabetes, gout consultation, renal disease, colchicine, diuretics, NSAIDS, cumulative allopurinol use
Renal disease	S3: Age, depression, cerebrovascular disease, coronary heart disease, gout consultation, analgesics, colchicine, diuretics, NSAIDS, alcohol consumption, BMI, SU level, smoking status, SU level*colchicine, S4: Age, diuretics, cumulative allopurinol use S5: Age S6: Age, diabetes, gout consultation, colchicine, diuretics, NSAIDS, smoking status, cumulative allopurinol use, SU level*colchicine

BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; S: Subclass; SMD: Standardised mean difference; SU: Serum urate

8.6 Treatment effect estimation

Originally, the complementary log-log model with a random intercept was used to estimate subclass-specific treatment effects however, not all models had converged, due to adjustment for follow-up time in the model or certain imbalanced covariates. Therefore, the random intercept was omitted, and robust standard errors accounting for clustering of repeated measurements from patients was utilised.

For half the outcomes (target SU level, mortality, joint replacement, peripheral vascular disease, the subclass-specific treatment effect estimates were homogenous. Allopurinol was associated with greater chance of reaching target SU level (4.89 (3.76, 6.37)) and increased risk of peripheral vascular disease (2.04 (1.39, 2.99)), however no association was observed for mortality and joint replacement (Table 8.16).

The subclass-specific treatment effect estimates for gout hospitalisation (Table 8.17), cerebrovascular disease (Table 8.18), coronary heart disease (Table 8.19), and renal disease (Table 8.20) were not homogenous, therefore are presented individually alongside with a summary of covariates.

Allopurinol increased the risk of gout hospitalisation in subclasses 3 and 4. Patients in subclass 4 had a larger treatment effect estimate (2.70 (2.22, 3.29)), and these patients in these intervals had fewer prescriptions for diuretics and was previously on allopurinol for a longer period than those in subclass 3 who had a lower treatment effect (1.84 (1.34, 2.51)).

Within subclass 3, allopurinol was protective against cerebrovascular disease (0.21 (0.11, 0.44)) and coronary heart disease (0.21 (0.11, 0.38)), in contrast with subclass 4 where allopurinol had increased risk of cerebrovascular disease (1.73 (1.26, 1.38)) and coronary heart disease 1.99 (1.58,

2.50)). In both outcomes, the distribution of covariates between subclasses 3 and 4 were similar but intervals in subclass 3 were older and had higher prevalence of SU level >360µmol/L, prescription for diuretics, and patients were previously on allopurinol for a shorter period of time than subclass 4.

Allopurinol use increased the risk of renal disease in subclass 4 (2.03 (1.61, 1.34)) whereas its use was protective in subclass 6 (0.58 (0.42, 0.81)). Subclass 6 contained intervals from patients who were older, resided in less deprived areas, and had higher prevalence of coronary heart disease, hypertension, ever smokers, ever drinkers, obesity, and fewer prescriptions for NSAIDS, and previously prescribed allopurinol for a longer period than intervals in subclass 4.

Table 8.16: Effect of allopurinol on outcome

	Non-allopurinol intervals N (%)	Allopurinol intervals N (%)	Unadjusted Hazard ratio (95% CI) Standard error	PS subclassification Hazard ratio (95% CI) Standard error
Target SU level				
Not reached target	4,381 (96)	2,513 (85)	4.57 (3.80, 5.48)	4.89 (3.76, 6.37)
Reached target	166 (4)	460 (15)	0.43	0.66
Mortality				
Alive	68,539 (97)	44,012 (96)	1.15 (1.08, 1.23)	0.97 (0.88, 1.08)
Death	2,317 (3)	1,630 (4)	0.04	0.05
Gout hospitalisation				
No	63,426 (99)	32,399 (97)	2.21 (2.01, 2.44)	*
Yes	780 (1)	1,049 (3)	0.11	
Joint replacement				
No	67,365 (99)	42,724 (99)	1.01 (0.87, 1.16)	0.77 (0.59, 1.01)
Yes	498 (1)	323 (1)	0.07	0.11
Cerebrovascular disease				
No	44,048 (98)	41,901 (99)	0.73 (0.65, 0.83)	*
Yes	677 (2)	438 (1)	0.05	
Coronary heart disease				
No	36,335 (96)	31,438 (97)	0.85 (0.78, 0.92)	*
Yes	1,350 (4)	870 (3)	0.04	
Peripheral vascular disease				
No	75,805 (99)	43,773 (99)	1.16 (0.97, 1.38)	2.04 (1.39, 2.99)
Yes	339 (1)	203 (1)	0.11	0.40
Renal disease				
No	51,930 (97)	35,969 (96)	1.21 (1.12, 1.31)	*
Yes	1,375 (3)	1,384 (4)	0.05	

*Subclass-specific treatment effects were not homogenous. Therefore they were not pooled; CI: Confidence interval; SU: Serum urate

Table 8.17: Estimated treatment effect of allopurinol on gout hospitalisation and distribution of covariates within each PS subclass

	Subclass 2 N=24,414	Subclass 3 N=24,413	Subclass 4 N=24,414	Subclass 5 N=24,413
Hazard ratio (95% CI)	0.69 (0.11, 4.20)	1.84 (1.34, 2.51)	2.70 (2.22, 3.29)	0.94 (0.61, 1.16)
Standard error	0.64	0.29	0.27	0.14
Demographics				
Age*	66.1 (13.8)	65.9 (13.7)	63.8 (13.6)	67.2 (12.6)
Sex: Female	5,149 (21)	4,960 (20)	4,291 (18)	4,888 (20)
Deprivation*	9.1 (5.5)	9.2 (5.5)	9.3 (5.5)	9.3 (5.5)
Comorbidities				
Anxiety	2,241 (9)	1,712 (7)	2,215 (9)	2,495 (10)
Depression	2,914 (12)	2,218 (9)	2,638 (11)	3,123 (13)
Cerebrovascular disease	1,610 (7)	1,127 (5)	1,045 (4)	1,698 (7)
Coronary heart disease	5,948 (24)	5,324 (22)	5,016 (21)	7,630 (31)
Diabetes	3,374 (14)	2,550 (10)	2,888 (12)	4,508 (18)
Gout consultation	2,147 (9)	4,510 (18)	5,876 (24)	4,729 (19)
Hyperlipidaemia	4,132 (17)	3,000 (12)	3,879 (16)	5,411 (22)
Hypertension	9,677 (40)	8,580 (35)	9,317 (38)	12,395 (51)
Osteoarthritis	4,360 (18)	3,521 (14)	4,356 (18)	5,666 (23)
Peripheral vascular disease	759 (3)	613 (3)	575 (2)	901 (4)
Renal disease	1,981 (8)	1,447 (6)	2,122 (9)	4,094 (17)
Lifestyle factors				
Alcohol consumption				
Ever drinker	18,442 (76)	16,246 (67)	17,998 (74)	19,318 (79)
Never drinker	1,207 (5)	1,138 (5)	974 (4)	1,144 (5)
Missing	4,765 (20)	7,029 (29)	5,442 (22)	3,951 (16)
Body mass index				
Normal	4,467 (18)	3,301 (14)	3,394 (14)	3,310 (14)
Overweight	8,714 (36)	8,296 (34)	8,784 (36)	9,137 (37)
Obese	6,578 (27)	6,151 (25)	7,035 (29)	8,522 (35)
Missing	4,655 (19)	6,665 (27)	5,201 (21)	3,444 (14)
Smoking status				
Ever smoker	14,598 (60)	12,525 (51)	13,648 (56)	15,920 (65)
Never smoker	6,968 (29)	6,671 (27)	6,974 (29)	6,276 (26)
Missing	2,848 (12)	5,217 (21)	3,792 (16)	2,217 (9)
SU level				
≤360µmol/L	304 (1)	248 (1)	388 (2)	294 (1)
>360µmol/L	7,268 (30)	10,535 (43)	10,422 (43)	11,575 (47)
Missing	16,842 (69)	13,630 (56)	13,604 (56)	12,544 (51)
Medication use				
Analgesics	7,569 (31)	8,645 (35)	8,172 (33)	9,470 (39)
Colchicine	469 (2)	1,585 (6)	2105 (9)	1862 (8)
Diuretics	7,316 (30)	10,049 (41)	7,205 (30)	9,528 (39)
NSAIDS	9,667 (40)	14,156 (58)	13,767 (56)	9,776 (40)
Cumulative allopurinol use*	0 (0)	0.1 (0.3)	1.21 (1.65)	4.9 (3.3)

Number and percentage presented unless otherwise stated; *Mean (Standard deviation) presented for continuous covariates; Unit of analysis is intervals; CI: Confidence interval; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

Table 8.18: Estimated treatment effect of allopurinol on cerebrovascular disease and distribution of covariates within each PS subclass

	Subclass 3 N=29,021	Subclass 4 N=29,022	Subclass 5 N=29,021
Hazard ratio (95% CI)	0.21 (0.11, 0.44)	1.73 (1.26, 1.38)	0.80 (0.47, 1.37)
Standard error	0.08	0.28	0.22
Demographics			
Age*	65.1 (14.1)	61.8 (13.7)	66.1 (12.8)
Sex: Female	5,099 (18)	4,176 (14)	5,564 (19)
Deprivation*	9.3 (5.5)	9.1 (5.6)	9.3 (5.6)
Comorbidities			
Anxiety	2,539 (9)	2,598 (9)	2,789 (10)
Depression	3,122 (11)	3,153 (11)	3,456 (12)
Coronary heart disease	6,880 (24)	5,620 (19)	8,159 (28)
Diabetes	3,791 (13)	3,403 (12)	5,091 (18)
Gout consultation	6,612 (23)	8,027 (28)	4,981 (17)
Hyperlipidaemia	4,769 (16)	4,803 (17)	6,240 (22)
Hypertension	11,728 (40)	11,336 (39)	14,045 (48)
Osteoarthritis	5,131 (18)	5,101 (18)	6,452 (22)
Peripheral vascular disease	772 (3)	511 (2)	890 (3)
Renal disease	2,768 (10)	2,815 (10)	4,516 (16)
Lifestyle factors			
Alcohol consumption			
Ever drinker	21,221 (73)	22,248 (77)	22,551 (78)
Never drinker	1,144 (4)	1,003 (3)	1,293 (4)
Missing	6,656 (23)	5,771 (20)	5,177 (18)
Body mass index			
Normal	3,987 (14)	3,901 (13)	3,762 (13)
Overweight	10,334 (36)	10,464 (36)	10,331 (36)
Obese	8,350 (29)	8,966 (31)	10,230 (35)
Missing	6,350 (22)	5,691 (20)	4,698 (16)
Smoking status			
Ever smoker	16,605 (57)	16,582 (57)	18,240 (63)
Never smoker	8,230 (28)	8,608 (30)	7,815 (27)
Missing	4,186 (14)	3,832 (13)	2,966 (10)
SU level			
≤360µmol/L	356 (1)	412 (1)	363 (1)
>360µmol/L	14,712 (51)	12,022 (41)	13,859 (48)
Missing	13,953 (48)	16,588 (57)	14,799 (51)
Medication use			
Analgesics	9,530 (33)	8,695 (30)	10,909 (38)
Colchicine	2,005 (7)	2,840 (10)	2,080 (7)
Diuretics	10,284 (35)	7,378 (25)	10,938 (38)
NSAIDS	15,858 (55)	16,510 (57)	11,429 (39)
Cumulative allopurinol use*	0.2 (0.5)	1.5 (1.7)	5.2 (3.5)

Number and percentage presented unless otherwise stated; *Mean (Standard deviation) presented for continuous covariates; Unit of analysis is intervals; CI: Confidence interval; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

Table 8.19: Estimated treatment effect of allopurinol on coronary heart disease and distribution of covariates within each PS subclass

	Subclass 3 N=23,331	Subclass 4 N=23,331	Subclass 5 N=23,331
Hazard ratio (95% CI)	0.21 (0.11, 0.38)	1.99 (1.58, 2.50)	0.84 (0.61, 1.19)
Standard error	0.06	0.23	0.14
Demographics			
Age*	63.4 (13.9)	60.1 (13.8)	64.0 (13.0)
Sex: Female	3,946 (17)	3,099 (13)	4,056 (17)
Deprivation*	9.1 (5.5)	9.1 (5.6)	9.2 (5.6)
Comorbidities			
Anxiety	1,998 (9)	2,027 (9)	2,187 (9)
Depression	2,001 (9)	2,447 (10)	2,601 (11)
Cerebrovascular disease	997 (4)	701 (3)	1,077 (5)
Diabetes	2,490 (11)	2,150 (9)	3,293 (14)
Gout consultation	5,048 (22)	6,315 (27)	4,295 (18)
Hyperlipidaemia	3,340 (14)	2,928 (13)	3,979 (17)
Hypertension	8,915 (38)	7,944 (34)	10,369 (44)
Osteoarthritis	3,694 (16)	3,593 (15)	4,549 (19)
PVD	256 (1)	322 (1)	436 (2)
Renal disease	1,698 (7)	1,742 (7)	2,922 (13)
Lifestyle factors			
Alcohol consumption			
Ever drinker	16,397 (70)	17,432 (75)	17,317 (74)
Never drinker	809 (3)	644 (3)	943 (4)
Missing	6,125 (26)	5,255 (23)	5,071 (22)
Body mass index			
Normal	2,862 (12)	3,100 (13)	2,821 (12)
Overweight	8,013 (34)	8,083 (35)	7,946 (34)
Obese	6,638 (28)	6,688 (29)	7,901 (34)
Missing	5,818 (25)	5,460 (23)	4,663 (20)
Smoking status			
Ever smoker	12,424 (53)	12,410 (53)	13,480 (58)
Never smoker	7,079 (30)	7,397 (32)	6,929 (30)
Missing	3,828 (16)	3,524 (15)	2,922 (13)
SU level			
≤360µmol/L	241 (1)	329 (1)	290 (1)
>360µmol/L	12,078 (52)	9,547 (41)	10,872 (47)
Missing	11,012 (47)	13,455 (58)	12,169 (52)
Medication use			
Analgesics	6,949 (30)	6,110 (26)	7,691 (33)
Colchicine	1,373 (6)	2,036 (9)	1,642 (7)
Diuretics	7,069 (30)	4,573 (20)	6,944 (30)
NSAIDS	12,961 (56)	13,456 (58)	9,855 (42)
Cumulative allopurinol use*	0.2 (0.4)	1.3 (1.6)	4.9 (3.5)

Number and percentage presented unless otherwise stated; *Mean (Standard deviation) presented for continuous covariates; Unit of analysis is intervals; CI: Confidence interval; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

Table 8.20: Estimated treatment effect of allopurinol on renal disease and distribution of covariates within each PS subclass

	Subclass 3 N=22,664	Subclass 4 N=22,665	Subclass 5 N=22,665	Subclass 6 N=22,664
Hazard ratio (95% CI)	0.46 (0.21, 1.00)	2.03 (1.61, 2.56)	1.14 (0.97, 1.34)	0.58 (0.42, 0.81)
Standard error	0.18	0.24	0.09	0.10
Demographics				
Age*	64.0 (14.2)	61.8 (13.9)	61.0 (13.6)	64.8 (12.3)
Sex: Female	4,169 (18.4)	3,531 (16)	3,218 (14)	3,695 (16)
Deprivation*	8.9 (5.5)	9.1 (5.5)	9.3 (5.6)	5.4 (3.3)
Comorbidities				
Anxiety	1,487 (7)	1,775 (8)	1,844 (8)	2,351 (10)
Depression	1,624 (7)	1,977 (9)	2,314 (10)	2,719 (12)
Cerebrovascular disease	979 (4)	781 (3)	803 (4)	1,269 (6)
Coronary heart disease	4,568 (20)	3,922 (17)	3,850 (17)	6,044 (27)
Diabetes	2,105 (9)	2,056 (9)	2,121 (9)	3,652 (16)
Gout consultation	5,009 (22)	3,932 (17)	6,715 (30)	4,076 (18)
Hyperlipidaemia	2,747 (12)	2,695 (12)	3,134 (14)	4,672 (21)
Hypertension	7,498 (33)	7,049 (31)	7,630 (34)	10,531 (46)
Osteoarthritis	2,982 (13)	3,021 (13)	3,245 (14)	4,841 (21)
Peripheral vascular disease	443 (2)	425 (2)	394 (2)	610 (3)
Lifestyle factors				
Alcohol consumption				
Ever drinker	15,547 (69)	15,860 (70)	16,595 (73)	17,997 (79)
Never drinker	985 (4)	805 (4)	737 (3)	861 (4)
Missing	6,132 (27)	6,000 (26)	5,333 (24)	3,806 (17)
Body mass index				
Normal	3,468 (15)	3,073 (14)	2,931 (13)	2,842 (13)
Overweight	7,592 (33)	7,561 (33)	7,899 (35)	8,246 (36)
Obese	5,576 (25)	5,966 (26)	6,529 (29)	7,974 (35)
Missing	6,028 (27)	6,065 (27)	5,306 (23)	3,602 (16)
Smoking status				
Ever smoker	11,394 (50)	11,753 (52)	12,034 (53)	14,389 (63)
Never smoker	6,721 (30)	6,567 (29)	6,702 (30)	6,192 (27)
Missing	4,549 (20)	4,345 (19)	3,929 (17)	2,083 (9)
SU level				
≤360µmol/L	174 (1)	389 (2)	308 (1)	255 (1)
>360µmol/L	8,187 (36)	9,885 (44)	9,773 (43)	10,582 (47)
Missing	14,303 (63)	12,391 (55)	12,584 (56)	11,827 (52)
Medication use				
Analgesics	6,720 (30)	6,598 (29)	6,810 (30)	7,841 (35)
Colchicine	1,059 (5)	1,369 (6)	2,189 (10)	1,532 (7)
Diuretics	7,411 (33)	6,784 (30)	5,861 (26)	7,306 (32)
NSAIDS	13,744 (61)	12,600 (56)	13,525 (60)	9,193 (41)
Cumulative allopurinol use*	0.002 (0.05)	0.3 (0.6)	1.6 (1.8)	5.4 (3.3)

Number and percentage presented unless otherwise stated; *Mean (Standard deviation) presented for continuous covariates; Unit of analysis is intervals; CI: Confidence interval; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SU: Serum urate

8.6.1 Comparison of included and excluded intervals

As intervals were excluded from outcome analysis due to no occurrence of outcome in the subclass(es) with the lowest PS amongst allopurinol intervals, the distribution of covariates were compared between these excluded intervals, and intervals that were included in outcome analysis. This was performed for target SU level, mortality, and coronary heart disease (as that analysis excluded the most number of intervals) (Table 8.21).

For the analysis of target SU level, 75% of intervals were used in treatment effect estimation. Intervals that were excluded from outcome analysis included older patients, less likely to reside in poorer areas, and had lower prevalence of gout consultation and prescription for NSAIDS, higher prevalence of ever smokers, lower mean baseline SU level and cumulative allopurinol use, than intervals that were included for analysis.

For the analysis of mortality, 75% of intervals were used in treatment effect estimation but in the analysis of coronary heart disease fewer intervals were analysed (60%). For both outcomes, excluded intervals contained more patients that were female, residing in poorer areas, had higher prevalence of normal BMI and SU level <360μmol/L, had lower prevalence of gout consultation, and fewer prescriptions for diuretics and NSAIDS, than intervals included for analysis. Similar comparisons between excluded and included intervals were also found in the analyses of gout hospitalisation, cerebrovascular disease, and renal disease (data not shown).

Table 8.21: Generalisability of results

Outcome	Target SU level		Mortality		Coronary heart disease	
	Excluded intervals N=2,507	Included intervals N=7,520	Excluded intervals N=38,333	Included intervals N=116,498	Excluded intervals N=46,663	Included intervals N=69,993
Demographics					,	
Age (Mean (SD))	63.1 (13.2)	60.5 (14.0)	65.4 (14.3)	64.6 (13.7)	62.4 (14.3)	62.5 (13.7)
Sex: Female	559 (22)	1,362 (18)	11,354 (29)	20,900 (18)	12,185 (26)	11,101 (16)
Deprivation (Mean (SD))	8.5 (5.1)	9.2 (5.7)	8.1 (5.5)	9.1 (5.5)	8.7 (5.5)	9.1 (5.6)
Comorbidities						
Anxiety	225 (9)	483 (6)	5,486 (14)	9974 (9)	4,811 (10)	6,212 (9)
Depression	262 (10)	593 (8)	7,026 (18)	11,862 (10)	6,232 (13)	7,049 (10)
Cerebrovascular disease	124 (5)	208 (3)	2,481 (6)	6,248 (5)	1,594 (3)	2,775 (4)
Coronary heart disease	561 (22)	1,456 (19)	8,310 (21)	27,038 (23)	-	-
Diabetes	322 (13)	788 (10)	6,304 (16)	15,590 (13)	5,097 (11)	7,933 (11)
Gout consultation	270 (11)	1,770 (24)	807 (2)	22,363 (19)	1,635 (4)	15,658 (22)
Hyperlipidaemia	485 (19)	930 (12)	7,935 (20)	19,617 (17)	5,159 (11)	10,247 (15)
Hypertension	1,036 (41)	2,513 (33)	15,597 (40)	47,152 (40)	13,279 (28)	27,228 (39)
Osteoarthritis	474 (19)	1,110 (15)	8,105 (21)	20,418 (18)	6,854 (15)	11,836 (17)
Peripheral vascular disease	65 (3)	142 (2)	1,342 (3)	3,171 (3)	1,019 (2)	1,014 (1)
Renal disease	166 (7)	606 (8)	3,560 (9)	12,021 (10)	2,073 (4)	6,362 (9)

Table 8.21 continued:

Lifestyle factors						
Alcohol consumption						
Ever drinker	1,878 (75)	5,554 (74)	30,800 (79)	85,879 (74)	32,812 (70)	51,146 (73)
Never drinker	134 (5)	305 (4)	2,165 (6)	4,790 (4)	2,508 (5)	2,396 (3)
Missing	495 (20)	1,661 (22)	5,868 (15)	25,829 (22)	11,343 (24)	16,451 (24)
Body mass index						
Normal	424 (17)	900 (12)	9,980 (26)	16,622 (14)	10,324 (22)	8,783 (13)
Overweight	860 (34)	2,565 (34)	14,101 (36)	40,810 (35)	15,225 (33)	24,042 (34)
Obese	823 (33)	2,324 (31)	9,474 (24)	34,247 (29)	10,105 (22)	21,227 (30)
Missing	400 (16)	1,731 (23)	5,278 (14)	24,819 (21)	11,009 (24)	15,941 (23)
Smoking status						
Ever smoker	1,570 (63)	4,242 (56)	25,195 (65)	66,327 (57)	24,863 (53)	38,314 (55)
Never smoker	754 (30)	2,050 (27)	11,337 (29)	32,933 (28)	14,717 (32)	21,405 (31)
Missing	183 (7)	1,228 (16)	2,301 (6)	17,238 (15)	7,083 (15)	10,274 (15)
SU level (Mean (SD))	454.3 (49.8)	504.5 (2.0)				
≤360µmol/L	-	-	7,719 (20)	1,386 (1)	6,147 (13)	860 (1)
>360µmol/L	-	-	5,833 (15)	49,163 (42)	7,701 (17)	32,497 (46)
Missing	-	-	25,281 (65)	65,949 (57)	32,815 (70)	36,636 (52)
Medication use						
Analgesics	634 (25)	2,393 (32)	10,439 (27)	38,554 (33)	11,132 (24)	20,750 (30)
Colchicine	36 (1)	712 (9)	254 (1)	7,825 (7)	418 (1)	5,051 (7)
Diuretics	517 (21)	2,282 (30)	6,330 (16)	38,831 (33)	8,281 (18)	18,586 (27)
NSAIDS	1,044 (42)	4,306 (57)	9,005 (23)	57,788 (50)	15,510 (33)	36,272 (52)
Cumulative allopurinol use (Mean (SD))	0 (0)	1.2 (2.0)	0 (0)	1.9 (3.0)	0 (0)	2.1 (3.0)

Number of percentage presented unless otherwise stated; NSAIDS: Non-steroidal-anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

8.7 Sensitivity analysis: impact of missing data

The main analysis was repeated by performing complete case analysis in the analysis of target SU level and mortality. For target SU level, 52% (N=909) had complete data and for mortality 21% (N=3,609) had complete data.

In the analysis of target SU level, allopurinol intervals had higher prevalence of coronary heart disease, gout consultation, renal disease, had higher mean baseline SU level, more prescriptions for analgesics, diuretics and NSAIDS, and longer mean cumulative allopurinol use than non-allopurinol intervals, where SMD >0.10 (Table 8.22). This analysis identified the same imbalanced covariates as the main analysis but had also found imbalance in coronary heart disease and colchicine between treatment groups.

For mortality, allopurinol intervals contained patients who were older, had higher prevalence of coronary heart disease, diabetes, gout consultation, hypertension, renal disease, obesity, SU level >360μmol/L, more prescriptions for analgesics, colchicine and diuretics, and longer mean previous cumulative allopurinol use than non-allopurinol intervals where SMD >0.10 (Table 8.22). Approximately the same number of imbalanced covariates identified here was also observed in the main analysis, albeit with larger SMD.

Distribution of PS between treatment groups is shown in Table 8.23 alongside with the number of subclasses required to achieve overall covariate balance across subclasses, and imbalanced covariates that remained within subclasses.

As the main analysis, subclass 1 had no occurrence of either outcome amongst allopurinol intervals thus that subclass was excluded from treatment effect estimation. In the analysis of target SU level, four PS subclasses were sufficient to achieve overall covariate balance between

treatment groups across subclasses. Within subclasses, there were a higher number of imbalanced covariates than was observed in the main analysis. For example, in subclass 3, age was now imbalanced between treatment groups when previously it was balanced in the main analysis.

For mortality, six subclasses were required to achieve overall covariate balance, whereas in the main analysis five subclasses were sufficient. In contrast, this analysis found a fewer number of imbalanced covariates between treatment groups within subclasses compared with the main analysis. For example, in subclass 2, in this analysis there were 7 imbalanced covariates compared with 12 in the main analysis.

The estimated HRs are presented in Table 8.24. Allopurinol use had higher chance of reaching target SU level (5.31 (3.67, 7.67)), and the estimated HR was larger than the HR estimated in the main analysis (5.31 vs. 4.89). On the other hand, no association was observed between allopurinol and mortality (1.12 (0.89, 1.41)), the same conclusion reached in the main analysis, although estimated HR was larger (1.12 vs. 0.97).

Standard errors were larger in this analysis compared with the main analysis due to smaller sample size (target SU level: 1.00 vs. 0.66; mortality: 0.13 vs. 0.05).

Table 8.22: Distribution of covariates by treatment for target SU level and mortality

	Target SU level		Mortality	
	Non-allopurinol intervals N=3,674	Allopurinol intervals N=287	No allopurinol N=21,946	Allopurinol N=1,026
Demographics				
Age (Mean (SD))	61.0 (13.5)	61.8 (13.4)	64.4 (13.6)	66.2 (12.9)
Sex: Female	731 (20)	303 (20)	5,732 (26)	2,466 (23)
Deprivation (Mean (SD))	9.0 (5.5)	9.4 (5.9)	9.1 (5.6)	9.5 (5.8)
Comorbidities				
Anxiety	299 (8)	132 (9)	2508 (11)	1247 (11)
Depression	365 (10)	117 (8)	2751 (13)	1542 (14)
Cerebrovascular disease	116 (3)	49 (3)	1123 (5)	784 (7)
Coronary heart disease	827 (23)	436 (28)	5805 (26)	3608 (33)
Diabetes	473 (13)	218 (14)	3340 (15)	2164 (20)
Gout consultation	683 (19)	378 (25)	2949 (13)	2215 (20)
Hyperlipidaemia	668 (18)	241 (16)	4479 (20)	2554 (23)
Hypertension	1494 (41)	629 (41)	9484 (43)	5763 (53)
Osteoarthritis	572 (16)	294 (19)	4260 (19)	2601 (24)
Peripheral vascular disease	107 (3)	44 (3)	627 (3)	443 (4)
Renal disease	250 (7)	178 (12)	2097 (10)	1902 (17)
Lifestyle factors				
Alcohol consumption				
Ever drinker	3490 (95)	1475 (96)	20811 (95)	10583 (97)
Never drinker	184 (5)	59 (4)	1135 (5)	360 (3)
Body mass index				
Normal	686 (19)	238 (16)	5074 (23)	1892 (17)
Overweight	1663 (45)	695 (45)	9731 (44)	4636 (42)
Obese	1325 (36)	601 (39)	7141 (33)	4415 (40)
Smoking status				
Ever smoker	2507 (68)	1008 (66)	15765 (72)	8150 (74)
Never smoker	1167 (32)	526 (34)	6181 (28)	2793 (26)
SU level	484.4 (65.3)*	514.2 (76.3)*		
≤360µmol/L	-	-	4477 (20)	324 (3)
>360µmol/L	-	-	17469 (80)	10619 (97)
Medication use				
Analgesics	1116 (30)	587 (38)	7096 (32)	4474 (41)
Colchicine	233 (6)	188 (12)	994 (5)	926 (8)
Diuretics	1032 (28)	550 (36)	6162 (28)	4263 (39)
NSAIDS	1925 (52)	827 (54)	9083 (41)	4764 (44)
Cumulative allopurinol use	0.3 (0.8)	2.2 (2.5)	0.4 (1.2)	4.0 (3.4)

N (%) presented unless otherwise stated; Cells highlighted in yellow indicate SMD >0.10; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Table 8.23: Distribution of PS, number of subclasses and imbalanced covariates within PS subclasses

Outcome	PS Median (Range) Allopurinol No allopurinol	Number of subclasses[a]	Smallest cell[b]	Imbalanced covariates (SMD >0.10)
Target SU level	0.03 (<0.01, 0.97) 0.82 (<0.01, 0.99)	4	2/1,534 (<1%)	S1: N/A S2: Deprivation, diabetes, hyperlipidaemia, renal disease, alcohol consumption, BMI, smoking status, SU level, analgesics, diuretics, NSAIDS, cumulative allopurinol use S3: Age, smoking status, cumulative allopurinol use S4: Age, sex, deprivation, diabetes, gout consultation, BMI, smoking status, SU level, analgesics, colchicine, NSAIDS
Mortality	0.35 (0.02, 0.75) 0.26 (0.01, 0.69)	6	30/1,026 (3%)	S1: N/A S2: Age, coronary heart disease, diabetes, gout consultation, hypertension, diuretics, SU level S3: Gout consultation, SU level S4: Gout consultation, renal disease, analgesics, colchicine, SU level

N/A: Balance not evaluated as this subclass was not used in treatment effect estimation due to no occurrence of outcome within allopurinol intervals; BMI: Body mass index; NSAIDS: Non-steroidal anti-inflammatory drugs; S: Subclass; PS: Propensity score; SU: Serum urate

Table 8.24: Treatment effect estimation in complete case analysis

Outcome	No allopurinol N (%)	Allopurinol N (%)	Unadjusted Hazard ratio (95% CI) Standard error	Adjusted Hazard Ratio (95% CI) Standard error
SU level				
Target level not met	2,281 (96)	1,285 (84)	4.51 (3.51, 5.79) 0.58	5.31 (3.67, 7.67) 1.00
Target level met	93 (4)	247 (16)		
Mortality				
Alive	13,323 (97)	10,554 (96)	1.25 (1.08, 1.44) 0.09	1.12 (0.89, 1.41) 0.13
Died	401 (3)	388 (4)		

CI: Confidence interval; SU: Serum urate

8.8 Summary

Use of time-varying PS subclassification had shown allopurinol increased the chance of reaching target SU level and increased risk of peripheral vascular disease. Treatment effect estimates for gout hospitalisation, cerebrovascular disease, coronary heart disease, and renal disease differed across PS subclasses; allopurinol was either shown to be protective and/or increased the risk of adverse outcome in certain subclasses. Although treatment effect estimates for target SU level and mortality from complete case analysis and the main analysis had the same conclusions, the magnitude of effects differed.

8.8.1 Comparison with baseline PS subclassification

Compared with treating allopurinol as a time-invariant measure, accounting for its possible change in the follow-up resulted in doubling of the HR for target SU level and peripheral vascular disease. As allopurinol has a direct effect on SU level, patients who may have reached target SU level later on during follow-up may have been due to being prescribed allopurinol near that time, which would not have been captured in baseline analysis. Correctly attributing reaching target SU to patients prescribed allopurinol later on, may have caused the estimated HR to increase. It is less clear why treatment effect had doubled for peripheral vascular disease. Increase in treatment effect of this magnitude was not observed for the other outcomes. For both outcomes, standard errors were much larger in time-varying PS subclassification than in baseline analysis.

Treatment effect was almost borderline statistically significant for joint replacement in both baseline and time-varying PS subclassification however, the direction of the HRs differed (1.15 in baseline analysis vs. 0.77 in time-varying analysis). There is the possibility that over time, as

patients continue taking allopurinol, their SU level decreases promoting crystal dissolution and tophi clearance, which may prevent further joint damage.

Time-varying PS subclassification had shown allopurinol use was not associated with premature mortality. The estimated HR was closer to the null value than the HR estimated from baseline PS subclassification, with both methods yielding similar standard errors.

A greater number of outcomes (gout hospitalisation, cerebrovascular disease, coronary heart disease, and renal disease), had subclass-specific treatment effect estimates that were not homogenous across subclasses, unlike baseline analysis where subclass-specific treatment effect estimates were not homogenous for gout hospitalisation only. Within time-varying PS subclassification, allopurinol was shown to be both protective and increases the risk of coronary heart disease and peripheral vascular disease, whereas in baseline analysis allopurinol increased the risk of coronary heart disease. Allopurinol was associated with gout hospitalisation in two of the four subclasses with estimated HR ranging between 1.84 – 2.70, that had similar magnitude as the HRs obtained from baseline analysis that ranged between 1.46 – 2.46.

8.8.2 Strengths and limitations

The strength of PS subclassification is that in theory it is a straightforward and intuitive approach to apply. However, in practice, issues were encountered that were not apparent when baseline PS subclassification was performed.

The main issue was that the positivity assumption was near violated, i.e., the PS were extremely close to zero, for the majority of non-users thus the PS distribution was heavily skewed; this was not observed for allopurinol users, hence common support was poor. Treatment effect estimation

could not be performed in the first subclass (indicating low propensity of treatment) as it contained a small number of allopurinol users who did not have any of the outcomes. Use of interaction terms and non-linear terms in PS modelling improved common support however, the PS distribution remained skewed and dissimilar between treatment groups. Alternative approaches to increase the number of outcomes within allopurinol intervals in the first subclass by removing patients outside of the regions of common support prior to subclassification, or creating PS subclasses based on the PS distribution for allopurinol intervals only, did not work (data not shown).

Use of random intercepts in PS estimation may have been one reason for near violation of the positivity assumption. It is known that random effects logistic model has better discrimination ability (i.e., higher and lower PS estimated for allopurinol users and non-users respectively) than logistic regression omitting random effects (Bouwmeester et al., 2013). This discrimination may result in no overlap of PS between treatment groups and lead to large standard errors in treatment effect estimation. As an exploratory analysis (data not shown), omitting random intercept in PS estimation caused the PS distribution to be normally distributed in both allopurinol users and non-users.

The optimal approach suggested by Leon (2011b) is to account for clustering effects of repeated measurements in PS estimation and treatment effect estimation via random effects. However, random effects in outcome analysis could not be used as the complementary log-log model did not converge. Therefore, random effects were omitted and robust standard errors accounting for clustering effects were instead estimated. Omitting random effects from PS estimation as well as in outcome analysis may potentially bias treatment effect (Leite, 2016, Li et al., 2013).

Due to poor common support and with some subclasses omitted from outcome analysis, generalisability of results was affected. Generally, observations excluded from analysis appeared to be from healthier patients as they had lower prevalence of gout consultations and hypertension, less prescriptions for diuretics and NSAIDS, and SU level was acceptable (<360 μmol/L). These patients did not all the indications for allopurinol treatment which may be why their PS were close to zero.

8.8.3 Conclusions

To conclude, PS subclassification did not perform well in this dataset. There was poor common support resulting in observations removed from outcome analysis affecting generalisability, and not all of patient follow-up was modelled in outcome analysis which is far from ideal.

Alternative methods such as MSM were fitted in the next chapter to see how well fitting these models work in practice.

9 Modelling simple mechanisms of allopurinol via MSM

In this chapter, marginal structural models (MSM) were used to model the effect of treatment on outcome accounting for treatment and covariate histories. It was assumed the association between covariates and treatment were constant regardless of treatment history i.e., whether patients were initiating or continuing with treatment. As alluded to in Section 6.4, making this assumption led to MSM not performing well. Therefore, analysis approach and results are presented, for demonstration purposes, for mortality only. The study sample described in Section 8.1 was used in this chapter.

9.1 Associations between covariates, allopurinol, and mortality

Table 9.1 shows the associations between covariates, treatment, and outcome. The majority of covariates with the exception of anxiety, gout consultation and hyperlipidaemia were associated with mortality. Diuretic use had the strongest association (hazard ratio (HR) 4.13), followed by renal disease (3.34), and cerebrovascular and peripheral vascular diseases (3.01). All covariates apart from non-steroidal anti-inflammatory drugs (NSAIDS) were strongly associated with allopurinol with odds ratios (OR) ranging from 0.25 to 3.93, with serum urate (SU) level having the largest association (OR 254.80). Covariates that were associated with both mortality and allopurinol were included in the propensity score (PS) model.

Standardised mean difference (SMD) for covariates between treatment groups in each interval and overall are shown in Table 9.2. The majority of covariates were imbalanced between treatment groups in at least one follow-up interval (SMD >0.10) with the exception of anxiety, depression, cerebrovascular disease and peripheral vascular disease. The largest differences

between treatment groups over time was observed for cumulative allopurinol use and baseline SU level with SMD >0.25 indicating severe imbalance.

Table 9.1: Associations between covariates, allopurinol, and outcome

	Association with mortality HR (95% CI)	Association with allopurinol OR (95% CI)	Covariates considered for weight estimation
Demographics			
Age	1.10 (1.10, 1.10)	1.12 (1.12, 1.13)	X
Sex: Female	1.73 (1.63, 1.84)	0.59 (0.50, 0.70)	X
Deprivation	1.02 (1.01, 1.02)	1.03 (1.02, 1.05)	X
Comorbidities			
Anxiety	1.04 (0.95, 1.15)	2.61 (2.31, 2.95)	
Depression	1.27 (1.17, 1.38)	1.91 (1.71, 2.13)	X
Cerebrovascular disease	3.01 (2.78, 3.26)	2.76 (2.39, 3.20)	X
Coronary heart disease	2.31 (2.18, 2.45)	3.52 (3.22, 3.85)	X
Diabetes	1.68 (1.57, 1.80)	2.70 (2.46, 2.98)	X
Gout consultation	1.01 (0.93, 1.09)	2.03 (1.93, 2.13)	
Hyperlipidaemia	1.00 (0.93, 1.08)	3.33 (3.07, 3.62)	
Hypertension	1.25 (1.18, 1.33)	3.56 (3.33, 3.80)	X
Osteoarthritis	1.50 (1.40, 1.60)	3.04 (2.79, 3.31)	X
Peripheral vascular disease	3.01 (2.71, 3.33)	2.44 (1.98, 2.99)	X
Renal disease	3.34 (3.10, 3.59)	3.93 (3.63, 4.27)	X
Lifestyle factors			
Alcohol consumption			X
Ever drinker	1.00	1.00	
Never drinker	1.90 (1.71, 2.11)	0.44 (0.37, 0.54)	
Missing	1.11 (1.03, 1.19)	0.36 (0.33, 0.39)	
Body mass index			X
Normal	1.00	1.00	
Overweight	0.57 (0.53, 0.62)	1.30 (1.18, 1.42)	
Obese	0.44 (0.40, 0.48)	2.14 (1.92, 2.40)	
Missing	0.64 (0.60, 0.70)	0.52 (0.46, 0.58)	
Smoking status			X
Ever smoker	1.00	1.00	
Never smoker	0.84 (0.79, 0.90)	0.47 (0.43, 0.51)	
Missing	0.84 (0.76, 0.92)	0.25 (0.23, 0.27)	
SU level			X
≤360µmol/L	1.00	1.00	
>360µmol/L	1.46 (1.26, 1.69)	254.80 (175.30, 370.35)	
Missing	1.61 (1.41, 1.89)	43.92 (30.53, 63.19)	
Medication use			
Analgesics	2.49 (2.36, 2.63)	1.59 (1.50, 1.68)	X
Colchicine	1.53 (1.38, 1.70)	3.00 (2.76, 3.26)	X
Diuretics	4.13 (3.90, 4.38)	1.94 (1.81, 2.08)	X
NSAIDS	0.66 (0.62, 0.70)	1.01 (0.97, 1.06)	X
Cumulative allopurinol use	1.02 (1.01, 1.04)	1.45 (1.43, 1.46)	X

CI: confidence interval; HR: Hazard ratio; NSAIDS: Non-steroidal anti-inflammatory drugs; OR: Odds ratio; SD: Standard deviation; SU: Serum urate; X indicates which covariates were associated with both mortality and allopurinol, thus were included in the propensity score model.

Table 9.2: SMD for each covariate between treatment groups over time

Follow-up year	Age	Sex	Deprivation	Anxiety	Depression	Cerebrovascular disease	Coronary heart disease	Diabetes	Gout consultation	Hyperlipidaemia
1	0.08	0.04	0.11	-0.02	-0.01	0.04	0.11	0.06		<0.01
2	0.11	0.02	0.09	-0.01	0.01	0.05	0.15	0.09	0.70	0.05
3	0.09	-0.02	0.08	-0.03	0.02	0.04	0.16	0.08	0.36	0.05
4	0.07	-0.04	0.07	-0.02	<0.01	0.02	0.14	0.07	0.29	0.03
5	0.06	-0.06	0.07	-0.03	-0.03	0.03	0.13	0.08	0.23	0.07
6	0.06	-0.06	0.05	-0.04	-0.03	0.03	0.11	0.08	0.18	0.06
7	0.03	-0.09	0.04	-0.04	-0.05	0.01	0.10	0.07	0.16	0.05
8	0.04	-0.13	0.03	-0.04	-0.08	<0.01	0.12	0.06	0.14	0.06
9	0.03	-0.14	0.03	-0.05	-0.08	0.04	0.10	0.06	0.13	0.06
10	0.01	-0.14	0.03	-0.06	-0.09	0.04	0.10	0.07	0.09	0.07
11	0.01	-0.16	0.02	-0.05	-0.08	<0.01	0.09	0.05	0.09	0.07
12	-0.02	-0.17	0.02	-0.04	-0.08	0.01	0.09	0.02	0.08	0.07
13	-0.08	-0.19	0.04	-0.07	-0.07	-0.04	0.04	0.03	0.07	0.07
14	-0.06	-0.23	0.06	-0.04	-0.05	<0.01	0.07	0.03	0.14	0.07
15	-0.10	-0.24	0.08	<0.01	-0.04	<0.01	0.09	0.03	0.16	0.07
16	-0.06	-0.29	0.12	0.02	-0.02	0.09	0.13	0.10	0.17	0.12
17	-0.11	-0.26	0.29	0.07	-0.06	0.09	0.20	0.14	0.12	0.13
Overall	0.08	-0.09	0.06	-0.01	-0.01	0.04	0.14	0.10	0.21	0.11

Table 9.2 continued:

					Alcohol consumption			Body mass index			
Follow-up year	Hypertension	Osteoarthritis	Peripheral vascular disease	Renal disease	Ever drinker	Never drinker	Missing	Normal	Overweight	Obese	Missing
1	0.03	0.10	0.05	0.12	0.04	0.02	-0.05	0.01	-0.01	0.09	-0.07
2	0.08	0.09	0.03	0.17	0.06	0.03	-0.07	0.03	-0.03	0.11	-0.09
3	0.09	0.08	0.03	0.17	0.10	0.02	-0.10	0.06	-0.05	0.14	-0.13
4	0.11	0.07	0.01	0.18	0.07	0.02	-0.08	0.05	-0.06	0.12	-0.10
5	0.13	0.06	<0.01	0.18	0.09	-0.02	-0.08	0.03	-0.09	0.14	-0.09
6	0.13	0.05	0.02	0.19	0.08	0.01	-0.08	0.03	-0.10	0.16	-0.12
7	0.15	0.03	0.01	0.20	0.09	<0.01	-0.10	0.03	-0.11	0.15	-0.11
8	0.17	0.04	<0.01	0.18	0.11	-0.04	-0.09	0.04	-0.13	0.16	-0.12
9	0.17	0.04	0.01	0.20	0.11	-0.04	-0.10	0.05	-0.16	0.17	-0.12
10	0.16	0.04	<0.01	0.18	0.11	-0.04	-0.10	0.01	-0.16	0.18	-0.10
11	0.14	0.04	0.01	0.17	0.12	-0.06	-0.10	0.02	-0.16	0.15	-0.07
12	0.16	0.03	-0.01	0.17	0.11	-0.05	-0.09	0.01	-0.16	0.15	-0.06
13	0.14	0.04	-0.01	0.19	0.14	-0.06	-0.12	0.01	-0.20	0.18	-0.06
14	0.14	0.04	0.02	0.18	0.19	-0.07	-0.17	0.01	-0.25	0.22	-0.08
15	0.16	0.07	0.06	0.11	0.20	-0.08	-0.18	-0.01	-0.25	0.23	-0.05
16	0.20	0.10	0.10	0.13	0.21	-0.07	-0.20	0.07	-0.30	0.23	-0.12
17	0.23	0.14	0.06	0.14	0.40	-0.10	-0.39	<0.01	-0.25	0.32	-0.26
Overall	0.18	0.10	0.02	0.22	0.18	-0.02	0.17	0.07	-0.09	0.19	-0.18

Table 9.2 continued:

	Smoking status			SU level							
Follow-up year	Ever smoker	Never smoker	Missing	≤360µmol/L	>360µmol/L	Missing	Analgesics	Colchicine	Diuretics	NSAIDS	Cumulative allopurinol use
1	0.03	-0.05	0.02	-0.28	0.39	-0.27	0.19	0.13	0.25	0.30	-
2	0.08	-0.05	-0.03	-0.29	0.35	-0.23	0.23	0.30	0.28	-0.03	2.12
3	0.06	-0.02	-0.04	-0.30	0.32	-0.19	0.19	0.25	0.24	0.24	2.35
4	0.04	-0.03	-0.01	-0.31	0.31	-0.18	0.16	0.20	0.22	0.13	2.29
5	0.05	-0.02	-0.04	-0.31	0.30	-0.16	0.14	0.20	0.23	0.09	2.32
6	0.07	-0.04	-0.05	-0.32	0.29	-0.15	0.15	0.18	0.22	0.05	2.38
7	0.06	-0.02	-0.06	-0.30	0.28	-0.14	0.14	0.14	0.22	0.04	2.31
8	0.05	-0.02	-0.08	-0.33	0.30	-0.15	0.10	0.12	0.18	0.01	2.31
9	0.08	-0.04	-0.11	-0.34	0.29	-0.14	0.11	0.14	0.17	-0.02	2.28
10	0.06	-0.02	-0.12	-0.33	0.30	-0.16	0.07	0.13	0.16	-0.02	2.30
11	0.07	-0.05	-0.08	-0.33	0.28	-0.13	0.06	0.08	0.18	-0.01	2.26
12	0.07	-0.05	-0.10	-0.33	0.31	-0.16	0.08	0.13	0.15	-0.01	2.23
13	0.08	-0.06	-0.10	-0.31	0.29	-0.15	0.12	0.12	0.12	0.02	2.30
14	0.12	-0.10	-0.09	-0.32	0.30	-0.15	0.10	0.14	0.10	0.02	2.24
15	0.14	-0.12	-0.11	-0.27	0.27	-0.14	0.04	0.16	0.06	<0.01	2.13
16	0.18	-0.14	-0.17	-0.15	0.25	-0.18	0.10	0.16	0.06	-0.11	2.05
17	0.27	-0.21	-0.23	-0.14	0.24	-0.19	<0.01	-0.02	0.07	-0.23	2.13
Overall	0.13	-0.04	-0.13	-0.31	0.30	-0.16	0.14	0.17	0.19	0.03	1.41

Yellow cells indicate absolute SMD >0.10; Red cells indicate absolute SMD >0.25; NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

9.2 Propensity score model and treatment effect estimation

The PS main effects model initially included main effects of covariates and linear functional form of continuous covariates. Unstabilised weights had an extremely skewed distribution with mean (standard deviation (SD)) $4.64*10^{12}$ ($1.16*10^{15}$), median (interquartile range (IQR)) 3.04 (1.83, 26.20), and range 1.03, $3.95*10^{17}$. Consequently, stabilised weights were estimated instead.

Figure 9.1 illustrates the distribution of the stabilised weights that was less skewed than unstabilised weights (but was still skewed), with mean (SD) weight 1.60 (55.9), median (IQR) 1.0 (0.6, 1.2), and range $4*10^{-5}$ to 20,893.76. For the first ten years of follow-up, mean weight was approximately 1 and SD was constant, but after 10 years of follow-up, mean weight and SD increased. In contrast, the median weight was initially approximately 1 for the first few years of follow-up, and thereafter decreased over time with the IQR widening (Table 9.3).

Figure 9.1: Distribution of stabilised weights

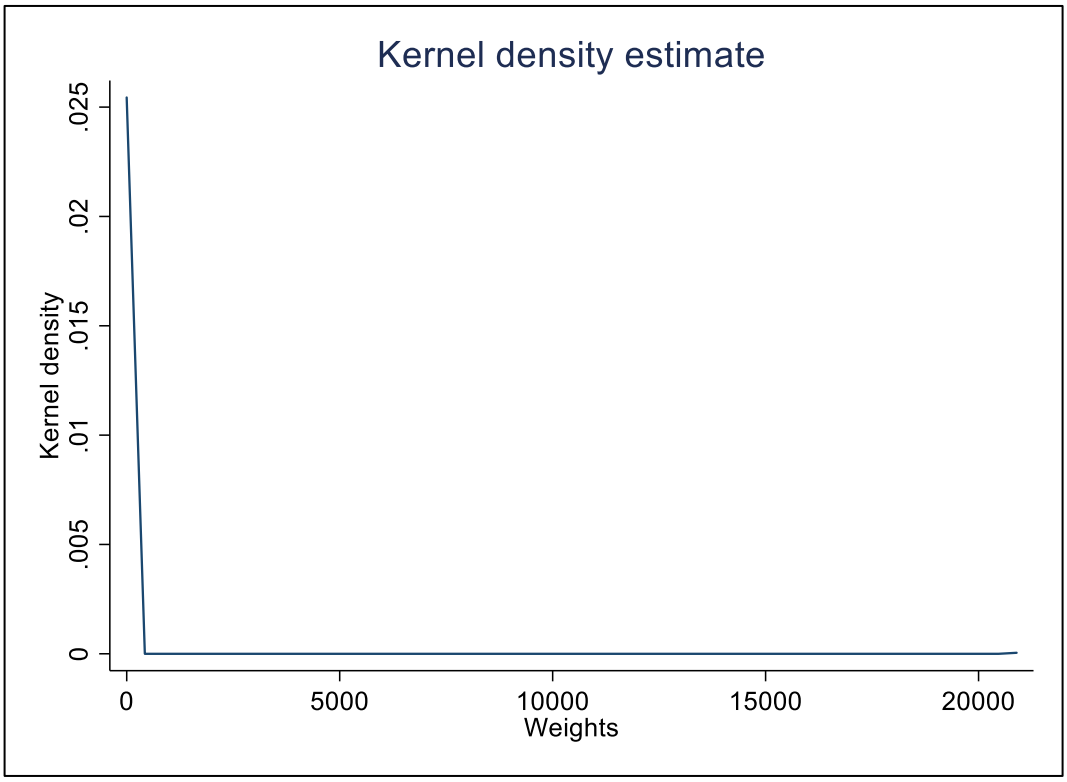


Table 9.3: Distribution of estimated weights over time

Follow-up interval	Mean (SD)	Median (IQR)	Range
1	0.98 (0.21)	1.04 (1.00, 1.09)	0.21, 1.80
2	1.02 (0.41)	1.09 (0.95, 1.24)	0.06, 3.09
3	1.03 (0.49)	1.09 (0.86, 1.28)	0.02, 5.47
4	1.02 (0.54)	1.07 (0.70, 1.27)	0.01, 6.71
5	1.00 (0.59)	1.03 (0.64, 1.25)	0.01, 11.67
6	0.98 (0.67)	0.98 (0.61, 1.21)	0.01, 24.61
7	0.97 (0.70)	0.93 (0.59, 1.16)	0.01, 13.57
8	0.98 (0.84)	0.86 (0.57, 1.11)	0.01, 14.29
9	1.01 (1.12)	0.80 (0.53, 1.07)	0.01, 25.43
10	1.09 (1.70)	0.74 (0.48, 1.02)	0.01, 53.69
11	1.29 (3.35)	0.67 (0.44, 0.98)	0.01, 194.56
12	1.68 (5.86)	0.62 (0.39, 0.95)	$2.35*10^{-3}$, 280.06
13	2.37 (9.62)	0.57 (0.35, 0.92)	$6.00*10^{-4}$, 276.42
14	3.30 (15.04)	0.52 (0.32, 0.91)	$1.36*10^{-4}$, 369.68
15	6.75 (38.45)	0.47 (0.30, 1.01)	$1.62*10^{-4}$, 875.40
16	17.17 (176.03)	0.43 (0.28, 1.17)	$1.89*10^{-4}$, 5469.65
17	73.99 (967.43)	0.39 (0.21, 1.31)	$3.50*10^{-5}$, 20893.76

IQR: Interquartile range; SD: Standard deviation

The unadjusted HR in the unweighted study sample was 1.19 (1.12, 1.26) with standard error 0.03. The treatment effect was then estimated in the weighted study sample and had shown allopurinol reduced risk of premature mortality by 22%; treatment effect was then adjusted for baseline covariates and cumulative allopurinol use, and there was little change in the estimated HR however standard error increased from 0.20 to 0.34 (Table 9.4).

Weights were re-estimated by restricting the PS model to only include covariates with SMD > 0.10 in at least half the follow-up intervals (Table 9.2). These covariates were sex, coronary heart disease, hypertension, renal disease, body mass index (BMI), baseline SU level, analgesics, colchicine, diuretics and cumulative allopurinol use. Compared with the main effects PS model, variability of weights was smaller although extreme weights remained; estimated HR was more conservative and standard error was smaller.

Table 9.4: Distribution of weights and treatment effect of allopurinol

PS model	Mean (SD) Range	Median (IQR)	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
Main effects model	1.55 (55.90) 3.50, 20893.76	1.00 (0.55, 1.17)	0.78 (0.52, 1.15) 0.20	0.77 (0.39, 1.51) 0.34
Restricted main effects model	1.44 (16.33) $4.88*10^{-5}$, 4468.60	1.00 (0.56, 1.16)	0.77 (0.60, 0.97) 0.12	0.85 (0.55, 1.31) 0.22

Main effects propensity score model included age, sex, deprivation, depression, cerebrovascular disease, coronary heart disease, diabetes, hypertension, osteoarthritis, peripheral vascular disease, renal disease, alcohol consumption, BMI, smoking status, baseline SU level, analgesics, colchicine, diuretics, NSAIDS, and cumulative allopurinol use; Restricted main effects model included sex, coronary heart disease, hypertension, renal disease, BMI, baseline SU level, analgesics, colchicine, diuretics and previous cumulative allopurinol use; *Adjusted for cumulative allopurinol use and baseline covariates; CI: Confidence interval; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: standard deviation

Based on the main effects PS model, attempts were made to reduce the variability of weights, i.e., reducing standard deviation. Inclusion of fractional polynomial (FP) terms for continuous covariates and interaction terms were considered, and whether there were problematic covariates that increased weight variability.

9.2.1 Fractional polynomials

Non-linear functions of continuous covariates were included in the PS model (Table 9.5). Use of FP1 terms (model specification 2) lowered weight SD to 4.31 from 55.90 that was obtained from the model specification 1 (the main effects PS model with linear terms). Use of FP2 terms (model specification 3) yielded higher weight SD of 40.06. Model specification 4 identified cumulative allopurinol use with FP2 terms and linear terms for the remaining continuous covariates reduced SD of weights the most (3.69). Despite improvements in reducing weight variability, extreme weights remained present although they were considerably smaller.

The estimated HRs differed across the different PS model specifications. Compared with model specification 1, model specification 4 estimated a stronger unadjusted HR (0.68 vs. 0.78) and had the smallest standard error (0.08 vs. 0.20). Adjustment for baseline covariates and cumulative

allopurinol use in the Cox model estimated similar HRs as the unadjusted HR in model specifications 1 and 2 however, in model specifications 3 and 4, the HRs had become stronger.

Model specification 4 was retained (as estimated weights were the least variable) and backwards selection was performed to identify problematic covariates in PS estimation.

Table 9.5: Distribution of weights and re-estimation of treatment effect of allopurinol

	PS model specification	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	1.55 (55.90)	1.00 (0.55, 1.17)	$3.55*10^{-5}$, 20893.76	0.78 (0.52, 1.15) 0.20	0.77 (0.39, 1.51) 0.34
2	Main effects model + FP1 terms for deprivation$^{(0)}$, cumulative allopurinol use$^{(-2)}$ + linear terms for age, follow-up time	1.40 (4.31)	0.87 (0.65, 1.16)	$9.11*10^{-7}$, 437.06	0.64 (0.53, 0.78) 0.10	0.63 (0.52, 0.76) 0.10
3	Main effects model + FP2 terms for deprivation$^{(-2, 1)}$, cumulative allopurinol use$^{(-2, -1)}$, age$^{(3, 3)}$, follow-up time$^{(-1, 0)}$	1.40 (40.06)	0.95 (0.68, 1.15)	$8.46*10^{-5}$, 13064.98	0.80 (0.67, 0.97) 0.09	0.47 (0.58, 0.94) 0.13
4	Main effects model + linear terms for age, deprivation, follow-up time + cumulative allopurinol use$^{(-2, -1)}$	1.12 (3.69)	0.97 (0.83, 1.10)	$5.07*10^{-4}$, 730.51	0.68 (0.58, 0.79) 0.08	0.57 (0.44, 0.72) 0.13

*Adjusted for cumulative allopurinol use and baseline covariates. Values in brackets (column 2) indicated the fractional polynomial terms used in PS model specification; CI: confidence interval; FP1: fractional polynomial terms of dimension 1; FP2: fractional polynomial terms of dimension 2; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: Standard deviation

9.2.2 Problematic covariates

Backwards selection to identify problematic covariates was next undertaken, which involved by first removing covariates one by one from the PS model and identifying which covariate reduced weight variability i.e., SD, the most. That covariate was then removed the PS model, and the process was repeated to identify the next covariate that reduced weight variability the most, and so on.

The weight distribution and treatment effect estimates are given in Table 9.6. Renal disease was identified as the most problematic covariate; omitting renal disease from the PS model resulted in a reduction in SD of weights from 3.69 to 2.19; unadjusted treatment effect estimate was closer to the null value (0.72 vs. 0.68) and its associated standard error decreased (0.06 vs. 0.08) compared with the PS model retaining renal disease; there was little change in the treatment effect and standard error when renal disease was adjusted for in the Cox model, and little change in HR when further adjusting for baseline covariates and cumulative allopurinol use, although standard error did increase.

Depression was next removed from the PS model and the SD of weights further decreased to 1.45; there was little change in the unadjusted treatment effect (0.71) but associated standard error decreased to 0.05. Once adjusted for depression, baseline covariates and cumulative allopurinol use in the Cox model, the estimated HR moved further away from the null value.

Further removal of covariates from the PS model led to the mean weight moving closer towards 1 and variability decreased. This resulted in HRs (unadjusted and adjusted) to move closer towards the null value and associated standard error decreased.

The decision was made to retain all covariates in the PS model. Although renal disease improved weight estimation, continuing to remove covariates from the PS model still led to large

improvements in weight variability and it had become unclear which covariates should be removed from the PS model, whether one should solely remove renal disease, or to remove more covariates.

Table 9.6: Weight distribution and treatment effect estimation of allopurinol after removing problematic covariates from the PS model

PS model specification	Covariates cumulatively removed from PS model	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error	**Adjusted MSM HR (95% CI) Standard error
Main effects model + linear terms for age, deprivation, follow-up time + FP2 terms for cumulative allopurinol use$^{(-2, -1)}$		1.12 (3.69)	0.97 (0.83, 1.10)	$5.07*10^{-4}$, 730.51	0.68 (0.58, 0.79) 0.08	0.57 (0.44, 0.72) 0.13	0.57 (0.44, 0.72) 0.13
2	Renal disease	1.09 (2.19)	0.97 (0.83, 1.11)	$1.91*10^{-3}$, 421.29	0.72 (0.65, 0.81) 0.06	0.74 (0.66, 0.82) 0.06	0.71 (0.60, 0.85) 0.10
3	Depression	1.08 (1.45)	0.97 (0.83, 1.11)	$1.91*10^{-3}$, 279.98	0.71 (0.64, 0.79) 0.05	0.72 (0.65, 0.80) 0.05	0.67 (0.57, 0.80) 0.09
4	Analgesic use	1.07 (1.06)	0.97 (0.84, 1.11)	$2.56*10^{-3}$, 136.42	0.74 (0.67, 0.81) 0.05	0.75 (0.68, 0.82) 0.05	0.71 (0.62, 0.83) 0.08
5	Diuretic use	1.04 (0.84)	0.99 (0.88, 1.09)	$7.70*10^{-3}$, 176.94	0.91 (0.85, 0.98) 0.04	0.85 (0.79, 0.91) 0.04	0.84 (0.76, 0.93) 0.05
6	Colchicine use	1.04 (0.84)	0.99 (0.88, 1.09)	0.01, 176.94	0.91 (0.85, 0.98) 0.04	0.85 (0.79, 0.91) 0.04	0.84 (0.76, 0.94) 0.05
7	Gout consultation	1.04 (0.56)	0.99 (0.89, 1.08)	0.01, 59.29	0.93 (0.87, 0.99) 0.04	0.86 (0.80, 0.92) 0.04	0.86 (0.78, 0.94) 0.05
8	Body mass index	1.02 (0.40)	0.99 (0.90, 1.07)	0.02, 39.60	0.92 (0.86, 0.99) 0.03	0.89 (0.83, 0.95) 0.03	0.87 (0.79, 0.95) 0.05

Values in brackets (column 1) indicated the fractional polynomial terms used; *Adjusted only for covariates that were cumulatively removed from the PS model; **Adjusted for covariates that were cumulatively removed from the PS model, cumulative allopurinol use and baseline variables; CI: confidence interval; FP2: fractional polynomial terms of dimension 2; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; PS: Propensity score; SD: standard deviation

9.2.3 Interaction terms

From the previous section, no covariates were removed from the PS model (Table 9.7). Starting with the PS model specification 4, two-way interaction terms that improved SD of weights were next added to the PS model. Adding an interaction term between hypertension and gout consultation reduced SD of weight from 3.69 to 2.73. For each extra interaction term added to the PS model, mean weight moved closer to 1 and variability decreased.

With each additional interaction term to the PS model, the unadjusted HR remained constant (0.6) however standard errors decreased from 0.08 to 0.05 after including nine interaction terms. The estimated adjusted HRs were lower (0.5) and standard errors were higher than unadjusted analyses regardless of which interaction terms were included in the PS model.

The process was then repeated but including three way interactions to the PS model (specification 4) (Table 9.8). Similarly, inclusion of a three-way interaction between alcohol consumption, diuretic use and colchicine use caused SD of the weights to decrease from 3.69 to 2.73. Further inclusion of interaction terms further reduced weight variability. Unadjusted and adjusted HRs were similar when adding further interaction terms to the PS model although standard errors did decrease.

It was expected inclusion of one three-way interaction term would reduce weight variability more including one two-way interaction term. However, this was not the case as they both reduced SD of weights by the same amount. Including four three-way interaction terms yielded a higher SD for weight than including nine two-way interaction terms to the PS model (1.20 vs. 1.15). Therefore, three-way interaction terms were no longer considered.

It was clear that regardless of including FP and interaction terms to the PS model, the weight distribution will remain skewed with mean deviating from one and presence of extreme weights.

The PS model that was next considered in sensitivity analyses included a two-way interaction term between hypertension and gout consultation.

Table 9.7: Forward selection of two-way interaction terms in PS model

PS model specification	Two-way interactions cumulatively added to the PS model	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
Main effects model + linear terms for age, deprivation, follow-up time + FP2 terms for cumulative allopurinol use$^{(-2, -1)}$		1.12 (3.69)	0.97 (0.83, 1.10)	$5.07*10^{-4}$, 730.51	0.68 (0.58, 0.79) 0.08	0.57 (0.44, 0.72) 0.13
2	Hypertension*gout consultation	1.11 (2.73)	0.97 (0.83, 1.11)	$4.54*10^{-4}$, 481.27	0.67 (0.58, 0.77) 0.07	0.54 (0.43, 0.68) 0.12
3	Diuretic use*follow-up time	1.11 (2.36)	0.96 (0.82, 1.13)	$7.83*10^{-4}$, 407.46	0.66 (0.58, 0.75) 0.07	0.54 (0.44, 0.67) 0.11
4	Osteoarthritis *depression	1.10 (1.90)	0.96 (0.82, 1.13)	$9.36*10^{-4}$, 332.88	0.67 (0.59, 0.75) 0.06	0.55 (0.45, 0.67) 0.10
5	Alcohol consumption*colchicine use	1.10 (1.62)	0.96 (0.82, 1.13)	$9.34*10^{-4}$, 271.55	0.68 (0.62, 0.76) 0.05	0.58 (0.48, 0.69) 0.09
6	Body mass index*hypertension	1.10 (1.48)	0.96 (0.82, 1.13)	$6.80*10^{-4}$, 195.02	0.67 (0.61, 0.75) 0.05	0.56 (0.47, 0.67) 0.09
7	Gout consultation*follow-up time	1.09 (1.35)	0.96 (0.81, 1.13)	$4.25*10^{-4}$, 177.21	0.69 (0.61, 0.76) 0.05	0.59 (0.49, 0.69) 0.09
8	Hypertension*coronary heart disease	1.09 (1.25)	0.96 (0.81, 1.13)	$6.54*10^{-4}$, 135.96	0.69 (0.62, 0.76) 0.05	0.58 (0.49, 0.69) 0.08
9	Body mass index*depression	1.09 (1.19)	0.96 (0.81, 1.13)	$6.89*10^{-4}$, 123.75	0.69 (0.62, 0.76) 0.05	0.58 (0.49, 0.68) 0.08
10	Diuretic use*depression	1.09 (1.15)	0.96 (0.81, 1.13)	$6.40*10^{-4}$, 134.29	0.69 (0.63, 0.76) 0.05	0.58 (0.50, 0.69) 0.08

*Adjusted for baseline variables and cumulative allopurinol use; CI: Confidence interval; FP2: Fractional polynomial terms of dimension 2; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; PS: Propensity score; SD: Standard deviation

Table 9.8: Forward selection of three-way interaction terms in PS model

PS model specification	Three-way interactions cumulatively added to the PS model	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
Main effects model + linear terms for age, deprivation, follow-up time + FP2 terms for cumulative allopurinol use$^{(-2, -1)}$		1.12 (3.69)	0.97 (0.83, 1.10)	$5.07*10^{-4}$, 730.51	0.68 (0.58, 0.79) 0.08	0.57 (0.44, 0.72) 0.13
2	Alcohol consumption*diuretic use*colchicine use	1.10 (2.26)	0.96 (0.82, 1.12)	$8.34*10^{-4}$, 351.34	0.69 (0.61, 0.77) 0.06	0.58 (0.47, 0.70) 0.10
3	Analgesic use*depression*follow-up time	1.10 (1.74)	0.96 (0.82, 1.12)	$1.76*10^{-3}$, 247.41	0.68 (0.61, 0.76) 0.06	0.58 (0.49, 0.69) 0.09
4	Renal disease*osteoarthritis *depression	1.09 (1.41)	0.96 (0.82, 1.12)	$2.89*10^{-3}$, 185.45	0.70 (0.64, 0.78) 0.05	0.61 (0.52, 0.72) 0.08
5	Alcohol consumption*hypertension*gout consultation	1.09 (1.20)	0.96 (0.82, 1.12)	$1.46*10^{-3}$, 86.92	0.70 (0.63, 0.77) 0.05	0.60 (0.52, 0.70) 0.08

*Adjusted for baseline variables and cumulative allopurinol use; CI: Confidence interval; FP2: Fractional polynomial terms of dimension 2; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; Propensity score; SD: Standard deviation

9.2.4 Assessment of covariate balance

Covariate balance between treatment groups was evaluated for the PS model that included FP2 terms for cumulative allopurinol use and one interaction term between hypertension and gout consultation. SMD was evaluated for each covariate between treatment groups in each year of follow-up and then overall (Table 9.9).

There were large differences between treatment groups across the majority of covariates with SMD >0.25 in at least one interval except for peripheral vascular disease and renal disease. Compared with the unweighted study sample (Table 9.2), within many covariates, covariate balance between treatment groups had worsened in the weighted study sample when it was expected covariate balance would be achieved. For example, for coronary heart disease, SMD ranged from 0.04 to 0.20 in the unweighted study sample whereas in the weighted study sample, SMD ranged from 0.02 to 0.54. In contrast, for some covariates SMD had improved in the weighted study sample. For example, for renal disease, SMD ranged from 0.11 to 0.20 and covariate imbalance was present in all intervals in the unweighted study sample however, in the weighted study sample, SMD ranged from 0.01 to 0.19 with only four intervals that had differences between treatment groups.

Covariate balance was also assessed by overall pooling intervals together. In the weighted study sample, 17 covariates were balanced between treatment groups (SMD<0.10), more than the number of balanced covariates in the unweighted study sample (13). There is some indication that overall balance between treatment groups had improved, but not within intervals.

Table 9.9: SMD over time in the weighted study sample

Follow-up year	Age	Sex	Deprivation	Anxiety	Depression	Cerebrovascular disease	Coronary heart disease	Diabetes	Gout consultation	Hyperlipidaemia	Hypertension
1	-0.03	0.01	0.11	-0.02	0.01	0.02	0.02	0.03	-	-0.02	-0.05
2	-0.03	-0.01	0.10	0.00	0.05	0.02	0.03	0.04	0.53	0.02	0.00
3	-0.06	-0.03	0.09	-0.02	0.06	0.00	0.02	0.02	0.23	0.02	-0.01
4	-0.08	-0.04	0.08	-0.01	0.05	-0.02	-0.02	-0.01	0.20	0.00	0.01
5	-0.10	-0.06	0.07	-0.01	0.02	-0.02	-0.05	-0.01	0.12	0.03	0.01
6	-0.11	-0.05	0.06	-0.01	0.03	-0.02	-0.08	-0.02	0.06	0.00	-0.02
7	-0.14	-0.07	0.05	0.01	0.04	-0.04	-0.12	-0.06	0.03	-0.02	-0.04
8	-0.16	-0.10	0.04	0.00	0.03	-0.06	-0.14	-0.07	-0.01	-0.02	-0.06
9	-0.21	-0.09	0.04	0.04	0.06	-0.03	-0.19	-0.11	-0.05	-0.02	-0.12
10	-0.21	-0.07	0.03	0.05	0.08	-0.03	-0.23	-0.13	-0.12	-0.02	-0.17
11	-0.22	-0.01	0.04	0.09	0.17	-0.06	-0.27	-0.18	-0.19	-0.01	-0.25
12	-0.27	0.08	0.10	0.18	0.26	-0.05	-0.36	-0.24	-0.17	0.01	-0.24
13	-0.36	0.13	0.18	0.33	0.42	-0.12	-0.48	-0.35	-0.30	-0.08	-0.28
14	-0.42	0.27	0.03	0.50	0.59	-0.11	-0.41	-0.23	-0.23	-0.07	-0.25
15	-0.52	0.64	0.00	0.92	0.98	-0.15	-0.29	-0.25	-0.17	0.02	-0.10
16	-0.29	0.55	0.40	0.83	0.81	0.08	-0.54	-0.21	0.10	0.39	-0.16
17	-0.38	-0.07	0.29	0.51	0.46	0.42	-0.35	-0.18	0.27	0.03	-0.43
Overall	-0.06	0.01	0.09	0.20	0.24	0.00	-0.11	-0.01	0.02	0.11	0.01

Table 9.9 continued:

				Alcohol consumption			Body mass index				Smoking status		
Follow-up year	Osteoarthritis	Peripheral vascular disease	Renal disease	Ever drinker	Never drinker	Missing	Overweight	Normal weight	Obese	Missing	Ever smoker	Never smoker	Missing
1	0.07	0.03	0.05	0.02	-0.06	0.01	0.03	0.03	-0.04	0.06	0.02	-0.07	0.05
2	0.04	0.01	0.09	0.02	-0.09	0.02	0.04	0.04	-0.07	0.11	0.05	-0.16	0.09
3	0.03	0.01	0.11	0.02	-0.12	0.05	0.02	0.06	-0.10	0.07	0.08	-0.16	0.11
4	0.01	-0.02	0.07	0.02	-0.08	0.04	0.01	0.03	-0.06	0.01	0.08	-0.11	0.07
5	0.00	-0.04	0.08	-0.02	-0.07	0.02	-0.01	0.03	-0.05	0.00	0.09	-0.11	0.08
6	-0.02	-0.02	0.06	0.01	-0.07	0.03	0.00	0.03	-0.06	0.01	0.07	-0.12	0.06
7	-0.05	-0.05	0.07	0.00	-0.07	0.02	0.02	0.00	-0.04	-0.02	0.07	-0.11	0.07
8	-0.06	-0.06	0.07	-0.05	-0.05	0.03	0.05	-0.04	-0.03	-0.05	0.10	-0.11	0.07
9	-0.08	-0.05	0.06	-0.03	-0.05	0.03	0.05	-0.07	-0.01	-0.07	0.10	-0.11	0.06
10	-0.08	-0.07	0.05	-0.03	-0.04	-0.01	0.13	-0.10	0.02	-0.11	0.14	-0.13	0.05
11	-0.10	-0.04	0.02	-0.01	-0.02	0.02	0.18	-0.20	0.05	-0.11	0.13	-0.07	0.02
12	-0.08	-0.11	0.02	-0.03	-0.01	0.01	0.24	-0.24	0.05	-0.12	0.14	-0.09	0.02
13	-0.09	-0.10	0.02	0.02	-0.04	-0.08	0.28	-0.21	0.09	-0.14	0.16	-0.08	0.02
14	0.07	-0.09	-0.04	0.12	-0.05	0.12	0.29	-0.39	0.06	-0.24	0.26	-0.10	-0.04
15	0.23	-0.06	-0.12	0.26	-0.08	0.14	0.56	-0.62	-0.02	-0.40	0.41	-0.08	-0.12
16	0.47	-0.13	-0.19	0.46	-0.24	0.39	0.13	-0.46	-0.16	-0.45	0.48	-0.16	-0.19
17	-0.12	-0.03	0.14	0.07	-0.20	0.53	0.16	-0.58	-0.20	0.00	0.03	-0.16	0.14
Overall	0.10	-0.03	0.22	0.02	-0.23	0.11	0.15	-0.02	-0.21	0.08	0.10	-0.25	0.22

Table 9.9 continued:

	SU level									
Follow-up year	SU level ≤360µmol/L	SU level >360µmol/L	SU level missing	Analgesic	Colchicine	Diuretic	NSAIDS	Cumulative allopurinol use (FP2 term 1)	Cumulative allopurinol use (FP2 term 2)	Hypertension* Gout consultation
1	-0.29	0.40	-0.27	0.05	0.07	0.02	0.27			
2	-0.29	0.33	-0.21	0.06	0.16	-0.02	-0.02	2.18	2.18	0.18
3	-0.29	0.31	-0.18	0.02	0.15	-0.06	0.20	0.95	1.58	0.12
4	-0.31	0.31	-0.18	0.00	0.11	-0.08	0.10	0.57	1.19	0.09
5	-0.31	0.29	-0.15	-0.03	0.10	-0.10	0.04	0.39	0.96	0.06
6	-0.31	0.28	-0.14	-0.04	0.05	-0.15	-0.01	0.29	0.82	0.02
7	-0.29	0.26	-0.13	-0.07	0.02	-0.21	-0.02	0.14	0.64	0.02
8	-0.31	0.29	-0.15	-0.15	-0.05	-0.30	-0.08	0.07	0.51	-0.01
9	-0.31	0.28	-0.14	-0.20	-0.06	-0.40	-0.11	0.00	0.38	-0.08
10	-0.28	0.25	-0.13	-0.24	-0.15	-0.46	-0.16	-0.06	0.27	-0.11
11	-0.24	0.22	-0.11	-0.25	-0.24	-0.48	-0.16	-0.14	0.16	-0.20
12	-0.20	0.27	-0.17	-0.30	-0.27	-0.58	-0.07	-0.22	0.02	-0.16
13	-0.09	0.23	-0.18	-0.34	-0.35	-0.62	-0.10	-0.28	-0.13	-0.36
14	-0.31	0.20	-0.08	-0.18	-0.26	-0.90	-0.12	-0.30	-0.22	-0.23
15	-0.25	0.17	-0.08	-0.39	-0.11	-1.24	-0.12	-0.44	-0.53	-0.14
16	-0.28	0.06	0.04	0.30	0.12	-0.84	-0.24	-0.37	-0.22	0.17
17	-0.22	0.39	-0.33	-0.38	0.20	-0.46	-0.55	-0.44	-0.39	0.27
Overall	-0.28	0.24	-0.12	-0.06	-0.01	-0.30	-0.10	0.29	0.61	-0.01

Yellow cells indicate absolute SMD >0.10; Red cells indicate absolute SMD >0.25; FP2: Fractional polynomial terms of dimension 2; NSAIDS: Non-steroidal anti-inflammatory drugs; SU: Serum urate; SMD: Standardised mean difference

Possible reasons for extreme weights

A summary of covariates was described for observations with large weights. A small proportion of patients had weights larger than 20; 320 intervals from 121 patients had weights larger than 10; 111 intervals from 43 patients had weights larger than 20.

The distribution of covariates were compared across observations with: weights <10, weights >10, and weights >20 (Table 9.10). Compared to observations with weights <10, those with larger weights consisted of older patients, higher percentage of females, comorbidities (depression, diabetes, hypertension, osteoarthritis, renal disease), ever drinker, ever smoker, normal/overweight BMI, and higher percentage of prescription for analgesics and NSAIDS, as well as higher mean duration of cumulative allopurinol use.

This suggests that larger weights were assigned to patients whose SU level was not measured, in worse health and in older females. It may that this combination of covariates has a very small propensity for treatment but actually received treatment thus resulting in large weights.

Table 9.10: Distribution and comparison of covariates in observations with large versus non-large weights

	Weight <10 N=155,011 time intervals from 16,876 patients	Weight >10 N=320 time intervals from 121 patients	Weight >20 N=111 time intervals from 43 patients
Demographics			
Age (Mean (SD)	64.8 (13.8)	72.2 (10.8)	73.2 (9.5)
Sex: Female	32,163 (21)	91 (28)	43 (39)
Deprivation (Mean (SD))	9.0 (5.5)	10.0 (5.4)	10.4 (5.7)
Comorbidities			
Anxiety	15,378 (10)	82 (26)	40 (36)
Depression	18,767 (12)	121 (38)	45 (41)
Cerebrovascular disease	8,703 (6)	26 (8)	5 (5)
Coronary heart disease	35,280 (23)	68 (21)	28 (25)
Diabetes	21,801 (14)	93 (29)	37 (33)
Gout consultation	23,130 (15)	40 (13)	17 (15)
Hyperlipidaemia	27,442 (18)	110 (34)	43 (39)
Hypertension	62,589 (40)	160 (50)	57 (51)
Osteoarthritis	28,384 (18)	139 (43)	64 (58)
Peripheral vascular disease	4,501 (3)	12 (4)	2 (2)
Renal disease	15,469 (10)	112 (35)	45 (41)
Lifestyle factors			
Alcohol consumption			
Ever drinker	11,6382 (75)	297 (93)	105 (95)
Never drinker	6,942 (4)	13 (4)	6 (5)
Missing	31,687 (20)	10 (3)	0 (0)
Body mass index			
Normal weight	26,495 (17)	107 (33)	35 (32)
Overweight	54,779 (35)	132 (41)	48 (43)
Obese	43,646 (28)	75 (23)	27 (24)
Missing	30,091 (19)	6 (2)	1 (1)
Smoking status	,		
Ever smoker	91,263 (59)	259 (81)	91 (82)
Never smoker	44,212 (29)	58 (18)	20 (18)
Missing	19,536 (13)	3 (1)	0 (0)
SU level			
≤360µmol/L	9,095 (6)	10 (3)	6 (5)
>360µmol/L	54,896 (35)	100 (31)	20 (18)
Missing	91,020 (59)	210 (66)	85 (77)
Medication use			
Analgesic	48,850 (32)	143 (45)	61 (55)
Colchicine	8,035 (5)	44 (14)	20 (18)
Diuretic	45,060 (29)	101 (32)	39 (35)
NSAIDS	66,728 (43)	65 (20)	19 (17)
Cumulative allopurinol use (Mean (SD))	1.4 (2.7)	8.6 (4.7)	8.9 (4.9)
Allopurinol	45,437 (29)	210 (66)	79 (64.23)

NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

9.3 Sensitivity analyses

The previous section highlighted that modifying the PS model was unlikely to estimate a reasonable distribution of weights without large variability. Therefore, a range of sensitivity analyses were performed to assess whether treatment effect estimate was altered by considering normalised weights, weight truncation, truncating follow-up time, and performing intention-to-treat analysis.

The PS model main effects model with FP2 terms for cumulative allopurinol use and an interaction term between hypertension and gout consultation was selected to perform sensitivity analyses on.

9.3.1 Normalised weights

The estimated stabilised weights were normalised, i.e., the mean weight was forced to be one in all intervals. The normalised weight distribution was less skewed with decreased variability; normalised weights had SD of 1.53 compared with 2.73 from stabilised weights. Large weights were still present, although these were not as extreme as stabilised weights (Table 9.11).

The estimated unadjusted HR was 0.71 (0.63, 0.80) which was closer to the null value and had smaller standard error compared with using stabilised weights (0.67 (0.58, 0.67)). The adjusted HR had a stronger protective effect of 0.58 and larger standard error of 0.10.

Table 9.11: Distribution of normalised weights

Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
1.00 (1.53)	0.94 (0.75, 1.06)	$1.97*10^{-4}$, 218.04	0.71 (0.63, 0.80) 0.06	0.58 (0.47, 0.71) 0.10

*Adjusted for baseline variables and cumulative allopurinol use; CI: Confidence interval; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: Standard deviation

9.3.2 Weight truncation

In order to visualise the distribution of weights and why they may need to be truncated, the box plot of the log transformed stabilised and normalised weights over time is shown in Figure 9.2 and Figure 9.3 respectively, illustrating the median (IQR) and range. The box plot highlights that both stabilised and normalised weights had become more dispersed over time.

Figure 9.2: Distribution of stabilised weights over time

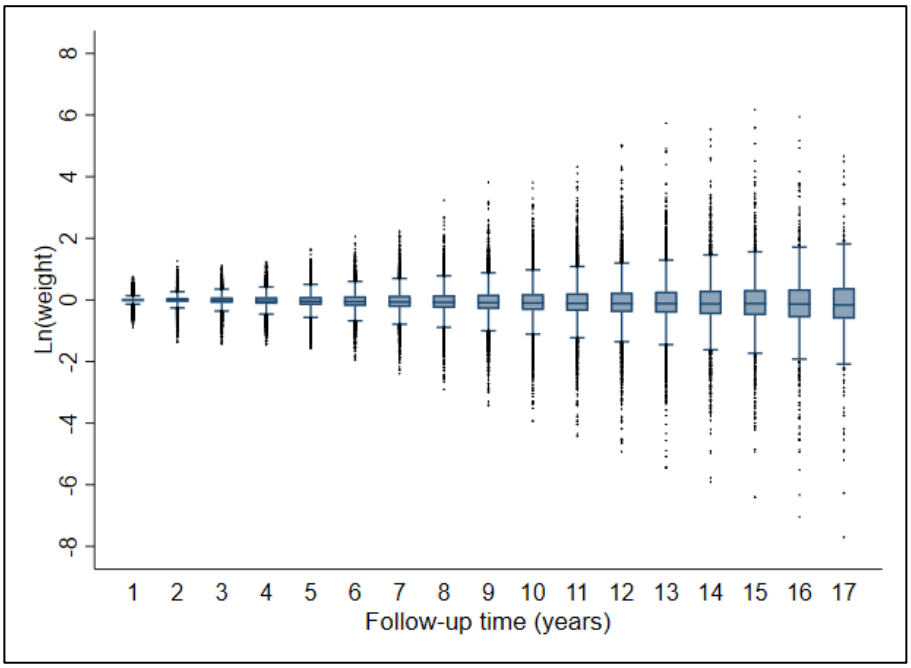
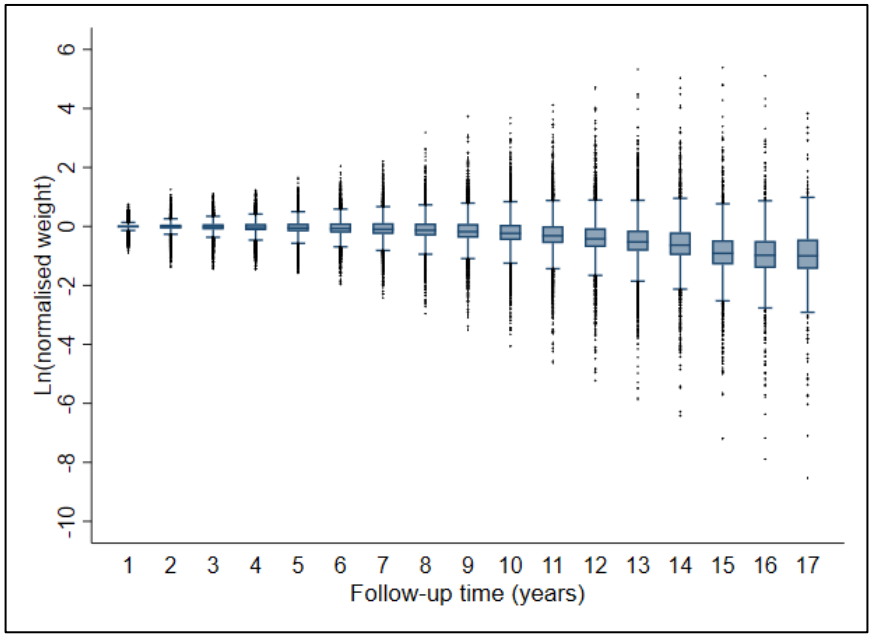


Figure 9.3: Distribution of normalised weights over time



Truncating the stabilised weight distribution at 1% eliminated large weights, with the maximum value being 3.85 reduced from 481.27 when no truncation (0%) had taken place; SD of weights reduced to 0.49 from 2.73, and mean weight reduced to 1.04 from 1.11. The estimated unadjusted HR was closer to the null value (0.77 vs. 0.67) and standard error halved in size (0.04 vs. 0.07). The estimated adjusted HR and standard error was similar to unadjusted results.

Similar results were also obtained for normalised weights when distribution was truncated at 1%; weight variability decreased, and estimated HRs (unadjusted and adjusted) was closer to the null value with smaller standard error compared with no truncation (Table 9.12).

Progressively increasing percentile at which truncation takes place (2%, 5% and 10%), there were small improvements in weight variability for both stabilised and normalised weights, with the estimated (unadjusted and adjusted) HRs moving closer towards the null whilst its standard error decreased. The estimated treatment effects when weighted by stabilised or normalised weights were similar.

Table 9.12: Distribution of truncated weights and its impact on treatment effect estimation of allopurinol

	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
% of truncated stabilised weights					
0%	1.11 (2.73)	0.97 (0.83, 1.11)	$4.53*10^{-4}$, 481.27	0.67 (0.58, 0.77) 0.07	0.54 (0.43, 0.68) 0.12
1%	1.04 (0.49)	0.97 (0.83, 1.11)	0.20, 3.85	0.77 (0.72, 0.83) 0.04	0.72 (0.66, 0.80) 0.05
2%	1.03 (0.41)	0.97 (0.83, 1.11)	0.31, 2.71	0.80 (0.74, 0.86) 0.04	0.76 (0.70, 0.84) 0.05
5%	1.00 (0.32)	0.97 (0.83, 1.11)	0.48, 1.84	0.85 (0.79, 0.91) 0.03	0.83 (0.76, 0.91) 0.04
10%	0.99 (0.24)	0.97 (0.83, 1.11)	0.63, 1.44	0.92 (0.86, 0.98) 0.03	0.90 (0.83, 0.98) 0.04
% of truncated normalised weights					
0%	1.00 (1.53)	0.94 (0.75, 1.06)	$1.97*10^{-4}$, 218.04	0.71 (0.63, 0.78) 0.06	0.58 (0.47, 0.71) 0.10
1%	1.00 (0.45)	0.95 (0.79, 1,08)	0.17, 3.84	0.78 (0.73, 0.84) 0.04	0.73 (0.66, 0.81) 0.05
2%	1.00 (0.39)	0.96 (0.80, 1.09)	0.28, 2.70	0.80 (0.75, 0.86) 0.04	0.77 (0.70, 0.84) 0.05
5%	1.00 (0.32)	0.97 (0.83, 1.10)	0.47, 1.84	0.85 (0.79, 0.91) 0.03	0.83 (0.76, 0.90) 0.04
10%	1.00 (0.24)	0.97 (0.84, 1.12)	0.62, 1.49	0.91 (0.86, 0.97) 0.03	0.90 (0.83, 0.98) 0.04

*Adjusted for cumulative allopurinol use and baseline covariates. CI: Confidence interval; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: Standard deviation

Standardised mean difference

SMD was evaluated for covariates between treatment groups in each year of follow-up. SMD was near identical when evaluated using stabilised and normalised weights, thus SMD was only presented for stabilised weights.

Table 9.13 shows the number of intervals where SMD was greater than 0.25 for each covariate when weight truncation was performed at 0%, 1%, 2%, 5%, and 10%. When no truncation had taken place (0%), large differences (SMD>0.25) between treatment groups were observed in over half the follow-up intervals for baseline SU level, previous cumulative allopurinol use, diuretics, and renal disease; all covariates had large imbalance in at least one interval. Performing 1% weight truncation reduced the number of intervals in which SMD >0.25, for example, renal disease had the number of affected intervals halved (from 8 to 4). Covariates that had a few affected intervals when there was no truncation (for example BMI and hypertension) had no large differences between treatment groups when 1% weight truncation was performed. However, SMD remained persistently high for baseline SU level and previous cumulative allopurinol use performing 10% weight truncation.

Table 9.14 shows for each covariate the number of intervals with SMD >0.10. When no truncation had taken place, over half the intervals had imbalance between treatment groups for baseline SU level, all medications, smoking status, age, coronary heart disease, gout consultation and renal disease. When 1% weight truncation was performed, the number of affected intervals had reduced for the majority of covariates (for example gout consultation, BMI, colchicine), however imbalance was still present. Weight truncation at 1% had little effect on reducing the number of the affected intervals for analgesics, diuretics, previous cumulative allopurinol use, SU level, age, and coronary heart disease. Substantial weight truncation at 10% had achieved balance for the

majority of covariates within intervals however, weight truncation did not improve balance for SU level, diuretics, previous cumulative allopurinol use, age, and sex.

Table 9.13: Number of follow-up intervals where SMD >0.25 after weight truncation

	Percentage of weight truncation				
	0%	1%	2%	5%	10%
Demographics					
Age	6	6	6	5	1
Sex: Female	3	2	4	2	2
Deprivation	2	1	1	1	1
Comorbidities					
Anxiety	5	0	0	0	0
Depression	6	0	0	0	0
Cerebrovascular disease	1	0	0	0	0
Coronary heart disease	7	1	0	0	0
Diabetes	1	0	0	0	0
Gout consultation	3	1	1	1	2
Hyperlipidaemia	1	0	0	0	0
Hypertension	2	0	0	0	0
Osteoarthritis	1	0	0	0	0
Peripheral vascular disease	0	0	0	0	0
Renal disease	8	4	2	0	0
Lifestyle factors					
Alcohol consumption					
Ever drinker	0	0	1	1	1
Never drinker	2	0	0	0	0
Missing	0	1	1	1	1
Body mass index					
Normal	2	0	0	0	0
Overweight	3	0	0	0	0
Obese	4	0	0	0	0
Missing	0	0	0	0	0
Smoking status					
Ever smoker	2	0	0	0	0
Never smoker	3	0	0	0	0
Missing	0	0	0	0	0
SU level					
≤360µmol/L	12	16	15	15	15
>360µmol/L	12	17	17	17	17
Missing	2	1	1	1	1
Medication use					
Analgesic	5	2	0	0	0
Colchicine	3	0	0	0	0
Diuretic	10	10	10	6	0
NSAIDS	2	1	1	1	1
Cumulative allopurinol use FP2 term 1	10	6	6	5	5
Cumulative allopurinol use FP2 term 2	11	10	10	12	12
Interactions					
Hypertension*gout consultation	2	0	0	0	0

FP2: Fractional polynomial terms of dimension 2; NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

Table 9.14: Number of follow-up intervals where SMD >0.10 after weight truncation

	Percentage of weight truncation				
	0%	1%	2%	5%	10%
Demographics					
Age	12	12	12	12	11
Sex: Female	5	10	10	11	11
Deprivation	5	4	4	4	3
Comorbidities					
Anxiety	6	3	2	2	1
Depression	7	4	3	0	0
Cerebrovascular disease	4	4	2	1	0
Coronary heart disease	11	10	10	5	1
Diabetes	9	7	6	1	0
Gout consultation	12	4	4	4	6
Hyperlipidaemia	1	0	0	0	0
Hypertension	9	9	7	1	0
Osteoarthritis	3	1	0	0	0
Peripheral vascular disease	3	0	0	0	0
Renal disease	9	9	8	6	4
Lifestyle factors					
Alcohol consumption					
Ever drinker	4	2	4	5	5
Never drinker	3	0	0	0	0
Missing	3	4	4	4	5
Body mass index					
Normal	4	0	0	0	0
Overweight	8	0	0	2	5
Obese	8	0	0	1	1
Missing	3	2	2	1	2
Smoking status					
Ever smoker	8	2	2	2	3
Never smoker	9	1	1	1	1
Missing	11	11	11	9	7
SU level					
≤360µmol/L	16	17	17	17	17
>360µmol/L	16	17	17	17	17
Missing	14	17	17	17	17
Medication use					
Analgesics	10	10	10	6	2
Colchicine	11	5	5	5	6
Diuretics	13	13	13	12	10
NSAIDS	10	8	7	5	5
Cumulative allopurinol use FP2 term 1	13	12	11	10	10
Cumulative allopurinol use FP2 term 2	15	13	15	15	15
Interaction terms					
Hypertension*gout consultation	10	3	2	3	4

FP2: Fractional polynomial terms of dimension 2; NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

Overall, weight truncation at 1% was sufficient to remove large weights and to estimate a more precise treatment effect estimate and removed some severe covariate imbalance between treatment groups. However, greater weight truncation was required to achieve more comparable treatment groups within intervals however, this led to HRs moving closer to the null value with smaller standard errors.

9.3.3 Truncating follow-up at 10 years

For many covariates, differences between treatment groups increased over time particularly, after 10 years of follow-up (with SMD >0.25) for covariates such as age, coronary heart disease, diuretics etc. in the weighted study sample (Table 9.9). Baseline covariates were compared between patients with 10 or less years of follow-up with patients with more than 10 years of follow-up to understand if there were possible reasons for this (Table 9.15). Treatment effects were then re-estimated with follow-up truncated at 10 years (Table 9.16).

Patients with ≤10 follow-up years were more likely to female (27% vs. 29%), were older (67 vs. 57 years) and resided in poorer areas (9.4 vs. 8.9), and had a higher prevalence of coronary heart disease (17% vs. 10%), more prescriptions for analgesics (40% vs. 27%) and diuretics (48% vs. 26%) than patients with more than 10 years of follow-up.

Truncating follow-up improved variability in both the stabilised and normalised weights and weights were less extreme. Unadjusted HRs of 0.7 were estimated however the adjusted HRs were lower of 0.5 when weighting the study sample using either stabilised or normalised weights (Table 9.16).

Table 9.15: Comparison of baseline covariates of patients with ≤10 years and >10 years of follow-up

	≤10 years of follow-up N=7,748	>10 years of follow-up N=9,128
Demographics		
Age (Mean (SD)	67.1 (15.1)	57.8 (13.0)
Sex: Female	2,117 (27)	1,764 (19)
Deprivation (Mean (SD))	9.4 (5.5)	8.9 (5.5)
Comorbidities		
Anxiety	312 (4)	360 (4)
Depression	427 (6)	415 (5)
Cerebrovascular disease	279 (4)	128 (1)
Coronary heart disease	1,283 (17)	884 (10)
Diabetes	624 (8)	423 (5)
Hyperlipidaemia	347 (4)	436 (5)
Hypertension	1,510 (19)	1,627 (18)
Osteoarthritis	578 (7)	528 (6)
Peripheral vascular disease	190 (2)	67 (1)
Renal disease	164 (2)	53 (1)
Lifestyle factors		
Alcohol consumption		
Ever drinker	4,289 (55)	5,199 (57)
Never drinker	483 (6)	373 (4)
Missing	2,976 (38)	3,556 (39)
Body mass index		
Normal weight	1,314 (17)	1,203 (13)
Overweight	2,208 (28)	2,725 (30)
Obese	1,367 (18)	1,852 (20)
Missing	2,859 (37)	3,348 (37)
Smoking status		
Ever smoker	3,035 (39)	3,401 (37)
Never smoker	2,225 (29)	2,622 (29)
Missing	2,488 (32)	3,105 (34)
SU level		
≤360µmol/L	409 (5)	542 (6)
>360µmol/L	2,767 (36)	3,295 (36)
Missing	4,572 (59)	5,291 (58)
Medication use		
Analgesics	3,098 (40)	2,480 (27)
Colchicine	214 (3)	175 (2)
Diuretics	3,732 (48)	2,410 (26)
NSAIDS	3,587 (46)	4,437 (49)

NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Prior to truncating follow-up, SMD >0.25 was previously observed in the majority of covariates. Truncating follow-up improved SMD. Balance was achieved on 13 covariates compared to only one covariate beforehand. SMD >0.25 was observed for three covariates compared to 21 covariates beforehand (Table 9.17).

Table 9.16: Effect of truncating follow-up on weight distribution and treatment effect estimation of allopurinol

	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
Stabilised weights					
All follow-up observed	1.11 (2.73)	0.97 (0.83, 1.11)	$4.54*10^{-4}$, 481.27	0.67 (0.58, 0.77) 0.07	0.54 (0.43, 0.68) 0.12
Follow-up truncated at 10 years	1.01 (0.78)	1.0 (0.85, 1.10)	0.03, 108.45	0.75 (0.64, 0.87) 0.08	0.57 (0.42, 0.78) 0.16
Normalised weights					
All follow-up observed	1.00 (1.53)	0.94 (0.75, 1.06)	$1.97*10^{-4}$, 218.04	0.71 (0.63, 0.80) 0.06	0.58 (0.47, 0.71) 0.10
Follow-up truncated at 10 years	1.00 (0.72)	1.00 (0.82, 1.10)	0.02, 95.12	0.76 (0.66, 0.88) 0.07	0.59 (0.44, 0.79) 0.15

*Adjusted for cumulative allopurinol use and baseline covariates CI: Confidence interval; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: standard deviation

Table 9.17: Number of intervals where SMD was greater than 0.10 or 0.25

	All follow-up observed		Follow-up truncated at 10 years	
	SMD >0.10	SMD >0.25	SMD >0.10	SMD >0.25
Demographics				
Age	13	1	4	0
Sex: Female	8	3	2	0
Deprivation	8	2	2	0
Comorbidities				
Anxiety	4	1	0	0
Depression	6	3	0	0
Cerebrovascular disease	2	1	0	0
Coronary heart disease	7	2	0	0
Diabetes	3	2	0	0
Gout consultation	12	3	7	1
Hyperlipidaemia	3	2	0	0
Hypertension	4	3	0	0
Osteoarthritis	2	2	0	0
Peripheral vascular disease	1	0	0	0
Renal disease	7	2	0	0
Lifestyle factors				
Alcohol consumption				
Ever drinker	2	0	0	0
Never drinker	0	0	0	0
Missing	2	0	0	0
Body mass index				
Overweight	4	1	0	0
Normal	4	2	0	0
Obese	1	0	0	0
Missing	2	0	0	0
Smoking status				
Ever smoker	5	3	0	0
Never smoker	5	3	0	0
Missing	6	0	1	0
SU level				
≤360µmol/L	17	16	9	9
>360µmol/L	16	15	9	8
Missing	17	4	7	1
Medication use				
Analgesics	8	2	2	0
Colchicine	7	4	1	0
Diuretics	11	6	3	0
NSAIDS	9	2	5	0
Cumulative allopurinol use	16	16	8	8

NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

9.3.4 Intention-to-treat analysis

MSM was modelled under the intention-to-treat principle in that the PS was estimated up to when patients first initiated treatment, thereafter the PS was assumed constant for the remainder of follow-up. Stabilised weights were estimated only.

Assuming patients remained on treatment after initiation, the estimated weights had mean weight of 1 and variability was small. The unadjusted effect of initiating treatment yielded HR of 1.06 (0.99, 1.13); the adjusted HR yielded similar HR with larger standard error (Table 9.18). Covariate imbalance between treatment groups did persist with differences observed across most intervals in gout consultation, target SU level, and cumulative allopurinol use (Table 9.19).

The issues surrounding weight estimation (in terms of mean weight deviating from 1, presence of large weights, and large covariate imbalance between treatment groups) in as-treated analysis did not persist when performing intention-to-treat analysis. This suggests that estimating the PS after patients initiate treatment, may be more complex than initially considered.

Table 9.18: Weight estimation and treatment effect estimation of allopurinol under intention-to-treat analysis

	Mean (SD) weight	Median (IQR) weight	Weight range	Unadjusted HR (95% CI) Standard error	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error
As-treated analysis	1.11 (2.73)	0.97 (0.83, 1.11)	$4.54*10^{-4}$, 481.27	1.19 (1.12, 1.26) 0.03	0.67 (0.58, 0.77) 0.07	0.54 (0.43, 0.68) 0.12
Intention-to-treat analysis	1.00 (0.55)	1.00 (0.70, 1.14)	0.12, 16.90	1.19 (1.12, 1.26) 0.03	1.06 (0.99, 1.13) 0.03	1.06 (0.96, 1.17) 0.05

*Adjusted for cumulative allopurinol use and baseline covariates; CI: Confidence interval; HR: Hazard ratio; IQR: Interquartile range; MSM: Marginal structural models; SD: Standard deviation

Table 9.19: Number of intervals where SMD was greater than 0.10 or 0.25

	As-treated analysis		Intention-to-treat analysis	
	SMD >0.10	SMD >0.25	SMD >0.10	SMD >0.25
Demographics				
Age	13	1	4	0
Sex: Female	8	3	11	4
Deprivation	8	2	6	0
Comorbidities				
Anxiety	4	1	0	0
Depression	6	3	1	0
Cerebrovascular disease	2	1	2	0
Coronary heart disease	7	2	0	0
Diabetes	3	2	0	0
Gout consultation	12	3	16	16
Hyperlipidaemia	3	2	2	0
Hypertension	4	3	4	0
Osteoarthritis	2	2	0	0
Peripheral vascular disease	1	0	0	0
Renal disease	7	2	4	0
Lifestyle factors				
Alcohol consumption				
Ever drinker	2	0	10	0
Never drinker	0	0	4	0
Missing	2	0	6	0
Body mass index				
Normal	4	2	9	0
Overweight	4	1	1	0
Obese	1	0	10	1
Missing	2	0	1	0
Smoking status				
Ever smoker	5	3	3	0
Never smoker	5	3	1	0
Missing	6	0	7	0
SU level				
≤360µmol/L	17	16	17	17
>360µmol/L	16	15	17	17
Missing	17	4	17	3
Medication use				
Analgesics	8	2	1	0
Colchicine	7	4	16	3
Diuretics	11	6	5	0
NSAIDS	9	2	12	0
Cumulative allopurinol use	16	16	16	16

NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

9.4 Summary

Application of MSM had shown that allopurinol was strongly protective of premature mortality. The HRs estimated via as-treated analyses in this chapter ranged from 0.5 to 0.8. These estimates are somewhat lower than what has been reported in the existing literature. A systematic review and meta-analysis of three PS matched cohort studies using electronic health records (EHR) (two of which used UK primary care data) found no association between allopurinol and mortality in gout, reporting an overall HR of 0.80 (0.60, 1.05). Although Dubreuil et al. (2015) had found allopurinol use was protective of premature mortality, their reported HR 0.81 (0.70, 0.92) was closer to the null value of one. It is unclear why exactly the analyses in this chapter provided lower estimates of allopurinol effect, but a possible reason could be differences in modelling approach.

In this chapter several approaches were taken in an attempt to understand exactly the source of issues encountered in estimation of PS modelling. The PS main effects model resulted in the distribution of weights that was skewed, mean weight deviated from 1 and extreme weights were present. This was an indication the PS model was possibly misspecified. A logical approach was to attempt to reduce the extremeness of weights, thus inclusion of non-linear terms and interaction terms in the PS model was considered as a mitigation measure suggested by Cole and Hernan (2008). However, regardless of which covariates and their composites were included in the PS model, distribution of weights remained skewed and extreme weights persisted. The resulting HRs were wide ranging, from 0.5 to 0.8, and it remained unclear which, if any, of the PS models considered were specified correctly. These large HRs were not expected to be observed as treatment of gout is suboptimal in primary care, and patients are generally prescribed low dosages of allopurinol; if a protective effect were to be observed, it was expected the HR would

be much closer to the null value. Balance of covariates between treatment groups was assessed throughout and it was envisaged that there would be little imbalance in the weighted study sample. It transpired that this was not the case with imbalance in fact worse in the weighted than in the unweighted study sample.

Various sensitivity analyses were conducted to evaluate how robust the main analysis treatment effect estimates were against approaches considered for weight variability reduction. Use of normalised weights and weight truncation improved the distribution of weights. The HRs remained strongly protective of premature mortality with unadjusted HR is the region of 0.7 and adjusted HR in the region of 0.5. Similarly, imbalance of covariates between treatment groups remained and only improved when substantial 10% percentile weight truncation was performed compared with 1%.

Truncating follow-up at 10 years drastically improved balance over time although HRs remained strongly protective. It appears that in estimating the as-treated treatment effect, the best approach was to truncate follow-up prior to differences between treatment groups becoming very large in an attempt to achieve covariate balance. This approach is only feasible if one is confident that the estimated HR is plausible and estimating the long-term effect of treatment is maintained and clinically relevant.

The HR estimated under the intention-to-treat principle lost statistical significance and was very close to the null value. Weight estimation (up to when patients initiated treatment) was more satisfactory than estimating weights in all follow-up intervals. This suggests that estimating PS after patients initiated treatment may be more complex. There is the possibility that the associations between covariates and initiating treatment may be different to the associations

between covariates and continuing with treatment which may explain why extreme weights were estimated in this analysis.

To conclude, there are several modelling choices analysts can make while fitting MSM and this study has shown that treatment effect estimates and their precision will depend on specific choices made. Sensitivity to missing data and unmeasured confounding was not assessed as it is believed the associations between covariates and treatment was not modelled correctly.

The next chapter considers modifications to weight estimation process, by allowing associations to differ between covariates and patients initiating allopurinol and between covariates and patients continuing with allopurinol.

10 Modelling complex mechanisms of allopurinol via MSM

In Chapter 9, covariates were assumed to have the same association with initiation and continuation of allopurinol. However, this led to presence and persistence of extreme weights and skewed weight distribution, which in turn may have affected magnitude and precision of treatment effect estimates. In this chapter, this assumption was relaxed, and results from application of marginal structural models (MSM) when associations were allowed to differ in patients initiating and continuing with treatment are presented.

Chapter 6, Section 6.5 described how the study sample described in Section 8.1 was essentially stratified by previous allopurinol use in the last year into two datasets for propensity score (PS) estimation; the first dataset contained observations that did not have allopurinol use in the previous interval, thus allowing one to estimate the probability of initiating allopurinol; the second dataset contained observations that did have allopurinol use in the previous interval, thus allowing one to estimate the probability of continuing treatment. After PS estimation, the datasets were combined together where weight estimation and treatment effect estimation proceeded.

Analyses were performed for all outcomes.

10.1 Covariates associated with initiation and continuation of allopurinol

Table 10.1 describes the distribution of covariates between allopurinol users and non-users stratified by allopurinol use in the previous year; standardised mean difference (SMD) was presented that assessed covariate balance between non-users and patients initiating allopurinol

(in intervals that were not prescribed allopurinol in the previous year), and between patients who discontinued allopurinol and patients who continued with allopurinol (in intervals that were prescribed allopurinol in the previous year).

In intervals that were not prescribed allopurinol in the previous year, patients initiating allopurinol were younger, and had a lower percentage of anxiety, depression, diabetes, gout consultation, hyperlipidaemia, hypertension, having baseline serum urate (SU) level ≥360μmol/L, lower percentage of ever drinkers, normal body mass index (BMI), ever smokers, and more prescriptions for analgesics, colchicine, diuretics, and non-steroidal anti-inflammatory drugs (NSAIDS) compared with non-users.

In intervals that were prescribed allopurinol in the previous year, patients continuing with allopurinol were older, and had a higher percentage of coronary heart disease, diabetes, hyperlipidaemia, hypertension, osteoarthritis, renal disease, ever drinkers, obesity, ever smokers, lower percentage of gout consultation, more prescriptions for diuretics and fewer prescriptions for colchicine and NSAIDS, and had higher mean cumulative allopurinol use than patients discontinuing with allopurinol.

The majority of covariates were associated with initiating and continuing allopurinol, although the direction of the odds ratio (ORs) differed. Older age, anxiety, cerebrovascular disease, diabetes, hyperlipidaemia, hypertension, osteoarthritis, and longer cumulative allopurinol use had increased odds of continuing with allopurinol (OR ranged from 1.04 to 2.15); in contrast, these same covariates had reduced odds of initiating allopurinol (OR ranged from 0.62 to 0.99).

On the other hand, gout consultation, never smokers, and NSAID use had decreased odds of continuing allopurinol (OR ranged from 0.51 to 0.76) but had increased odds of initiating allopurinol (OR ranged from 1.21 to 2.17).

Renal disease was only associated with continuing allopurinol. Baseline SU level above target and colchicine use were only associated with initiating allopurinol.

Table 10.1: Distribution and association of covariates with initiating and continuing with allopurinol (n=16,876 patients)

	Allopurinol non-users in the previous year				Allopurinol users in the previous year			
	Total number of intervals N=115,345	Intervals not on allopurinol N=105,919	Intervals initiated allopurinol N=9,426	OR (SE) of initiating allopurinol	Total number of intervals N=39,986	Intervals discontinued allopurinol N=3,765	Intervals continuing allopurinol use N=36,221	OR (SE) of continuing allopurinol
Demographics								
Age (Mean (SD))	64.5 (14.0)	64.6 (14.0)	62.7 (14.2)	0.99 (<0.01)*	65.9 (13.2)	61.7 (14.7)	66.3 (12.9)	1.04 (<0.01)*
Sex: Female	25,006 (22)	23,268 (22)	1,738 (18)	0.80 (0.04)*	7,248 (18)	640 (17)	6,608 (18)	1.06 (0.10)
Deprivation (Mean (SD))	8.9 (5.5)	8.9 (5.5)	9.3 (5.6)	1.02 (<0.01)*	9.2 (5.6)	9.5 (5.6)	9.2 (5.6)	0.99 (0.01)
Comorbidities								
Anxiety	11,262 (10)	10,620 (10)	642 (7)	0.62 (0.04)*	4,198 (10)	345 (9)	3,853 (11)	1.27 (0.14)*
Depression	13,795 (12)	13,008 (12)	787 (8)	0.57 (0.04)*	5,093 (13)	432 (11)	4,661 (13)	1.17 (0.11)
Cerebrovascular disease	6,049 (5)	5,684 (5)	365 (4)	0.72 (0.06)*	2,680 (7)	177 (5)	2,503 (7)	1.67 (0.22)*
Coronary heart disease	24,092 (21)	22,310 (21)	1,782 (19)	0.93 (0.04)	11,256 (28)	736 (20)	10,520 (29)	2.03 (0.15)*
Diabetes	14,814 (13)	13,914 (13)	900 (10)	0.65 (0.04)*	7,080 (18)	423 (11)	6,657 (18)	2.00 (0.17)*
Gout consultation	13,826 (12)	12,128 (11)	1,698 (18)	1.75 (0.06)*	9,344 (23)	1,704 (45)	7,640 (21)	0.54 (0.03)*
Hyperlipidaemia	18,677 (16)	17,549 (17)	1,128 (12)	0.69 (0.04)*	8,875 (22)	577 (15)	8,298 (23)	2.04 (0.16)*
Hypertension	43,002 (37)	40,033 (38)	2,969 (31)	0.76 (0.03)*	19,747 (49)	1,393 (37)	18,354 (51)	2.15 (0.13)*
Osteoarthritis	19,522 (17)	18,237 (17)	1,285 (14)	0.76 (0.04)*	9,001 (23)	694 (18)	8,307 (23)	1.50 (0.12)*
Peripheral vascular disease	3,166 (3)	2,964 (3)	202 (2)	0.70 (0.08)*	1,347 (3)	93 (2)	1,254 (3)	1.43 (0.26)
Renal disease	9,043 (8)	8,368 (8)	675 (7)	1.05 (0.06)	6,538 (16)	406 (11)	6,132 (17)	1.85 (0.15)*

Table 10.1 continued:

Lifestyle factors								
Alcohol consumption								
Ever drinker	84,756 (73)	78,371 (74)	6,385 (68)	1.00	31,923 (80)	2,823 (75)	29,100 (80)	1.00
Never drinker	5,283 (5)	4,889 (5)	394 (4)	0.98 (0.09)	1,672 (4)	155 (4)	1,517 (4)	0.82 (0.13)
Missing	25,306 (22)	22,659 (21)	2,647 (28)	1.52 (0.06)*	6,391 (16)	787 (21)	5,604 (15)	0.60 (0.04)*
Body mass index								
Normal	21,085 (18)	19,818 (19)	1,267 (13)	1.00	5,517 (14)	543 (14)	4,974 (14)	1.00
Overweight	40,326 (35)	37,206 (35)	3,120 (33)	1.39 (0.08)*	14,585 (36)	1,305 (35)	13,280 (37)	1.11 (0.10)
Obese	29,812 (26)	27,330 (26)	2,482 (26)	1.61 (0.10)*	13,909 (35)	1,108 (29)	12,801 (35)	1.41 (0.14)*
Missing	24,122 (21)	21,565 (20)	2,557 (27)	2.00 (0.12)*	5,975 (15)	809 (21)	5,166 (14)	0.66 (0.07)*
Smoking status								
Ever smoker	65,783 (57)	61,109 (58)	4,674 (50)	1.00	25,739 (64)	2,136 (57)	23,603 (65)	1.00
Never smoker	33,394 (29)	30,683 (29)	2,711 (29)	1.21 (0.05)*	10,876 (27)	1,110 (29)	9,766 (27)	0.76 (0.05)*
Missing	16,168 (14)	14,127 (13)	2,041 (22)	1.88 (0.08)*	3,371 (8)	519 (14)	2,852 (8)	0.46 (0.04)*
SU level								
≤360µmol/L	8,547 (7)	8,408 (8)	139 (1)	1.00	558 (1)	60 (2)	498 (1)	1.00
>360µmol/L	36,718 (32)	32,497 (31)	4,221 (45)	27.30 (3.84)*	18,278 (46)	1,695 (45)	16,583 (46)	1.72 (0.49)
Missing	70,080 (61)	65,014 (61)	5,066 (54)	11.11 (1.53)*	21,150 (53)	2,010 (53)	19,140 (53)	1.61 (0.46)
Medication use								
Analgesics	34,382 (30)	31,151 (29)	3,231 (34)	1.35 (0.05)*	14,611 (37)	1,292 (34)	13,319 (37)	1.17 (0.06)*
Colchicine	4,682 (4)	3,864 (4)	818 (9)	2.81 (0.16)*	3,397 (8)	534 (14)	2,863 (8)	0.89 (0.07)
Diuretics	31,542 (27)	28,114 (27)	3,428 (36)	2.19 (0.08)*	13,619 (34)	974 (26)	12,645 (35)	1.74 (0.11)*
NSAIDS	50,013 (43)	44,335 (42)	5,678 (60)	2.17 (0.06)*	16,780 (42)	2,363 (63)	14,417 (40)	0.51 (0.02)*
Cumulative allopurinol use (Mean (SD))	0.3 (1.0)	0.3 (1.0)	0.4 (1.3)	0.67 (0.01)*	4.8 (3.3)	2.9 (2.5)	4.9 (3.3)	1.21 (0.01)*

Number (percentage) presented unless otherwise stated; Cells highlighted in yellow indicate SMD>0.10; *Statistically significant p-value <0.05; NSAIDS: Non-steroidal anti-inflammatory drugs; SE: Standard error; SD: Standard deviation; Standardised mean difference; SU: Serum urate

10.2 Propensity score model

Across all outcome analyses, the same covariates used to estimate PS in Chapter 8 for time-varying PS subclassification (Table 8.7) was used to estimate the probability of initiating allopurinol and the probability of continuing allopurinol. To briefly recap, age, sex, deprivation, renal disease, colchicine, NSAIDS, diuretics, SU level, gout consultation, cumulative allopurinol use, and follow-up time were chosen to be in the PS models a priori. Covariates that were associated with outcome, by modelling the association between covariate and outcome via the complementary log-log model, were also included in the PS model if the p-value<0.05; generally, the majority of demographics, comorbidities, lifestyle factors and medication use were associated with most outcomes. In estimating the PS for initiating allopurinol and continuing allopurinol, the PS models considered main effects of all covariates and continuous covariates had a linear functional form, and this was considered as the main analysis. All covariates were entered into the PS model as main effects with linear functional form for continuous outcomes.

Stabilised weights vs. basic weights

After PS estimation, initially stabilised weights were derived using the whole study sample. The distribution of weights across all outcomes was satisfactory with mean close to 1, small standard deviation (SD), and narrow range. For example, in the analysis of target SU level, mean (SD) weight was 1.00 (0.44) with range 0.05 to 7.18; the distribution was not skewed as median (interquartile range (IQR)) was 1.02 (0.72, 1.15). Similar findings were also observed in the analyses of the secondary outcomes with the exception for morality, that had a larger weight range of <0.01 to 188, although mean (SD) was 1 (0.99).

As the stabilised weight distribution was satisfactory, basic weights were then derived and how it compared with stabilised weights. Across all outcomes, the basic weight distribution was

satisfactory with mean weight close to one although variability had increased compared with stabilised weights. For example, in the analysis of target SU level, basic weights had mean (SD) 0.99 (0.68) with range <0.01 to 13.05. With the exception of mortality, similar findings were observed in secondary outcome analyses that basic weight variability was larger than stabilised weights, for example in the analysis of cerebrovascular disease, basic weights had mean (SD) 1.00 (0.45) whereas for stabilised weights it was 1.00 (0.29). In the analysis of mortality, the opposite was found that the variability of basic weights was lower than stabilised weights (mean (SD) 1.00 (0.81) vs. 1.00 (0.99)).

The decision was taken to use basic weights for the main analysis. This was because the confounding effects of baseline covariates and treatment history were fully adjusted for within weight estimation, which was not the case for stabilised weights. Therefore, in treatment effect estimation, weighting the study sample using basic weights did not need to adjust for baseline covariates and treatment history in the Cox model, which would be required to do so if the study sample was weighted using stabilised weights.

Assessment of covariate balance

SMD evaluated covariate balance between treatment groups in each year of follow-up and overall in the weighted study sample. This was presented for the analysis for mortality (Table 10.2). For all other outcomes, overall SMD for each covariate was presented (Table 10.3).

As observed in Table 10.2, covariate balance was achieved for deprivation and peripheral vascular disease across all intervals. Severe imbalance (SMD>0.25) between treatment groups was present for gout consultation and cumulative allopurinol use in the majority of intervals; colchicine was also imbalanced across the majority of intervals but was not as severe. For the remaining

covariates, balance tended to be achieved in the first 12 years of follow-up, with small imbalances (SMD≤0.25) appearing thereafter.

From Table 10.3, overall covariate balance across time found covariate imbalance remained in the weighted study sample. Across all outcome analyses, the largest difference between treatment groups was observed for cumulative allopurinol use (SMD>0.76). Sex, deprivation, anxiety, depression, and NSAIDS were balanced between treatment groups, whereas small imbalance was observed for most lifestyle factors, renal disease, and hypertension.

Table 10.2: SMD over time in the weighted sample in analysis of mortality

Follow-up year	Age	Sex	Deprivation	Anxiety	Depression	Cerebrovascular disease	Coronary heart disease	Diabetes	Gout consultation	Hyperlipidaemia	Hypertension
1	0.13	0.10	0.05	-0.02	0.01	0.06	0.08	0.09	-	-0.01	0.03
2	0.12	0.08	0.05	<0.01	0.03	0.06	0.09	0.09	0.70	0.04	0.07
3	0.08	0.03	0.04	-0.02	0.04	0.03	0.08	0.07	0.33	0.04	0.06
4	0.03	0.02	0.04	<0.01	0.04	<0.01	0.06	0.06	0.34	0.02	0.08
5	0.03	0.02	0.03	0.01	0.01	0.03	0.06	0.07	0.29	0.10	0.10
6	0.02	0.01	0.01	-0.01	0.02	0.03	0.03	0.06	0.25	0.07	0.08
7	-0.02	-0.03	<0.01	0.01	0.01	0.02	<0.01	0.04	0.25	0.05	0.07
8	-0.02	-0.10	-0.02	-0.01	-0.05	<0.01	0.02	0.04	0.26	0.04	0.08
9	-0.02	-0.10	-0.03	-0.01	-0.06	0.04	0.01	0.03	0.26	0.04	0.10
10	-0.05	-0.10	-0.03	-0.01	-0.09	0.03	0.01	0.02	0.28	0.06	0.07
11	-0.06	-0.11	-0.03	-0.03	-0.09	-0.01	<0.01	-0.01	0.21	0.04	0.02
12	-0.11	-0.11	-0.03	0.02	-0.01	0.03	-0.02	-0.09	0.25	0.07	0.02
13	-0.17	-0.14	<0.01	0.10	0.08	-0.07	-0.07	-0.08	0.19	0.07	0.05
14	-0.14	-0.24	0.07	0.19	0.19	-0.05	-0.06	-0.20	0.43	0.12	0.07
15	-0.14	-0.33	0.03	0.10	0.09	0.04	-0.03	-0.14	0.40	0.04	-0.01
16	-0.13	-0.35	0.09	0.04	-0.01	0.18	-0.07	-0.09	0.36	0.04	-0.09
17	-0.29	-0.17	0.03	-0.25	-0.37	-0.06	-0.18	-0.14	0.42	0.06	-0.16
Overall	0.08	-0.07	0.01	0.08	0.08	0.06	0.09	0.11	0.28	0.18	0.21

Table 10.2 continued:

				Alcohol consumption			Body mass index				Smoking status		
Follow-up year	Osteoarthritis	Peripheral vascular disease	Renal disease	Ever drinker	Never drinker	Missing	Overweight	Normal weight	Obese	Missing	Ever smoker	Never smoker	Missing
1	0.08	0.05	0.05	0.04	0.04	-0.05	0.02	0.04	0.05	-0.06	0.06	0.01	-0.06
2	0.08	0.02	0.10	0.05	0.05	-0.07	0.02	0.02	0.06	-0.07	0.09	0.02	-0.10
3	0.08	0.03	0.10	0.06	0.05	-0.08	0.05	<0.01	0.06	-0.09	0.05	0.04	-0.09
4	0.06	<0.01	0.11	0.03	0.04	-0.05	0.04	-0.01	0.04	-0.06	0.01	0.02	-0.03
5	0.05	<0.01	0.09	0.07	<0.01	-0.07	0.03	-0.04	0.06	-0.06	0.02	0.03	-0.07
6	0.03	<0.01	0.07	0.05	0.01	-0.06	0.02	-0.03	0.10	-0.09	0.03	0.01	-0.07
7	0.01	-0.02	0.09	0.06	0.01	-0.06	0.02	-0.04	0.09	-0.08	0.03	0.01	-0.09
8	0.01	-0.03	0.06	0.08	-0.06	-0.05	0.04	-0.07	0.09	-0.10	0.02	0.03	-0.11
9	0.03	-0.02	0.10	0.07	-0.04	-0.06	0.03	-0.09	0.09	-0.07	0.05	-0.01	-0.13
10	0.02	-0.02	0.05	0.08	-0.03	-0.07	-0.01	-0.07	0.10	-0.06	<0.01	0.03	-0.10
11	0.01	<0.01	0.01	0.08	-0.06	-0.05	0.02	-0.05	0.02	<0.01	0.03	-0.02	-0.06
12	0.02	0.01	-0.01	0.06	-0.06	-0.04	<0.01	-0.02	<0.01	0.02	0.02	<0.01	-0.08
13	-0.01	-0.01	-0.01	0.03	0.10	-0.10	0.03	-0.07	0.01	0.02	0.08	-0.06	-0.10
14	0.04	-0.03	0.01	0.07	0.13	-0.18	<0.01	-0.13	0.09	0.01	0.07	-0.05	-0.10
15	0.02	0.06	-0.21	0.24	-0.08	-0.23	0.08	-0.12	-0.03	0.09	0.13	-0.11	-0.12
16	-0.03	-0.05	-0.23	0.21	0.07	-0.27	0.22	-0.08	-0.12	-0.06	0.18	-0.14	-0.17
17	-0.08	0.08	-0.31	0.19	0.08	-0.25	-0.01	-0.09	0.07	0.02	0.10	-0.08	-0.14
Overall	0.14	0.02	0.19	0.28	<0.01	-0.29	0.12	0.01	0.16	-0.28	0.21	<0.01	-0.29

Table 10.2 continued:

	SU level							
Follow-up year	SU level ≤360µmol/L	SU level >360µmol/L	SU level missing	Analgesic	Colchicine	Diuretic	NSAIDS	Cumulative allopurinol use
1	-0.03	0.14	-0.12	0.07	<0.01	0.09	0.02	-
2	-0.02	0.09	-0.07	0.14	0.22	0.10	-0.08	2.54
3	<0.01	0.03	-0.03	0.08	0.19	0.05	0.18	2.24
4	-0.01	0.02	-0.01	0.09	0.17	0.04	0.14	2.04
5	0.03	-0.02	<0.01	0.04	0.17	0.03	0.09	1.90
6	0.01	-0.03	0.02	0.07	0.16	0.03	0.07	1.92
7	0.05	-0.06	0.03	0.05	0.13	0.03	0.05	1.72
8	-0.04	-0.03	0.05	-0.02	0.12	-0.03	0.06	1.56
9	-0.09	-0.03	0.07	-0.01	0.13	-0.02	0.04	1.43
10	-0.07	-0.02	0.05	-0.05	0.14	-0.04	0.10	1.32
11	-0.06	-0.05	0.08	-0.04	0.10	-0.05	0.09	1.15
12	-0.05	0.01	0.01	-0.04	0.15	-0.09	0.10	1.00
13	0.13	-0.03	-0.04	-0.07	0.17	-0.14	0.10	0.90
14	0.24	-0.08	-0.06	-0.05	0.37	-0.18	0.20	0.72
15	0.19	-0.10	-0.01	-0.07	0.28	-0.15	0.24	0.41
16	0.17	-0.12	0.04	-0.09	0.41	-0.24	0.07	0.19
17	-0.01	-0.30	0.30	-0.38	0.28	-0.55	0.10	<0.01
Overall	<0.01	-0.02	0.02	0.04	0.17	-0.04	<0.01	1.18

Yellow cells indicate absolute SMD >0.10; Red cells indicate absolute SMD >0.25; NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate

Table 10.3: Overall SMD in the weighted study sample

Outcome	Target SU level	Mortality	Gout hospitalisation	Joint replacement	Cerebrovascular disease	Coronary heart disease	Peripheral vascular disease	Renal disease
Demographics								
Age	0.01	0.08	0.10	0.14	0.14	0.11	0.14	0.08
Sex: Female	0.02	-0.07	0.01	0.02	0.01	<0.01	0.01	-0.02
Deprivation	0.05	0.01	0.04	0.06	0.06	0.06	0.06	0.06
Comorbidities								
Anxiety	0.06	0.08	0.02	0.02	0.01	0.03	0.02	0.04
Depression	0.01	0.08	0.01	<0.01	<0.01	0.03	<0.01	0.03
Cerebrovascular disease	0.04	0.06	0.05	0.13		0.06	0.13	0.09
Coronary heart disease	-0.07	0.09	0.11	0.07	0.13		0.06	0.05
Diabetes	0.08	0.10	0.10	0.08	0.11	0.09	0.11	0.08
Gout consultation	0.28	0.29	0.22	0.22	0.22	0.25	0.22	0.24
Hyperlipidaemia	0.06	0.18	0.06	0.10	0.09	0.09	0.09	0.07
Hypertension	0.07	0.21	0.14	0.15	0.16	0.16	0.15	0.12
Osteoarthritis	0.06	0.14	0.09	0.10	0.12	0.10	0.11	0.10
Peripheral vascular disease	-0.01	0.01	0.04	0.03	0.04	0.04		0.03
Renal disease	0.10	0.19	0.20	0.22	0.22	0.18	0.22	
Lifestyle factors								
Alcohol consumption[a]	-0.19	-0.29	0.12	0.15	-0.15	0.14	0.15	0.14
Body mass index[a]	-0.19	-0.28	0.12	0.15	-0.15	0.14	-0.15	0.13
Smoking status[a]	-0.21	-0.29	0.10	0.11	0.12	0.10	0.11	0.10
SU level[a]	-0.01[b]	0.02	0.22	0.19	0.19	0.20	-0.13	0.19
Medication use								
Analgesics	0.01	0.04	0.13	0.16	0.16	0.14	0.15	0.13
Colchicine	0.12	0.17	0.15	0.16	0.16	0.16	0.15	0.16
Diuretics	-0.04	-0.04	0.15	0.18	0.17	0.12	0.17	0.12
NSAIDS	-0.03	<0.01	0.02	0.02	0.02	0.07	0.02	0.07
Cumulative allopurinol use	0.76	1.10	1.24	1.24	1.25	1.19	1.26	1.20

Cells highlighted in red indicate severe covariate balance with SMD>0.25; Cells highlighted in yellow indicate small covariate imbalance with SMD between 0.10 and 0.25; [a]For categorical variables, the largest SMD was presented; [b]SU level was a continuous covariate; NSAIDS: Non-steroidal anti-inflammatory drugs; SMD: Standardised mean difference; SU: Serum urate.

10.3 Treatment effect estimation

Treatment effect estimates were derived in the weighted study sample using basic weights (Table 10.4). Allopurinol use was strongly associated with reaching target SU level (HR 4.73 (3.89, 5.76)) and increased risk of gout hospitalisation (2.66 (2.40, 2.94)), coronary heart disease (1.18 (1.09, 1.29)), and renal disease (1.42 (1.32, 1.53)). No association between allopurinol and mortality, joint replacement, cerebrovascular disease, and peripheral vascular disease were found.

All treatment effect estimates were then adjusted for cumulative allopurinol use as there were large differences (SMD >0.25) between treatment groups; gout consultation was also adjusted for in the analysis of mortality as this covariate was largely imbalanced between treatment groups as well (SMD >0.25). Across all outcomes, adjusted HRs estimates either increased or decreased but the standard errors increased compared with unadjusted results. For example, for target SU level HR increased from 4.73 to 5.01 and standard error increased from 0.47 to 0.57.

Treatment effect estimates were then further adjusted for all covariates whose overall SMD >0.10. There was no/little change in HR for target SU level and mortality; HRs decreased further for gout hospitalisation (2.49 vs. 2.22), cerebrovascular disease (1.07 vs. 1.01), coronary heart disease (1.20 vs. 1.11), peripheral vascular disease (1.23 vs. 1.13); and renal disease (1.43 vs. 1.27).

Table 10.4: Treatment effect estimation of allopurinol

Outcome	No allopurinol N (%)	Allopurinol N (%)	Weight Mean (SD) Range	Unadjusted HR (95% CI) Standard error	Unadjusted MSM HR (95% CI) Standard error	*Adjusted MSM HR (95% CI) Standard error	**Adjusted MSM HR (95% CI) Standard error
Target SU level			0.99 (0.68)	4.63 (3.95, 5.42)	4.73 (3.89, 5.76)	5.01 (4.02, 6.25)	5.00 (4.00, 6.23)
Not reached target	6,805 (96.5)	2,515 (84.5)	<0.01, 16.13	0.37	0.47	0.57	0.56
Reached target	247 (3.5)	460 (15.5)					
Mortality			1.00 (0.81)	1.19 (1.12, 1.26)	0.96 (0.87, 1.06)	0.91 (0.80, 1.04)	0.93 (0.83, 1.06)
Alive	106,338 (97.0)	44,017 (96.4)	<0.01, 85.67	0.04	0.05	0.06	0.06
Death	3.346 (3.1)	1.630 (3.6)					
Gout hospitalisation			1.00 (0.41)	2.92 (2.67, 3.19)	2.66 (2.40, 2.94)	2.49 (2.15, 2.87)	2.22 (1.91, 2.58)
No	87,694 (99.0)	32,402 (96.9)	<0.01, 16.99	0.13	0.14	0.18	0.17
Yes	923 (1.0)	1,049 (3.1)					
Joint replacement			1.00 (0.46)	1.14 (0.99, 1.30)	1.14 (0.99, 1.32)	0.98 (0.81, 1.20)	0.93 (0.76, 1.13)
No	104,135 (99.3)	42,731 (99.3)	<0.01, 26.40	0.08	0.08	0.10	0.09
Yes	692 (0.7)	323 (0.8)					
Cerebrovascular disease			1.00 (0.45)	0.98 (0.88, 1.10)	1.03 (0.90, 1.19)	1.07 (0.87, 1.31)	1.01 (0.82, 1.25)
No	101,655 (98.9)	41,901 (99.0)	<0.01, 23.86	0.06	0.07	0.11	0.11
Yes	1,114 (1.1)	438 (1.0)					
Coronary heart disease			0.99 (0.51)	1.18 (1.09, 1.28)	1.18 (1.09, 1.29)	1.20 (1.08, 1.34)	1.11 (1.01, 1.23)
No	82,268 (97.5)	31,438 (97.3)	<0.01, 31.06	0.05	0.05	0.07	0.06
Yes	2,080 (2.5)	870 (2.7)					
Peripheral vascular disease			1.00 (0.45)	1.19 (1.01, 1.41)	1.20 (1.00, 1.44)	1.23 (0.95, 1.58)	1.13 (0.87, 1.46)
No	105,741 (99.6)	43,775 (99.5)	<0.01, 25.55	0.10	0.11	0.16	0.15
Yes	432 (0.4)	203 (0.5)					
Renal disease			0.99 (0.45)	1.42 (1.33, 1.52)	1.42 (1.32, 1.53)	1.43 (1.27, 1.61)	1.27 (1.13, 1.43)
No	96,077 (97.4)	35,986 (96.3)	<0.01, 26.43	0.05	0.05	0.08	0.07
Yes	2,541 (2.6)	1,384 (3.7)					

*Adjusted for imbalanced covariates with SMD>0.25; **Adjusted for imbalanced covariates with SMD>0.10; HR: Hazard ratio; CI: Confidence interval; MSM: Marginal structural models; SU: Serum urate;

10.4 Sensitivity analyses

10.4.1 Various PS model specifications

As previously seen in Chapter 9, treatment effect estimates may change depending on which covariates were included in the PS model. In this analysis, four different PS models were considered in addition to the PS model considered above:

Model specification 1: Main effects PS model with linear terms for continuous covariates (as described above).

Model specification 2: Main effects PS model with fractional polynomial terms of dimension 2 (FP2) terms for continuous covariates.

Model specification 3: Main effects PS model with fractional polynomials of dimension 1 (FP1) terms for continuous covariates.

Model specification 4: Main effects PS model with FP terms that reduced weight variability the most.

Model specification 5: Main effects PS model with FP terms and two-way interaction terms that reduced weight variability the most.

Choice of FP terms and interactions terms was based on which terms reduced SD of weights the most. Appendix N shows the PS model specification that was used to estimate the probability of initiating allopurinol and the probability of continuing with allopurinol. For outcomes target SU level, coronary heart disease, gout hospitalisation, no interaction terms were identified that reduced the SD of weights.

Table 10.5 shows the distribution of the weights and the unadjusted treatment effect in the weighted study sample for each PS model. In comparison with model specification 1 (the main analysis), adding FP terms and/or interaction terms did not cause the mean weight to deviate

from one. However, for the majority of outcomes, except for mortality and gout hospitalisation, the SD of weight increased for model specification 2, for example for joint replacement, SD increased from 0.46 to 1.27.

Treatment effects were then estimated. Note that these estimates were not adjusted for any covariates. For target SU level, estimated treatment effect differed slightly in model specifications 2-4 compared with the main analysis; estimated HRs were lower ranging between 4.60 and 4.69 compared with 4.73 obtained from the main analysis; standard errors were larger in model specifications 2 (SE=0.52) and 3 (SE=0.59) compared with the main analysis (SE=0.47), although model specification 4 yielded the same standard error (0.47).

Similarly, for secondary outcomes, estimated treatment effects in model specifications 2-5 differed slightly to the main analysis and standard errors were comparable to the main analysis. Overall, treatment effect estimates were robust the choice of PS model.

Table 10.5: Distribution of weights and treatment effect estimation of allopurinol

	Model specification 1	Model specification 2	Model specification 3	Model specification 4	Model specification 5
Target SU level					
Weight (mean (SD))	0.99 (0.68)	1.00 (0.79)	1.00 (0.85)	0.99 (0.64)	-
MSM Hazard ratio (95% CI)	4.73 (3.89, 5.76)	4.60 (3.69, 5.74)	4.69 (3.66, 6.00)	4.66 (3.82, 5.69)	
Standard error	0.47	0.52	0.59	0.47	
Mortality					
Weight (mean (SD))	1.00 (0.81)	1.01 (0.86)	1.01 (0.91)	0.99 (0.75)	0.99 (0.70)
MSM Hazard ratio (95% CI)	0.96 (0.87, 1.06)	0.98 (0.89, 1.08)	1.02 (0.92, 1.12)	1.02 (0.93, 1.11)	1.03 (0.95, 1.13)
Standard error	0.05	0.05	0.05	0.05	0.05
Gout hospitalisation					
Weight (mean (SD))	1.00 (0.41)	1.00 (0.42)	1.00 (0.46)	1.00 (0.41)	-
MSM Hazard ratio (95% CI)	2.66 (2.40, 2.94)	2.70 (2.43, 3.00)	2.70 (2.43, 3.00)	2.67 (2.41, 2.95)	
Standard error	0.14	0.14	0.15	0.14	
Joint replacement					
Weight (mean (SD))	1.00 (0.46)	1.01 (1.27)	1.00 (0.50)	1.00 (0.45)	1.00 (0.44)
MSM Hazard ratio (95% CI)	1.14 (0.99, 1.32)	1.09 (0.93, 1.28)	1.06 (0.91, 1.24)	1.15 (0.99, 1.32)	1.15 (0.99, 1.32)
Standard error	0.08	0.09	0.08	0.08	0.08
Cerebrovascular disease					
Weight (mean (SD))	1.00 (0.45)	1.01 (1.34)	1.00 (0.50)	1.00 (0.45)	1.00 (0.44)
MSM Hazard ratio (95% CI)	1.03 (0.90, 1.19)	1.00 (0.87, 1.15)	0.99 (0.87, 1.12)	1.04 (0.91, 1.18)	1.03 (0.91, 1.18)
Standard error	0.07	0.07	0.06	0.07	0.07
Coronary heart disease					
Weight (mean (SD))	0.99 (0.51)	1.01 (1.38)	1.00 (0.53)	0.99 (0.51)	-
MSM Hazard ratio (95% CI)	1.18 (1.09, 1.29)	1.12 (1.01, 1.23)	1.16 (1.06, 1.26)	1.16 (1.06, 1.26)	
Standard error	0.05	0.06	0.05	0.05	
Peripheral vascular disease					
Weight (mean (SD))	1.00 (0.45)	1.01 (1.42)	1.00 (0.50)	1.00 (0.45)	1.00 (0.44)
MSM Hazard ratio (95% CI)	1.20 (1.00, 1.44)	1.13 (0.94, 1.35)	1.13 (0.94, 1.36)	1.19 (1.00, 1.42)	1.18 (0.98, 1.41)
Standard error	0.11	0.11	0.11	0.11	0.11
Renal disease					
Weight (mean (SD))	0.99 (0.45)	1.01 (1.72)	0.99 (0.49)	0.99 (0.44)	0.99 (0.41)
MSM Hazard ratio (95% CI)	1.42 (1.32, 1.53)	1.33 (1.18, 1.49)	1.41 (1.30, 1.52)	1.39 (1.29, 1.51)	1.39 (1.29, 1.51)
Standard error	0.05	0.08	0.06	0.06	0.06

CI: Confidence interval; SD: Standard deviation; SU: Serum urate

10.4.2 Weight truncation

Sensitivity analysis was performed truncating weights at 0.5% of its percentile (Table 10.6). A low percentile was chosen as weights were not extreme. Across all outcome analyses, mean weight remained the same and SD decreased. For example, for mortality, the maximum weight value reduced from 85.67 to 4.57 and SD of weights reduced from 0.81 to 0.58. Estimated treatment effects using truncated weights changed very little compared to the main analysis and standard errors were very similar as well.

Table 10.6: Distribution of truncated weights and its impact on treatment effect estimation of allopurinol

	0% weight truncation		0.5% weight truncation	
Outcome	Weight Mean (SD) Range	MSM HR (95% CI) Standard error	Weight Mean (SD) Range	MSM HR (95% CI) Standard error
Target SU level	0.99 (0.68) <0.01, 16.13	4.73 (3.89, 5.76) 0.47	0.98 (0.60) 0.08, 4.25	4.72 (3.91, 5.69) 0.45
Mortality	1.00 (0.81) <0.01, 85.67	0.96 (0.87, 1.06) 0.05	0.98 (0.58) 0.05, 4.57	1.04 (0.96, 1.13) 0.04
Gout hospitalisation	1.00 (0.41) <0.01, 16.99	2.66 (2.40, 2.94) 0.14	0.99 (0.33) 0.14, 2.94	2.65 (2.41, 2.93) 0.13
Joint replacement	1.00 (0.46) <0.01, 26.40	1.14 (0.99, 1.32) 0.08	0.99 (0.33) 0.12, 2.82	1.16 (1.00, 1.33) 0.08
Cerebrovascular disease	1.00 (0.45) <0.01, 23.86	1.03 (0.90, 1.19) 0.07	0.99 (0.34) 0.10, 2.97	1.01 (0.90, 1.14) 0.06
Coronary heart disease	0.99 (0.51) <0.01, 31.06	1.18 (1.09, 1.29) 0.05	0.99 (0.40) 0.07, 3.42	1.17 (1.07, 1.27) 0.05
Peripheral vascular disease	1.00 (0.45) <0.01, 25.55	1.20 (1.00, 1.44) 0.11	0.99 (0.34) 0.13, 2.88	1.19 (0.99, 1.42) 0.11
Renal disease	0.99 (0.45) <0.01, 26.43	1.42 (1.32, 1.53) 0.05	0.99 (0.32) 0.09, 2.74	1.40 (1.30, 1.50) 0.05

HR: Hazard ratio; MSM: Marginal structural models; SD: Standard deviation; SU: Serum urate

10.4.3 Missing data

Complete case analysis was performed for all outcomes (Table 10.7). Missing data was present in alcohol consumption, BMI, smoking status, and SU level (SU level was only missing for secondary outcomes). For the analysis of target SU level, 52% of the study sample had complete data. For the secondary outcomes, approximately 20% of the study sample had complete data.

Compared with the main analysis, weights estimated in complete case analysis were found to have mean weight of 1 although, they were more variable across most outcomes; for joint replacement, the weights had become large with the maximum weight value increasing to 92.57 whilst for other outcomes (mortality, gout hospitalisation, and renal disease), the maximum weight value had decreased.

Compared with the main analysis, the direction of HRs and statistical significance remained the same for two outcomes although the magnitude of HRs decreased; for target SU the HR reduced from 4.73 to 4.42 and for gout hospitalisation the HR reduced from 2.66 to 2.47. The HR for coronary heart disease had reduced from 1.18 to 1.01 and lost statistical significance compared with the main analysis. The HRs for renal disease were similar between the main analysis and complete case analysis.

Although allopurinol was not associated with four outcomes in the main analysis, the direction of the HRs changed in complete case analysis but remained insignificant for three of the outcomes; for mortality HR had changed from 0.96 to 1.28; for joint replacement the HR changed from 1.14 to 0.85; for cerebrovascular disease the HR had increased from 1.03 to 1.92. In contrast, for peripheral vascular disease, the HR had doubled and had become statistically significant (2.44 (1.28, 4.67)).

As expected, due to reduced sample size, standard errors of all treatment effect estimates were larger compared to the main analysis. The largest increase in standard error was observed for target SU level (0.47 vs. 1.00) and cerebrovascular disease (0.07 vs. 0.86).

Table 10.7: Impact of missing data on treatment effect estimates of allopurinol

Outcome	Main analysis				Complete case analysis			
	Sample size	Weight Mean (SD) Range	Unadjusted HR (95% CI) Standard error	Unadjusted MSM HR (95% CI) Standard error	Sample size	Weight Mean (SD) Range	Unadjusted HR (95% CI) Standard error	Unadjusted MSM HR (95% CI) Standard error
Target SU level	1,742	0.99 (0.68) <0.01, 16.13	4.63 (3.95, 5.42) 0.37	4.73 (3.89, 5.76) 0.47	909	1.00 (0.84) <0.01, 15.63	4.32 (3.02, 6.18) 0.79	4.42 (2.84, 6.88) 1.00
Mortality	16,876	1.00 (0.81) <0.01, 85.67	1.19 (1.12, 1.26) 0.04	0.96 (0.87, 1.06) 0.05	3,609	0.99 (1.10) <0.01, 39.93	1.37 (1.11, 1.70) 0.15	1.28 (0.91, 1.80) 0.22
Gout hospitalisation	14,087	1.00 (0.41) <0.01, 16.99	2.92 (2.67, 3.19) 0.13	2.66 (2.40, 2.94) 0.14	3,044	0.99 (0.46) <0.01, 7.92	2.72 (1.98, 3.74) 0.44	2.47 (1.70, 3.60) 0.47
Joint replacement	16,644	1.00 (0.46) <0.01, 26.40	1.14 (0.99, 1.30) 0.08	1.14 (0.99, 1.32) 0.08	3,555	1.00 (1.30) 0.01, 92.57	0.96 (0.59, 1.57) 0.24	0.85 (0.48, 1.50) 0.25
Cerebrovascular disease	16,253	1.00 (0.45) <0.01, 23.86	0.98 (0.88, 1.10) 0.06	1.03 (0.90, 1.19) 0.07	3,473	0.99 (0.67) 0.01, 38.43	1.10 (0.73, 1.65) 0.23	1.92 (0.79, 4.63) 0.86
Coronary heart disease	14,063	0.99 (0.51) <0.01, 31.06	1.18 (1.09, 1.28) 0.05	1.18 (1.09, 1.29) 0.05	2,802	0.99 (0.70) <0.01, 26.22	1.09 (0.81, 1.46) 0.16	1.01 (0.70, 1.44) 0.18
Peripheral vascular disease	16,519	1.00 (0.45) <0.01, 25.55	1.19 (1.01, 1.41) 0.10	1.20 (1.00, 1.44) 0.11	3,517	1.00 (0.79) <0.01, 45.58	2.41 (1.33, 4.39) 0.74	2.44 (1.28, 4.67) 0.81
Renal disease	16,508	0.99 (0.45) <0.01, 26.43	1.42 (1.33, 1.52) 0.05	1.42 (1.32, 1.53) 0.05	3,512	0.99 (0.52) <0.01, 16.61	1.61 (1.31, 1.97) 0.17	1.44 (1.10, 1.89) 0.20

CI: Confidence interval; SD: Standard deviation; HR: Hazard ratio; MSM: Marginal structural models; SU: Serum urate

10.4.4 Unmeasured confounding variables

Sensitivity of treatment effect estimates (from the main analysis) against unmeasured confounding was assessed. E-values were derived which measures the minimum association required between an unmeasured confounding variable with treatment and outcome, to explain away (or nullify) the treatment effect; E-values were also calculated such that the confidence interval for the treatment effect estimate contains the null value. E-values are measured on the risk ratio (RR) scale.

E-values were computed for target SU level, gout hospitalisation, coronary heart disease, and renal disease and are presented in Table 10.8. The largest association on the RR scale between a measured covariate with outcome, allopurinol initiation, and allopurinol continuation are also presented.

For target SU level, to nullify the estimated HR of 4.73, an E-value of 5.15 was required and an E-value of 4.47 for the confidence interval to contain 1. The largest RR was observed between renal disease and outcome (RR 1.53); colchicine and allopurinol initiation (RR 1.61); gout consultation and allopurinol continuation (RR 1.61). As the E-values were larger than the observed RRs, it is unlikely an unmeasured confounding variable exists that is associated with both treatment and outcome with a RR of least 4.47, to cause the treatment effect estimate to lose statistical significance.

Similarly for gout hospitalisation, an E-value of 4.76 was required to nullify the treatment effect estimate of 2.66, and an E-value of 4.23 in order for the confidence interval to contain 1. The largest associations between any covariate with gout hospitalisation, allopurinol initiation, and allopurinol continuation ranged from 1.34 to 2.41. Therefore, as the E-value was outside of this range, it is unlikely a strong confounding variable exists that could cause the treatment effect estimate to lose statistical significance.

For coronary heart disease and renal disease, smaller E-values of 1.49 and 1.87 respectively, were required to nullify the estimated treatment effects; also, E-values of 1.32 and 1.72 respectively, would be required to cause the confidence interval to include the null value. The RRs observed between covariate and outcome (coronary heart disease/renal disease), allopurinol initiation, and allopurinol continuation ranged from 1.39 to 3.31. Therefore, it is likely an unmeasured confounding variable exists that could cause the estimated treatment effects to lose statistical significance.

Table 10.8: Assessment of unmeasured confounding using E-values

Outcome	MSM HR (95% CI)	E-value RR (95% LCl)	Largest RR between (covariate) and outcome	Largest RR between (covariate) and allopurinol initiation	Largest RR between (covariate) and allopurinol continuation
Target SU level	4.73 (3.89, 5.76)	5.15 (4.47)	1.53 (renal disease)	1.61 (colchicine use)	1.22 (gout consultation)
Gout hospitalisation	2.66 (2.40, 2.94)	4.76 (4.23)	2.41 (renal disease)	2.25 (colchicine use)	1.34 (NSAID use)
Coronary heart disease	1.18 (1.09, 1.29)	1.49 (1.32)	2.29 (diuretic use)	2.32 (colchicine use)	1.41[a] (no hypertension)
Renal disease	1.42 (1.32, 1.53)	1.87 (1.72)	3.31 (diuretic use)	2.30 (colchicine use)	1.39[a] (no diabetes)

[a]RR was <1.00 thus the inverse was taken and presented; HR: Hazard ratio; CI: Confidence interval; LCI: Lower bound of the confidence interval: MSM: Marginal structural models; RR: Risk ratio; SU: Serum urate

10.5 Summary

This chapter aimed to estimate the causal association between allopurinol use and long-term outcomes using MSM. In this analysis, the associations between covariates and treatment initiation and continuation were modelled separately.

This analysis had shown allopurinol users had higher chance of reaching target SU level and had higher risk of gout hospitalisation, coronary heart disease and renal disease than non-users. There was no evidence allopurinol use was associated with mortality, joint replacement, cerebrovascular disease, and peripheral vascular disease. The estimated treatment effects were robust to various PS model specifications and large weights, indicating that the PS

models were correctly specified. The treatment effect estimates from complete case analysis differed in terms of magnitude, direction, and statistical significance across outcomes, although this was primarily due to analysing a much smaller sample and selection bias.

Treatment estimates for target SU level and gout hospitalisation were likely to be robust against unmeasured confounding as strong confounding variables would be required for the treatment estimates to lose statistical significance. However, the same cannot be said for coronary heart disease and renal disease that will require weak confounding to explain away the treatment effect estimates.

Comparison with previous studies

Although there are no comparable studies that had evaluated the time-varying effect of allopurinol on outcome, one study had evaluated the changes in SU level on outcome in gout using MSM (Desai et al., 2018). That study is of interest due to the direct effects of allopurinol on lowering SU level that are subsequently expected to lower the risk of poor outcome. Desai et al. (2018) found per 3 mg/dL reduction in SU level was not associated with cardiovascular disease (coronary heart and cerebrovascular diseases) (HR 1.01 (0.81, 1.27)) but was associated with renal function decline (HR 0.89 (95% CI 0.81, 0.98)). The HRs differed to the HRs estimated in this analysis but that could be due to differing outcome and exposure definitions, different study population (USA), data was obtained from health insurance claims, and included more covariates in PS estimation.

Residual confounding

In the main analysis, it was found in the weighted study sample covariate balance was not achieved for all covariates in all follow-up periods. Some residual confounding was expected as the associations between covariates and patients initiating treatment for the first time and patients resuming treatment were assumed to be the same when they may instead differ; the

same assumption was also made between patients discontinuing treatment for the first time versus those discontinuing treatment repeatedly. However, using separate PS models to allow these associations to differ may lead to larger standard error in the treatment effects (Platt et al., 2013). Other sources of residual confounding will be from incomplete adjustment for lifestyle factors due to missing data, unable to model SU level as a time-varying covariate as it was infrequently measured, and under reporting of hypertension and renal disease found in this dataset.

Estimating the probability of initiating and continuing with allopurinol separately vastly improved the distribution of weights compared with weights estimated in Chapter 9. Estimated treatment effects were more in line with treatment effects estimated from baseline and time-invariant PS subclassification.

11 Discussion

Electronic health records (EHR) enable real life patterns of drug prescribing and other aspects of healthcare to be captured over a long period of time. This in turn allows for the possibility of treatment effects to be estimated under conditions of time-varying treatment and covariates. However, this may give rise to time-varying confounding as covariates may predict future treatment and outcome over time. Controlling for time-varying covariates that are affected by past treatment can be challenging in practice as one may inadvertently adjust out the effect of treatment that is mediated through covariates, resulting in biased effects.

The work presented in this thesis employed two advanced propensity score (PS) based approaches, namely PS subclassification and marginal structural models (MSM), to complex EHR data from the Clinical Practice Research Datalink (CPRD), to accurately estimate effects of allopurinol on a range of outcomes in patients with gout. Multiple clinical and statistical decisions have had to be made at various stages of data preparation and analysis. Sensitivity of the findings to various assumptions was assessed and some important implementation challenges have been identified.

This chapter summarises key findings and highlights the strengths and limitations of this work. Potential future research and clinical implications are also discussed.

11.1 Summary of key findings

Literature review

The literature review conducted in October 2014 was performed to identify published observational studies evaluating the effect of allopurinol on gout outcomes, and to establish the range of study designs and statistical methods used to control for confounding variables

(Chapter 3). The review highlighted a small number of studies that had used primary care EHR to evaluate effects of allopurinol. Studies tended to measure allopurinol and covariates at a single time point and the methods used to account for these covariates were generally suboptimal. No studies modelled time-varying allopurinol and covariates, nor had considered assessing sensitivity of treatment effect estimates to modelling assumptions made, such as absence of unobserved confounding.

Given that considerable time has elapsed since the initial review was conducted and there have since been published studies highlighting the benefits of using EHR for research (Herrett et al., 2015), the literature review was updated in June 2020. A further 56 studies were identified that evaluated effect of allopurinol, of which five studies had used primary care EHR data (Abdul Sultan et al., 2017, Sultan et al., 2019, Sultan et al., 2018, Roughley et al., 2018, Vargas-Santos et al., 2018). Twenty-two studies analysed EHR from insurance claims. PS matching at baseline to create comparable treatment groups was used in 11 studies. One study had used inverse probability treatment weights that were estimated using generalised boosted regression models to account for confounding (Chung et al., 2019). Some studies had reported results for outcomes not considered in earlier similar studies, such as hepatoxicity (Lee et al., 2019), dementia (Singh and Cleveland, 2018b), cancer (Shih et al., 2017, Chen et al., 2016), erectile dysfunction (Abdul Sultan et al., 2017), and fracture (Sultan et al., 2018, Tzeng et al., 2016). Kang et al. (2019) had used competing risk models to account for the competing risk of death in the evaluation of non-fatal cardiovascular events. Few studies had censored patient follow-up when treatment status had changed (Vargas-Santos et al., 2018, Foody et al., 2017, Kang et al., 2019, Kim et al., 2018, Zhang et al., 2018). Although studies had stated unmeasured confounding was a limitation, no studies had formally re-estimated the treatment effect.

No studies were identified that used advanced methods to evaluate the time-varying effect of allopurinol on outcome in the presence of time-varying confounding. Hence, new findings from this thesis adds to the current literature.

Baseline PS subclassification

Chapter 7 estimated the conditional effect of initiating allopurinol on outcome using the one-year landmark method. Initial descriptive statistics showed allopurinol users more commonly had SU level above target, coronary heart disease, and renal disease, and were prescribed pain relief and diuretics more compared to non-users. These differences between treatment groups suggested confounding by indication could be present.

Baseline confounding was controlled by creating homogenous subclasses via PS subclassification and by adjusting for remaining imbalanced covariates in outcome analysis. PS subclassification was relatively straightforward to implement. Estimated PS distribution was satisfactory with adequate common support between treatment groups and balance was achieved for the majority of covariates within subclasses.

Allopurinol was shown to be associated with higher chance of reaching target serum urate (SU) level ≤360μmol/L and fewer number of primary care gout consultations, while it also increased risk of premature mortality, gout hospitalisation, coronary heart disease, and renal disease compared with non-use. The magnitude of the effect of allopurinol differed across subclasses for gout consultations and gout hospitalisation. No association was observed between allopurinol and joint replacement, cerebrovascular disease, and peripheral vascular disease.

Analyses were subsequently stratified by presence of renal disease and severe hyperuricaemia. It was not possible to reliably estimate treatment effect estimates among those with renal disease due to very low sample size. However, among those without renal

disease, allopurinol remained associated with higher chance of reaching target SU level and fewer number of primary care gout consultations, and increased risk of gout hospitalisation, coronary heart disease, and renal disease than non-use, and now also with a higher risk of joint replacement; the direction and magnitude of the hazard ratios were similar to the estimated hazard ratios in the whole study sample analysis.

Allopurinol was associated with reaching target SU level, gout hospitalisation, joint replacement, coronary heart disease and fewer gout consultations in those with non-severe hyperuricaemia, but only reaching target SU level and fewer gout consultations in those with severe hyperuricaemia.

The key limitation of baseline PS subclassification approach was that PS was estimated as the probability of allopurinol at baseline conditional on covariates at baseline, thus assuming that these measures remained constant throughout follow-up. These assumptions may be unrealistic; the one-year landmark method used for the main analyses did not capture the majority of patients consulting for gout that were prescribed allopurinol later on. Using a two-year landmark period did not change treatment effect estimates substantially, suggesting that future similar studies analysing effect of allopurinol initiation following gout diagnosis should consider a longer landmark period. However, this simplistic approach provided a base set of results, against which estimates based on more elaborate methods could be compared.

Time-varying PS subclassification

Chapter 8 extended PS subclassification to a repeated measures setting that captured changes in allopurinol status and covariates over time, thus estimating the conditional effect of actual treatment. Descriptive analyses showed the prevalence of key comorbidities that are indications for allopurinol (renal disease and hypertension) had increased over time and to a smaller extent colchicine use although there was a reduction in prescriptions for NSAIDS. The

proportion of patients prescribed allopurinol was double that observed in the baseline analysis. It was also observed that once patients initiated allopurinol, 40% discontinued treatment, with nearly half of those patients subsequently restarting treatment. Differences in comorbidities between treatment groups identified in the baseline analysis were more pronounced and now included hypertension and obesity; differences in baseline SU level and medication use between treatment groups identified in baseline analysis persisted.

There were several issues encountered during implementation of time-varying PS estimation. The initial model estimated PS that near violated the positivity assumption with values close to zero in the majority of non-allopurinol intervals in the subclass with the lowest propensity for allopurinol. This resulted in poor common support of PS between treatment groups with few allopurinol intervals found in the subclass with the lowest PS, and no occurrence of outcome in that subclass. Various modifications, such as inclusion of interaction terms and non-linear terms, and omission of covariates, were made to the PS model however none resolved the issue. Reducing the number of subclasses from five to four and/or removing patients outside the region of common support prior to subclassification, still resulted in very few allopurinol intervals in the first subclass and no occurrence of outcome. With regards to covariate balance, overall balance was achieved across subclasses however within subclasses, a number of covariates were not balanced and had to be adjusted for in subclass-specific treatment effect estimation. Modifications made to the PS model and increasing the number of subclasses did little to improve covariate balance.

As a result, across all outcomes, the first subclass could not be used and was excluded from all outcome analyses. The implication of this was that not all repeated measurements from a patient were analysed. Intervals that were excluded contained observations from older patients and resided in less deprived areas, were female, had lower SU level, were less likely

to consult for gout, had fewer prescriptions for diuretics, and shorter mean duration of previous allopurinol use, than intervals that were included in analysis.

In terms of the findings, allopurinol was found to be associated with increased chance of reaching target SU level, and increased risk of gout hospitalisation, coronary heart disease, peripheral vascular disease and renal disease.

Compared with the baseline PS subclassification approach, some findings remained similar, however overall the estimated hazard ratios were different and standard errors were larger, to varying extent, across all outcomes. For example, the hazard ratios for target SU level and peripheral vascular disease doubled; effect of allopurinol on mortality lost statistical significance; cerebrovascular disease, coronary heart disease and renal disease now had treatment effects that varied across PS subclasses.

Overall application of time-varying PS subclassification method did not perform well in estimating PS, and consequently in the outcome analysis stage. The primary reason is suspected to be the use of random intercept (used to account for within patient repeated measurements) in the PS estimated model. Use of random effects is known to have better discrimination ability (i.e., allopurinol users and non-users have higher and lower estimated PS respectively) than logistic regression without random effects (Bouwmeester et al., 2013), which may have been the reason for poor common support of PS between treatment groups.

Random intercept terms had to be omitted from the estimation of subclass-specific treatment effects as models failed to converge successfully, potentially because adjustment for time component and/or adjustment for imbalanced covariates. Various specifications of variance-covariance structure were considered, and none resolved the issue of non-convergence. Therefore, in the end, robust standard errors were estimated instead to allow for correlated repeated measurements of patients within subclasses.

Marginal structural models

MSM were used to estimate the marginal treatment effect accounting for treatment and covariate histories. MSM were fitted under two different assumptions.

In Chapter 9, naïve scenario was considered, where the direction and magnitude of associations between covariates and allopurinol initiation and continuation were assumed to be the same. Only mortality outcome was considered. The initial PS model yielded extremely large stabilised weights with mean weight deviating from one. Consequently, unexpectedly low magnitude of the estimated treatment effect was observed, indicating that allopurinol use was strongly protective of premature mortality (which was not observed in previous chapters), although the standard error was smaller than the corresponding standard error derived from time-varying PS subclassification but larger via than baseline PS subclassification.

Although one primary care EHR study had shown allopurinol use was protective of premature mortality (Dubreuil et al., 2015), the estimated hazard ratios in this analysis was considerably lower and unexpected, as prescription for allopurinol dose is often low (100-300mg/day) and infrequently escalated to the optimal dose within primary care (Cottrell et al., 2013). Similar magnitude and direction of treatment effect estimates of allopurinol were also obtained for other secondary outcomes, though reporting of the results was restricted to mortality only in this analysis.

Various modifications to PS models were made in an attempt to reduce weight variability. Inclusion of non-linear terms for continuous covariates yielded the largest increase in the hazard ratio, thus estimating an even stronger protective effect against premature mortality although, standard error did reduce. Cumulatively including interaction terms between covariates to the PS model resulted in a decrease in hazard ratio towards the null and reduction in standard error that was similar to the standard error estimated in baseline PS

subclassification approach. Regardless of the choice of PS model, imbalance appeared in all covariates and worsened over time; direct adjustment for these imbalanced covariates in outcome analysis generally led to the hazard ratio moving further away from the null and increased standard error. In all instances, allopurinol use remained protective of premature mortality.

Various common approaches for reducing extremeness of weights were adopted. Using normalised weights and weight truncation for both stabilised and normalised weights yielded smaller hazard ratios that were closer to the null value, and smaller standard errors than the treatment effect estimates obtained from the main analysis. However, despite these remedial procedures, the issue of covariate imbalance between treatment groups remained, particularly after ten years of follow-up. When follow-up was truncated at 10 years, the treatment effect estimate moved closer to the null with a larger standard error compared with the main analysis. Notably, covariate balance was achieved over time; this may have been due to certain covariate combinations occurring after 10 years that may have been the cause of extreme weights initially observed not being captured. Furthermore, extreme weights that may have appeared in earlier follow-up were prevented to be cumulatively multiplied over a longer period that would have amplified extreme weights. Out of these various approaches used, weight truncation estimated the hazard ratios closest to the null and smallest standard errors although covariate balance could only be achieved with substantial weight truncation.

Lastly, intention-to-treat analysis that estimated the effect of initiating allopurinol and assumed patients remained on treatment until the end of follow-up was performed. Estimated stabilised weights were more satisfactory as there were no extreme weights and mean weight did not deviate from one. Estimated treatment effects was comparable with baseline PS subclassification although the standard error was slightly larger. This suggests estimating weights after treatment initiation may be complex.

Overall, it was concluded that the PS model was likely misspecified, primarily because of the assumption that the associations between covariates and allopurinol initiation and continuation were the same. There appeared to be a trade-off between bias and precision of the estimated treatment effect; use of different approaches to reduce extreme weights generally led to treatment effect estimates that were close to the null value (indicating increase in bias), while standard error decreased (increased precision).

In Chapter 10, the direction and magnitude of associations between covariates and allopurinol initiation and continuation were assumed to differ. This assumption was plausible as the associations between covariates and allopurinol continuation and discontinuation differed in magnitude, direction, and statistical significance. For example, hyperlipidaemia, hypertension, and diabetes had increased odds of continuing allopurinol compared with discontinuing allopurinol; in contrast, these same covariates had reduced odds of initiating allopurinol compared with non-use.

Estimating PS separately for patients initiating and continuing allopurinol improved weight estimation; the estimated weight distribution was not skewed, mean weight was close to one, and extreme weights were not present. Outcome analysis showed that allopurinol was associated with higher chance of reaching target SU level, and increased risk of gout hospitalisation, coronary heart disease, and renal disease. As a sensitivity analysis, truncating weights by a small percentage resulted in very small changes in the hazard ratios and standard errors for all outcomes suggesting any larger weights (although not extreme) did not have an undue impact on treatment effect estimates. Imbalance remained in some covariates and these were adjusted for in the outcome analysis; adjustment for large imbalanced covariates caused small changes in treatment effect estimates and small increase in standard errors, except for target SU level where hazard ratio and standard error increased by a large amount.

Although not statistically significant, treatment effect estimate had shown allopurinol generally increased risk of premature mortality in contrast to Chapter 9, where allopurinol was protective of mortality and had larger standard error due to larger weight variability.

Generally, similar conclusions regarding the effectiveness of allopurinol can be made based on application of MSM and time-varying PS subclassification approaches; allopurinol increased the chance of reaching target SU level, and increased risk of gout hospitalisation, coronary heart disease, and renal disease; however, PS subclassification had found allopurinol increased the risk of peripheral vascular disease as well. Magnitude of treatment effect estimates differed between the two methods; hazard ratios estimated from MSM were greater for target SU level and joint replacement, but smaller for peripheral vascular disease compared to estimates obtained via time-varying PS subclassification; as PS subclass-specific treatment effects could not be pooled for gout hospitalisation, cerebrovascular disease, coronary heart disease and renal disease, most estimates were smaller than the estimates in MSM, with the exception of one PS subclass that had greater estimates. Precision of treatment effect estimates resulting from MSM approach was generally smaller compared with time-varying PS subclassification.

Comparison of treatment effect estimates obtained from MSM and baseline PS subclassification (Chapter 7) are not comparable; MSM estimated the effect of actual treatment, whereas baseline PS subclassification estimate the effect of initiating treatment. Regardless, both methods found allopurinol increased the chance of reaching target SU level, and increased risk of gout hospitalisation, coronary heart disease, and renal disease. Notably, the largest difference observed between the two methods were that estimated hazard ratios for target SU level and gout hospitalisation in MSM were double of the hazard ratios obtained from baseline analysis, and also had larger standard errors. Overall, no reliable conclusions can be made regarding comparison of point and precision estimates between the two

methods as similarities and differences were dependent upon frequency of outcome events, consideration of time-invariant and time-varying treatment and covariates, and how soon outcome occurs after receiving treatment, and how confounding effects were accounted for.

Missing data

Missing data were present for body mass index, alcohol consumption, smoking status and baseline SU level. For the main analysis, the missing indicator method (MIM) and the last observation carried forward (LOCF) were utilised allowing the entire study sample to be analysed. Complete case analysis was performed restricting analysis to patients who did not have missing data in those covariates across all three methods to assess sensitivity of treatment effect estimates regarding target SU level and mortality in (baseline and time-varying) PS subclassification, and then across all outcomes in MSM.

In complete case analysis for target SU level, the estimated hazard ratio had reduced in baseline PS subclassification whereas in time-varying PS subclassification and MSM, the hazard ratios were greater than the hazard ratios obtained from analysing the whole study sample. However, significance of estimates of allopurinol effect remained throughout. In complete case analysis for mortality, hazard ratio resulting from baseline PS subclassification approach, had reduced compared to the hazard ratio estimated from analysing the whole study sample, and significance was lost. In contrast, hazard ratios obtained via time-varying PS subclassification and MSM increased in complete case analysis. Standard errors of treatment effect estimates across all three methods had increased as expected due to smaller sample available for analysis; however, the increase in standard error was the greatest in MSM, followed by time-varying PS subclassification and then baseline PS subclassification.

Larger differences in treatment effect estimates obtained from complete case analysis and the whole study sample analysis was found in MSM and time-varying subclassification

compared with baseline PS subclassification. This may be due to using MIM over time with LOCF that may have biased treatment effect estimates more than using MIM at baseline. However, treatment effect estimates from complete case analysis may be biased due to analysing a more selective group of patients not representative of the whole study sample thus introducing selection bias.

Complete case analysis was performed for the remaining outcomes in MSM only. Allopurinol remained associated with greater chance of reaching target SU level and greater risk of gout hospitalisation, peripheral vascular disease and renal disease; statistical significance was lost for coronary heart disease. Hazard ratios were smaller in complete case analysis for gout hospitalisation, joint replacement and coronary heart disease, and were larger for the remaining outcomes compared to the hazard ratios estimated in the whole study sample. As before, all standard errors were higher in complete case analysis.

Unmeasured confounding

PS-based methods make a strong assumption of no unmeasured confounding which in practice is unlikely to be satisfied. Impact of confounding was considered for baseline PS subclassification, for target SU level and mortality. In MSM, it was considered for outcomes that had statistically significant associations with allopurinol. Unmeasured confounding was not considered for time-varying PS subclassification as the approach had to exclude some repeated measurements within patients from analysis in treatment effect estimation. Therefore, differences in the treatment effect obtained in the main analysis compared with the revised treatment effect in the presence of unmeasured confounding will be attributed to both unmeasured confounding and residual confounding from exclusion of repeated measurements from analysis.

In baseline PS subclassification, treatment effect estimates were revised in light of an unmeasured binary covariate with differing prevalence rates within treatment groups using the approach by Lin et al. (1998). Regardless of the distribution of an unmeasured binary confounding variable with treatment, allopurinol use remained associated with increased chance of reaching target SU level; treatment effect estimate was robust to unmeasured confounding as the magnitude of the hazard ratio was smaller but remained large and statistically significant. However, for mortality, treatment effect estimate lost significance in the assumed presence of an unmeasured binary confounding variable that had small difference in prevalence rates between treatment groups. When the difference of prevalence rates of the unmeasured confounding variable was large between treatment groups, allopurinol was shown it could be protective of premature mortality. The disadvantage of this method is that it does not consider the association between an unmeasured confounding variable and outcome, and the standard error of treatment effect assumed to be constant.

Within MSM, E-value approach to assessing unmeasured confounding was used. E-value quantifies the association (risk ratio) between an unmeasured confounding variable with treatment and outcome that would be required to explain away the treatment effect. As large hazard ratios were estimated for target SU level and gout hospitalisation in the main analysis, a strong association of at least 4 on the risk ratio scale between an unmeasured confounding variable and allopurinol would be required to explain away or nullify the hazard ratios; such an unmeasured confounding variable is unlikely to exist as the strongest observed association (risk ratio) between any covariate with allopurinol and outcome was 2.4. In contrast, for coronary heart disease and renal disease, as the estimated hazard ratios were smaller, a weaker association of at least 1.67 (risk ratio) between an unmeasured confounding variable with outcome and treatment was required to explain away the treatment effect; thus these estimates may be non-robust.

Evidently, generally larger estimated treatment effects require stronger unmeasured confounding variables to explain away the treatment effect. This thesis had considered strong confounding variables such as renal disease and baseline SU level thus finding another strong single confounding variable is unlikely; however, in the presence of cumulative confounding effects from two or more unmeasured confounding variables could potentially explain away small treatment effects.

Comparison with previous research

Published literature mostly considered allopurinol use at baseline thus the results are only comparable with baseline PS subclassification. As discussed in Section 7.7.1, there are few published EHR studies that had used comparable methods that were used in this thesis. Compared to published studies that had used EHR from CPRD and the one-year landmark method, estimated hazard ratios for mortality and renal disease in this thesis were larger and statistically significant although precision was similar (Kuo et al., 2015a, Roughley et al., 2018). Kuo et al. (2018) had used a case-control study design with CPRD data and concluded allopurinol use was not associated with joint replacement. One study using EHR from a Taiwan administrative database had concluded allopurinol increased risk of coronary heart disease but not cerebrovascular disease, the same conclusions derived in this thesis albeit with smaller hazard ratios and standard errors (Kok et al., 2014). In contrast, a study using EHR from an American administrative database had shown allopurinol was protective of peripheral vascular disease (Singh and Cleveland, 2018a), which was not verified in this thesis. No existing studies using EHR databases have evaluated outcomes target SU level, repeated gout consultations and gout hospitalisation. A RCT evaluating efficacy of nurse-led care vs. general practitioner (GP)-led care, found nurse-led care achieved greater number of patients achieving target SU level and reduced gout flares over two years than GP-led care (Doherty et al., 2018). Other smaller studies have shown allopurinol was associated with greater chance of reaching

target SU level (Dalbeth et al., 2006, Roddy et al., 2007b) and reduced flares potentially leading to fewer consultations (Neogi et al., 2014, Zhang et al., 2012, Zhang et al., 2014) the same conclusions as derived in this thesis. However, this thesis found allopurinol use was associated with higher risk of gout hospitalisation unlike a small study case-control study in New Zealand that found allopurinol use had lower odds of hospitalisation (Hutton et al., 2009).

As stated above, no studies had evaluated the effect of time-varying allopurinol; few studies had censored follow-up when treatment had changed or switched treatment groups (Vargas-Santos et al., 2018, Kim et al., 2013b). However, one study had evaluated the association between changes in SU level and risk of renal disease and cardiovascular disease in people with gout using MSM (Desai et al., 2018). This study is of relevance to this thesis as changes in SU level due to taking allopurinol were expected to have an effect of outcome. That study found reduction in SU level was associated with declining renal function but was not associated with cardiovascular disease.

11.2 Strengths, limitations, and future research

Strengths

There are several strengths of the work presented in this thesis to note. This was the first retrospective cohort study that used complex methods to model time-varying treatment effect using a large UK primary care EHR database on a range of clinically important long-term outcomes in a heterogeneous group of patients with gout likely to be representative of those seen in day-to-day clinical practice. The study sample included incident and prevalent cases of gout to maximise generalisability of treatment effect estimates to the UK gout population.

Several aspects of how such data could be used for treatment-effectiveness research, particularly how to incorporate time dependency of covariates and treatment during data preparation, were discussed. Careful consideration was given to ensure temporal ordering between covariates, treatment and outcome across all methods employed, by using the landmark approach to define baseline period and subsequently splitting the follow-up into equally spaced intervals within which these measures can be ascertained. Initial baseline analysis employed the one-year landmark period which was extended to two years to capture more patients prescribed allopurinol. This extension made little difference to treatment effect estimates. Within time-varying PS subclassification, initially time was subdivided into six-month intervals which were then extended to yearly intervals. Decision to use one-year intervals for measurement of covariates and treatment was based on achieving a balance between increased computational intensity, that would result from use of small intervals, and capturing the real-life frequency of treatment and covariate measurement as accurately as is practically feasible.

The positivity assumption was plausible in the estimation of baseline PS however, was near violated when PS was estimated over time. Other core assumptions to infer causal treatment effects were generally deemed plausible; for the consistency assumption, i.e., that treatment was well defined, patients prescribed urate-lowering or uricosuric drugs in the two years prior to gout consultation were excluded to ensure there was no interference from the effect of those drugs and the effect of incident allopurinol use could then be evaluated. The plausibility of exchangeability and model misspecification were continually assessed. Various PS models were fitted to assess sensitivity of estimated treatment effects to these modelling variations. Furthermore, sensitivity analyses were conducted to assess robustness of treatments effects estimates to missing data and unmeasured confounding.

Limitations and future work

This thesis had defined allopurinol use as a total of three months of allopurinol prescription in the landmark period for baseline analysis, and subsequently in each year of follow-up. Sensitivity analysis could have been performed using alternative definitions for treatment such as classifying allopurinol users as those with six or more months of prescription (Kuo et al., 2015a), and/or taking precise allopurinol dosage into account.

As shown in this thesis, allopurinol use did not lead to protective effects against poor outcomes which may be due to suboptimal management of gout, such as patients not generally being prescribed a high enough allopurinol dose needed to reach target SU level (Cottrell et al., 2013) and discontinuation of treatment (Scheepers et al., 2018). Studies had shown higher allopurinol doses were protective of renal disease (Vargas-Santos et al., 2018), cardiovascular events (Kok et al., 2014), and greater chance of reaching target SU level (Rees et al., 2013). Further research could consider effectiveness of time-varying allopurinol dosages on outcome using MSM (Lipkovich et al., 2012).

Future research could evaluate the direct effect of SU level on outcome to avoid the need to model suboptimal allopurinol dose. However, within CPRD, SU level was infrequently measured and the majority of patients who did have a recorded SU level was over 360µmol/L. If CPRD data was used for such an analysis, future work at the most can only consider SU level at baseline.

Patients adhering with allopurinol for a longer period has been shown to be associated with reduced flares (Kim et al., 2013b). Future work could also include evaluation of effectiveness of different treatment regimens within MSM, such as continuous use of allopurinol for three years versus one year, or comparisons of effectiveness made between patients fully adhering with treatment versus those partially adhering versus those not treated. Sophisticated flexible

MSM could be used to determined how much cumulative exposure to allopurinol patients are required to have in order to see the benefits of treatment on outcome (Xiao et al., 2014).

Other sensitivity analyses could have been performed to assess robustness of treatment effect estimates obtained in this study. For example, length of intervals in which covariates and treatment were repeatedly measured in the follow-up could have been varied. However as noted above, use of smaller intervals may lead to computational issues such as non-convergence of regression models. Use of larger intervals on the other hand may be inappropriate if covariates are frequently measured, such as prescriptions; using 'out-of-date' information would have weaker associations with treatment and outcome than more recent measurements, thus confounding may not be adequately controlled for. Alternatively, one could have used unequal intervals where length of interval is based on the period the patient is on treatment, with the interval ending once treatment has changed or discontinued; this approach was used by Leon (2011a) and may be a solution to having a large number of intervals that may cause modelling issues.

Correct model specification and positivity assumptions are two of the assumptions that PS methods are based on. In considering time-varying PS, estimated PS in non-allopurinol intervals were close to zero (in the subclass with the lowest PS) which subsequently led to presence of extreme weights in MSM, which indicate that the PS model may have been misspecified. In the initial application of MSM, incorrect model specification may have stemmed from the assumption that the influence of covariates was the same amongst those initiating and continuing with treatment when in fact, they had differed. Yang et al. (2015b) simulation study had shown modelling the complex mechanisms of treatment use estimated unbiased treatment effects although it was limited to two time points. Further work could extend this simulation study to include more time points.

Within time-varying PS subclassification, it was difficult to assess whether model misspecification had an impact on treatment effect. Attempts to improve common support of PS between treatment groups was largely unproductive; patients for whom common support improved after modifying the PS model ended up being excluded regardless due to small number of allopurinol users and no occurrence of outcome in the subclass corresponding to the lowest PS. A possible solution could be to estimate PS estimated separately for patients initiating and continuing with allopurinol, as this approach improve PS estimation in MSM. Simulation studies would be required to assess how well this approach would adequately control for confounding effects.

In PS models, it was assumed the decision to treat was based on covariate values measured in the interval prior to treatment and previous cumulative exposure to treatment, and that covariates measured two or more years ago had no influence. In practice, GPs are likely to prescribe treatment based on cumulative effects of past treatment, covariate history, and patient response to treatment. However, the majority of time-varying covariates only changed value once, it was unlikely that adjustment for covariates observed in earlier intervals would be strongly associated with treatment. This is more likely to apply to covariates that change more frequently over time, such as prescription data; however, this was not explored in this thesis.

PS methods make a strong assumption of no unmeasured confounding which in practice is unlikely to be satisfied. In this thesis, effects of confounding were minimised by considering a large number of clinically relevant covariates. Covariates that were not considered in this thesis may potentially be indications for allopurinol (or urate-lowering therapy more generally), include recurrent flares, tophi, chronic gouty arthritis, bone erosion, and urolithiasis (Hui et al., 2017, Richette et al., 2017). Patients with these indications are likely to have SU level above target, which in turn is associated with outcomes considered in this thesis,

for example renal disease (Desai et al., 2018). However, these measures are infrequently recorded in primary care (Kuo et al., 2014). Robustness of treatment effect estimates against unmeasured confounding was evaluated. An alternative method to account for unmeasured confounding is use of instrumental variables (IV). IV are used to remove the correlation of treatment with unmeasured covariates in the estimation of causal effects. For a covariate to be an IV, it needs to be associated with treatment, act on outcome through treatment, and be independent of the unmeasured covariate. Finding a covariate that satisfies the assumption of IV may be difficult although directed acyclic graphs may be helpful in identifying such an IV (Brookhart et al., 2010a). For example, the IV approach has been used to evaluate the associations of hyperuricaemia and gout with myocardial infarction. Multiple IVs were considered such as age, renal function, diuretic use, and the metabolic syndrome, which were assumed to affect outcome though hyperuricaemia and gout (Krishnan et al., 2006).

Residual confounding may remain from incomplete adjustment of observed covariates. In this thesis the strongest confounding variables were considered to be SU level and renal disease. Ideally, SU level should have been treated as a time-varying covariate as it is a strong predictor for allopurinol and outcome. BSR and EULAR guidelines state SU levels should be monitored in line with titrating allopurinol dose, and ensuring patients reached target SU level (Jordan et al., 2007b, Hui et al., 2017). In practice, SU level was only measured at baseline, around the time of gout consultation for a small proportion of patients. Consequently, SU level is likely to have been imbalanced between treatment groups over time. Renal disease was treated as time-varying covariate assuming that once a patient was diagnosed with renal disease, they had it for the rest of follow-up. A more accurate measure could have been based on eGFR, a measure of renal function, to gauge severity of renal disease; unfortunately, eGFR was not measured routinely in clinical practice prior to 2004.

A large proportion of patients did not have complete data on lifestyle factors, perhaps due to intermitted recording of information when patients consult (Jordan et al., 2007a). This thesis had used the MIM and LOCF, however simulation studies have shown that use of these methods can bias estimates (Cook et al., 2004, Knol et al., 2010). As described above, complete case analysis had shown treatment effect estimates from baseline analysis were more robust to missing data than time-varying analyses.

An alternate approach, multiple imputation (MI) by chained equations, could have been used to impute missing data. MI is a frequently used approach to deal with missing data, and has been adopted in PS literature; a systematic review of 167 studies that had used PS-based methods published between 2010 and 2017, found that found MI was used in 19% of these studies (Malla et al., 2018). However, it was only until recently simulation studies had shown PS estimation (at a single time point) and treatment effect analysis should be performed within each imputed data set separately prior to combining treatment effect estimates (Granger et al., 2019, Leyrat et al., 2019); and that MI provided unbiased estimates unlike complete case analysis (Leyrat et al., 2019) and MIM (Choi et al., 2019) (within the context of PS methods). Similarly, in MSM under the intention-to-treat principle, simulation studies had shown MI and inverse probability missing weights performed better than LOCF and complete case analysis as estimates were less biased (Vourli and Touloumi, 2015, Mojaverian et al., 2015). For EHR where the number of follow-up time points is large, further MI approaches have been developed to impute missing data over time (Welch et al., 2014, Kontopantelis et al., 2017); however further research is required to assess in how well these methods work compared with complete case analysis, MIM and LOCF when estimating actual treatment effect via time-varying PS subclassification and MSM. Another issue to consider is that MI assumes that data are missing at random. This assumption may not hold in EHR databases as lifestyle factors are likely to be recorded if particular information is relevant to patient care.

Thus, missing data are more likely to be missing not at random. Indeed, one study had shown MI for smoking status and alcohol consumption using EHR data from THIN led to a higher proportion of current smokers and non-drinkers compared with other external databases, suggesting the missing at random assumption was not met (Marston et al., 2010). Due to the complexity and time taken fitting PS models in this thesis, MI unfortunately was not considered.

Various sensitivity analyses that tested the assumptions made in outcome analysis could have been performed in this thesis. Non-informative censoring, i.e., each patient's censoring time is independent to their outcome time, was assumed when implementing Cox models. Competing risks occur when a competing event such as death hinders or changes the risk of the outcome of interest being observed. These censored patients may systematically differ to uncensored patients introducing bias, for example censored patients due to death may be less healthy than uncensored patients. This assumption may not be plausible with an aging study sample, where increase in number of comorbidities and increased likelihood of death are observed. Competing risks models could have potentially been used with baseline PS approach, although it was only until recently a simulation study had shown how the two methods could be combined (Austin and Fine, 2019). Within MSM, use of censoring weights could be used to balance covariate distribution between censored and uncensored patients thus the pseudo-population would contain patients who had complete follow-up data. Future work could explore use of these two methods in treatment effect estimation.

11.3 Implications and concluding remarks

This thesis evaluated the effect of allopurinol in patients with gout on a range of clinically important long-term outcomes. Differing complexity of statistical models adjusting for

confounding were considered from simple baseline PS subclassification to time-varying PS subclassification and MSMs. Choice of statistical models used in this thesis is generally largely dependent on whether there is a priori belief that measured covariates are potentially time-varying confounders affected by past treatment, adherence to treatment, and the type of effect one wants to estimate.

Correctly modelling time-varying covariates that are affected by past treatment can be challenging. MSM remove the association between covariates and treatment therefore removing confounding effects; covariates are no longer on the causal pathway between past treatment and outcome, however the association between treatment history and outcome still exists as in the original population. This allows one to specifically estimate the overall unbiased marginal (or population average) effect of treatment on outcome i.e., the direct effect of current treatment. On the other hand, time-varying PS subclassification assumes past treatment and covariates are independent in order to estimate an unbiased treatment effect; within a PS subclass it only ensures the distribution of covariates and past treatment are balanced across treatment groups. Thus the two methods estimate differing effects of treatment that are only equivalent if all confounding has been removed, which is unlikely to be the case in practice.

Studies that want to estimate the actual effect of treatment should be aware of the number of assumptions needed to infer casual effect. The largest challenge encountered in this PhD project was satisfying the two important assumptions of positivity and correct model specification. Incorrect PS model specification, that had led to near violation of the positivity assumption, may not necessarily be a result of the choice of covariates included in the PS model, but due to inadequately modelling the complex mechanism of the influence of covariates on treatment. Although the complexity of treatment use can be modelled within

MSM, it has not been shown whether the same can be achieved via time-varying PS subclassification.

It seems plausible to suggest that extensive simulation studies are required to: assess whether unbiased estimates of treatment effects can be obtained when influence of covariates on treatment use varies over time; provide guidance on the optimal approach to imputation of missing time-varying covariates; investigate under which scenarios (considering for example sample size, omission of lagged covariates with various associations with treatment and outcome) PS model misspecification may lead to biased estimates; understand how accurately the time-varying PS subclassification estimates conditional treatment effect in the presence of covariates affected by past treatment, particularly if subclasses need to be excluded from analysis; assess the impact of length of time intervals used to convert continuous time line into discrete time measurements; rigorously compare various methods available for direct estimation of time-varying treatment effect, such as G-methods (G-computation and G-estimation) (Danaei et al., 2013) and regression based models that directly adjust for covariate history (Achy-Brou et al., 2010, Keogh et al., 2018).

Complex models (as described above and used in this thesis) for assessment of time-varying treatment effect are worth considering; insight is gained into how the mechanism of treatment assignment works in real life which would otherwise be ignored in the common intention-to-treat analysis approach. The magnitude of actual treatment effect is generally larger than the effect of initiating treatment, reflecting clinical practice that may otherwise be under-estimated. In this project, this was particularly found to be the case for target SU level and gout hospitalisation where the actual treatment effect was double the effect of initiating treatment. More generally, as-treated analysis is more likely to capture small effects of actual treatment effect than effects measured from intention-to-treat analysis. These models should

be applied in different populations using EHR data to gage how successful they are in practice more generally.

To conclude, PS methodology is a valuable approach to control for confounding at baseline and over time. Time-varying PS subclassification is a direct extension of baseline PS subclassification and in theory is relatively easy to understand and implement, however in practice this was not the case. However, this method was found to be limited if common support of PS between treatment groups cannot be achieved. Alternatively, one could use MSM however analysts should be mindful that there are considerable practical challenges in ensuring correct specification of PS and outcome models, particularly when application is to complex EHR data. Use of either of these methods will rely on a large number of decisions that have to be based on both statistical and clinical arguments, including defining covariates and treatment in time-varying manner, selecting the most appropriate length of follow-up intervals and PS covariate selection procedure among others. Despite the number of considerations that needed to be made in using EHR for analysis and challenges encountered, EHR is valuable resource that is linked to a number of external health care databases, large number of covariates and outcomes recorded, and continual supply of patient and healthcare information that accurately represents what actually happens in practice which will inform what areas of healthcare needs improving.

Allopurinol use was consistently associated with greater chance of reaching target SU level and fewer flares but greater risk of gout hospitalisation, coronary heart disease and renal disease. Firstly, these adverse effects of allopurinol could be explained by residual confounding by indication from unmeasured covariates, or incomplete adjustment of covariates with missing data or covariates that were under-reported. Secondly, the target SU level was reached by only 57% of allopurinol users suggesting that allopurinol dosing and/or adherence was sub-therapeutic and it is likely that the risk of these adverse outcomes would

have been lower if allopurinol treatment was more optimal. Despite these associations, GPs should continue to treat patients with gout should continue with allopurinol for the long-term benefits of complete crystal dissolution such as cessation of gout flares and resolution of tophi (Doherty et al., 2018).

12 References

ABDUL SULTAN, A., MALLEN, C., HAYWARD, R., MULLER, S., WHITTLE, R., HOTSTON, M. & RODDY, E. 2017. Gout and subsequent erectile dysfunction: a population-based cohort study from England. *Arthritis Research & Therapy,* 19**,** 123.

ACHY-BROU, A. C., FRANGAKIS, C. E. & GRISWOLD, M. 2010. Estimating treatment effects of longitudinal designs using regression models on propensity scores. *Biometrics,* 66**,** 824-33.

AHERN, M. J., REID, C., GORDON, T. P., MCCREDIE, M., BROOKS, P. M. & JONES, M. 1987. Does colchicine work? The results of the first controlled study in acute gout. *Aust N Z J Med,* 17**,** 301-4.

ALVAREZ-NEMEGYEI, J., CEN-PISTE, J. C., MEDINA-ESCOBEDO, M. & VILLANUEVA-JORGE, S. 2005. Factors associated with musculoskeletal disability and chronic renal failure in clinically diagnosed primary gout. *Journal of Rheumatology,* 32**,** 1923-1927.

ANDERSEN, P. K. & GILL, R. D. 1982. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics,* 10**,** 1100-1120.

ANDERSON, A. H., YANG, W., TOWNSEND, R. R., PAN, Q., CHERTOW, G. M., KUSEK, J. W., CHARLESTON, J., HE, J., KALLEM, R., LASH, J. P., MILLER, E. R., 3RD, RAHMAN, M., STEIGERWALT, S., WEIR, M., WRIGHT, J. T., JR. & FELDMAN, H. I. 2015. Time-updated systolic blood pressure and the progression of chronic kidney disease: a cohort study. *Ann Intern Med,* 162**,** 258-65.

ANDERSON, D. W., KISH, L. & CORNELL, R. G. 1980. On Stratification, Grouping and Matching. *Scandinavian Journal of Statistics,* 7**,** 61-66.

ANGLEMYER, A., HORVATH, H. T. & BERO, L. 2014. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials (Review). *Cochrane Database Syst Rev.,* 4.

ANNEMANS, L., SPAEPEN, E., GASKIN, M., BONNEMAIRE, M., MALIER, V., GILBERT, T. & NUKI, G. 2008. Gout in the UK and Germany: prevalence, comorbidities and management in general practice 2000–2005. *Annals of the Rheumatic Diseases,* 67**,** 960.

AUSTIN, P. C. 2009a. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med,* 28**,** 3083-107.

AUSTIN, P. C. 2009b. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biom J,* 51**,** 171-84.

AUSTIN, P. C. 2010. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. *American Journal of Epidemiology,* 172**,** 1092-1097.

AUSTIN, P. C. 2011a. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res,* 46**,** 399-424.

AUSTIN, P. C. 2011b. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics,* 10**,** 150-61.

AUSTIN, P. C. 2013. The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in medicine,* 32**,** 2837-2849.

AUSTIN, P. C. 2014. A comparison of 12 algorithms for matching on the propensity score. *Stat Med,* 33**,** 1057-69.

AUSTIN, P. C. 2017. A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. *Int Stat Rev,* 85**,** 185-203.

AUSTIN, P. C. & FINE, J. P. 2019. Propensity-score matching with competing risks in survival analysis. *Statistics in Medicine,* 38**,** 751-777.

AUSTIN, P. C., GROOTENDORST, P. & ANDERSON, G. M. 2007a. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med,* 26**,** 734-53.

AUSTIN, P. C., GROOTENDORST, P., NORMAND, S. L. & ANDERSON, G. M. 2007b. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat Med,* 26**,** 754-68.

AZEVEDO, V. F., BUIAR, P. G., GIOVANELLA, L. H., SEVERO, C. R. & CARVALHO, M. 2014. Allopurinol, benzbromarone, or a combination in treating patients with gout: Analysis of a series of outpatients. *International Journal of Rheumatology,* 2014.

BADVE, S. V., BROWN, F., HAWLEY, C. M., JOHNSON, D. W., KANELLIS, J., RANGAN, G. K. & PERKOVIC, V. 2011. Challenges of conducting a trial of uric-acid-lowering therapy in CKD. *Nat Rev Nephrol,* 7**,** 295-300.

BADVE, S. V., PASCOE, E. M., TIKU, A., BOUDVILLE, N., BROWN, F. G., CASS, A., CLARKE, P., DALBETH, N., DAY, R. O., DE ZOYSA, J. R., DOUGLAS, B., FAULL, R., HARRIS, D. C., HAWLEY, C. M., JONES, G. R. D., KANELLIS, J., PALMER, S. C., PERKOVIC, V., RANGAN, G. K., REIDLINGER, D., ROBISON, L., WALKER, R. J., WALTERS, G. & JOHNSON, D. W. 2020. Effects of Allopurinol on the Progression of Chronic Kidney Disease. *New England Journal of Medicine,* 382**,** 2504-2513.

BECKER, M. A., SCHUMACHER, H. R., ESPINOZA, L. R., WELLS, A. F., MACDONALD, P., LLOYD, E. & LADEMACHER, C. 2010. The urate-lowering efficacy and safety of febuxostat in the treatment of the hyperuricemia of gout: the CONFIRMS trial. *Arthritis Res Ther,* 12**,** R63.

BECKER, M. A., SCHUMACHER, H. R., JR., WORTMANN, R. L., MACDONALD, P. A., EUSTACE, D., PALO, W. A., STREIT, J. & JOSEPH-RIDGE, N. 2005. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Engl J Med,* 353**,** 2450-61.

BEDSON, J., BELCHER, J., MARTINO, O. I., NDLOVU, M., RATHOD, T., WALTERS, K., DUNN, K. M. & JORDAN, K. P. 2013. The effectiveness of national guidance in changing analgesic prescribing in primary care from 2002 to 2009: an observational database study. *Eur J Pain,* 17**,** 434-43.

BENSON, K. & HARTZ, A. J. 2000. A comparison of observational studies and randomized, controlled trials. *N Engl J Med,* 342**,** 1878-86.

BHASKARAN, K., FORBES, H. J., DOUGLAS, I., LEON, D. A. & SMEETH, L. 2013. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open,* 3**,** e003389.

BHATTARAI, N., CHARLTON, J., RUDISILL, C. & GULLIFORD, M. C. 2012. Coding, recording and incidence of different forms of coronary heart disease in primary care. *PLoS One,* 7**,** e29776.

BHOLE, V., DE VERA, M., RAHMAN, M. M., KRISHNAN, E. & CHOI, H. 2010. Epidemiology of gout in women: Fifty-two–year followup of a prospective cohort. *Arthritis & Rheumatism,* 62**,** 1069-1076.

BLACK, N. 1996. Why we need observational studies to evaluate the effectiveness of health care. *BMJ (Clinical research ed.),* 312**,** 1215-1218.

BLAGOJEVIC-BUCKNALL, M., MALLEN, C., MULLER, S., HAYWARD, R., WEST, S., CHOI, H. & RODDY, E. 2019. The Risk of Gout Among Patients With Sleep Apnea: A Matched Cohort Study. *Arthritis Rheumatol,* 71**,** 154-160.

BOOTH, H. P., PREVOST, A. T. & GULLIFORD, M. C. 2013. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf,* 22**,** 1357-61.

BOUWMEESTER, W., TWISK, J. W. R., KAPPEN, T. H., VAN KLEI, W. A., MOONS, K. G. M. & VERGOUWE, Y. 2013. Prediction models for clustered data: comparison of a random intercept and standard regression model. *BMC Medical Research Methodology,* 13**,** 19.

BOWEN-DAVIES, Z., MULLER, S., MALLEN, C. D., HAYWARD, R. A. & RODDY, E. 2018. Gout Severity, Socioeconomic Status, and Work Absence: A Cross-Sectional Study in Primary Care. *Arthritis Care Res (Hoboken),* 70**,** 1822-1828.

BROOKHART, M. A., RASSEN, J. A. & SCHNEEWEISS, S. 2010a. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf,* 19**,** 537-54.

BROOKHART, M. A., SCHNEEWEISS, S., ROTHMAN, K. J., GLYNN, R. J., AVORN, J. & STURMER, T. 2006. Variable selection for propensity score models. *Am J Epidemiol,* 163**,** 1149-56.

BROOKHART, M. A., STURMER, T., GLYNN, R. J., RASSEN, J. & SCHNEEWEISS, S. 2010b. Confounding control in healthcare database research: challenges and potential approaches. *Med Care,* 48**,** S114-20.

BURSILL, D., TAYLOR, W. J., TERKELTAUB, R., ABHISHEK, A., SO, A. K., VARGAS-SANTOS, A. B., GAFFO, A. L., ROSENTHAL, A., TAUSCHE, A.-K., REGINATO, A., MANGER, B., SCIRÉ, C., PINEDA, C., VAN DURME, C., LIN, C.-T., YIN, C., ALBERT, D. A., BIERNAT-KALUZA, E., RODDY, E., PASCUAL, E., BECCE, F., PEREZ-RUIZ, F., SIVERA, F., LIOTÉ, F., SCHETT, G., NUKI, G., FILIPPOU, G., MCCARTHY, G., DA ROCHA CASTELAR PINHEIRO, G., EA, H.-K., TUPINAMBÁ, H. D. A., YAMANAKA, H., CHOI, H. K., MACKAY, J., ODELL, J. R., VÁZQUEZ MELLADO, J., SINGH, J. A., FITZGERALD, J. D., JACOBSSON, L. T. H., JOOSTEN, L., HARROLD, L. R., STAMP, L., ANDRÉS, M., GUTIERREZ, M., KUWABARA, M., DEHLIN, M., JANSSEN, M., DOHERTY, M., HERSHFIELD, M. S., PILLINGER, M., EDWARDS, N. L., SCHLESINGER, N., KUMAR, N., SLOT, O., OTTAVIANI, S., RICHETTE, P., MACMULLAN, P. A., CHAPMAN, P. T., LIPSKY, P. E., ROBINSON, P., KHANNA, P. P., GANCHEVA, R. N., GRAINGER, R., JOHNSON, R. J., TE KAMPE, R., KEENAN, R. T., TEDESCHI, S. K., KIM, S., CHOI, S. J., FIELDS, T. R., BARDIN, T., UHLIG, T., JANSEN, T., MERRIMAN, T., PASCART, T., NEOGI, T., KLÜCK, V., LOUTHRENOO, W. & DALBETH, N. 2019. Gout, Hyperuricaemia and Crystal-Associated Disease Network (G-CAN) consensus statement regarding labels and definitions of disease states of gout. *Annals of the Rheumatic Diseases,* 78**,** 1592.

BURTON, C., CAMPBELL, P., JORDAN, K., STRAUSS, V. & MALLEN, C. 2013. The association of anxiety and depression with future dementia diagnosis: a case-control study in primary care. *Family Practice,* 30**,** 25-30.

CADZOW, M., MERRIMAN, T. R. & DALBETH, N. 2017. Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. *Arthritis research & therapy,* 19**,** 181-181.

CAMPBELL, J., DEDMAN, D., EATON, S., GALLAGHER, A. & WILLIAMS, T. 2013. Is the GPRD GOLD population comparable to the UK population? *Pharmacoepidemiology and Drug Safety,* 22**,** 280.

CAMPION, E. W., GLYNN, R. J. & DELABRY, L. O. 1987. Asymptomatic hyperuricemia. Risks and consequences in the Normative Aging Study. *Am J Med,* 82**,** 421-6.

CHANDRATRE, P., MALLEN, C., RICHARDSON, J., MULLER, S., HIDER, S., ROME, K., BLAGOJEVIC-BUCKNALL, M. & RODDY, E. 2018. Health-related quality of life in gout in primary care: Baseline findings from a cohort study. *Seminars in Arthritis and Rheumatism,* 48**,** 61-69.

CHARLSON, M., SZATROWSKI, T. P., PETERSON, J. & GOLD, J. 1994. Validation of a combined comorbidity index. *J Clin Epidemiol,* 47**,** 1245-51.

CHARLSON, M. E., POMPEI, P., ALES, K. L. & MACKENZIE, C. R. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis,* 40**,** 373-83.

CHEN-XU, M., YOKOSE, C., RAI, S. K., PILLINGER, M. H. & CHOI, H. K. 2019. Contemporary Prevalence of Gout and Hyperuricemia in the United States and Decadal Trends: The National Health and Nutrition Examination Survey, 2007-2016. *Arthritis Rheumatol,* 71**,** 991-999.

CHEN, C.-J., HSIEH, M.-C., LIAO, W.-T., CHAN, Y.-T. & CHANG, S.-J. 2016. Allopurinol and the incidence of bladder cancer: a Taiwan national retrospective cohort study. *European Journal of Cancer Prevention,* 25**,** 216-223.

CHEYOE, N., KUNING, M. & LIM, A. 2012. The prevalence of chronic kidney disease among gout patients in Nongjik hospital, Pattani province. *Thai Journal of Pharmaceutical Sciences,* 36**,** 144-149.

CHOE, J. Y., PARK, S. H. & KIM, S. K. 2010. Serum Cystatin C is a Potential Endogenous Marker for the Estimation of Renal Function in Male Gout Patients with Renal Impairment. *Journal of Korean Medical Science,* 25**,** 42-48.

CHOI, H., NEOGI, T., STAMP, L., DALBETH, N. & TERKELTAUB, R. 2018. New Perspectives in Rheumatology: Implications of the Cardiovascular Safety of Febuxostat and Allopurinol in Patients With Gout and Cardiovascular Morbidities Trial and the Associated Food and Drug Administration Public Safety Alert. *Arthritis & rheumatology (Hoboken, N.J.),* 70**,** 1702-1709.

CHOI, H. K. & FORD, E. S. 2008. Haemoglobin A1c, fasting glucose, serum C-peptide and insulin resistance in relation to serum uric acid levels—the Third National Health and Nutrition Examination Survey. *Rheumatology,* 47**,** 713-717.

CHOI, H. K., FORD, E. S., LI, C. & CURHAN, G. 2007. Prevalence of the metabolic syndrome in patients with gout: The Third National Health and Nutrition Examination Survey. *Arthritis Care & Research,* 57**,** 109-115.

CHOI, H. K., GAO, X. & CURHAN, G. 2009. Vitamin C intake and the risk of gout in men: a prospective study. *Arch Intern Med,* 169**,** 502-7.

CHOI, H. K., MOUNT, D. B., REGINATO, A. M., AMERICAN COLLEGE OF PHYSICIANS & AMERICAN PHYSIOLOGICAL SOCIETY 2005. Pathogenesis of gout. *Ann Intern Med,* 143**,** 499-516.

CHOI, H. K., SORIANO, L. C., ZHANG, Y. & RODRÍGUEZ, L. A. G. 2012. Antihypertensive drugs and risk of incident gout among patients with hypertension: population based case-control study. *BMJ,* 344**,** d8190.

CHOI, J., DEKKERS, O. M. & LE CESSIE, S. 2019. A comparison of different methods to handle missing data in the context of propensity score analysis. *European journal of epidemiology,* 34**,** 23-36.

CHUNG, T.-T., YU, K.-H., KUO, C.-F., LUO, S.-F., CHIOU, M.-J., LAN, W.-C., CHEN, J.-S., TSENG, W.-Y., HSIEH, A.-H. & WANG, L.-C. 2019. Impact of urate-lowering drugs on the

progression and recovery from chronic kidney disease among gout patients. *Arthritis Research & Therapy,* 21**,** 210.

CLARSON, L. E., HIDER, S. L., BELCHER, J., HENEGHAN, C., RODDY, E. & MALLEN, C. D. 2015. Increased risk of vascular disease associated with gout: a retrospective, matched cohort study in the UK Clinical Practice Research Datalink. *Ann Rheum Dis,* 74**,** 642-7.

CLARSON, L. E., HIDER, S. L., BELCHER, J., RODDY, E. & MALLEN, C. D. 2017. Factors Influencing Allopurinol Initiation in Primary Care. *Annals of family medicine,* 15**,** 557-560.

COCHRAN, W. 1968. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics,* 24**,** 295-313.

COLE, S. R. & HERNAN, M. A. 2008. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol,* 168**,** 656-64.

COLE, S. R., HERNÁN, M. A., MARGOLICK, J. B., COHEN, M. H. & ROBINS, J. M. 2005. Marginal Structural Models for Estimating the Effect of Highly Active Antiretroviral Therapy Initiation on CD4 Cell Count. *American Journal of Epidemiology,* 162**,** 471-478.

COOK, D. G., SHAPER, A. G., THELLE, D. S. & WHITEHEAD, T. P. 1986. Serum uric acid, serum glucose and diabetes: relationships in a population study. *Postgraduate medical journal,* 62**,** 1001-1006.

COOK, N. R., COLE, S. R. & BURING, J. E. 2012. Aspirin in the primary prevention of cardiovascular disease in the Women's Health Study: effect of noncompliance. *Eur J Epidemiol,* 27**,** 431-8.

COOK, R. J., ZENG, L. & YI, G. Y. 2004. Marginal Analysis of Incomplete Longitudinal Binary Data: A Cautionary Note on LOCF Imputation. *Biometrics,* 60**,** 820-828.

COTTRELL, E., CRABTREE, V., EDWARDS, J. J. & RODDY, E. 2013. Improvement in the management of gout is vital and overdue: an audit from a UK primary care medical practice. *BMC Family Practice,* 14**,** 170.

COX, D. R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological),* 34**,** 187-220.

CRITTENDEN, D. B., LEHMANN, R. A., SCHNECK, L., KEENAN, R. T., SHAH, B., GREENBERG, J. D., CRONSTEIN, B. N., SEDLIS, S. P. & PILLINGER, M. H. 2012. Colchicine use is associated with decreased prevalence of myocardial infarction in patients with gout. *Journal of Rheumatology,* 39**,** 1458-1464.

CULLIFORD, D., MASKELL, J., JUDGE, A., COOPER, C., PRIETO-ALHAMBRA, D. & ARDEN, N. K. 2015. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. *Osteoarthritis Cartilage,* 23**,** 594-600.

CUONG, N. V. 2013. Which covariates should be controlled in propensity score matching? Evidence from a simulation study. *Statistica Neerlandica,* 67**,** 169-180.

CURRIE, C. J., BERNI, E. R., BERNI, T. R., JENKINS-JONES, S., SINSAKUL, M., JERMUTUS, L., AMBERY, P. & JAIN, M. 2019. Major adverse cardiovascular events in people with chronic kidney disease in relation to disease severity and diabetes status. *PLoS One,* 14**,** e0221044.

D'AGOSTINO, R. B. & KWAN, H. 1995. Measuring effectiveness. What to expect without a randomized control group. *Med Care,* 33**,** As95-105.

DAFNI, U. 2011. Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes,* 4**,** 363-71.

DALBETH, N., HOUSE, M. E., HORNE, A., PETRIE, K. J., MCQUEEN, F. M. & TAYLOR, W. J. 2012. Prescription and dosing of urate-lowering therapy, rather than patient behaviours, are the key modifiable factors associated with targeting serum urate in gout. *BMC Musculoskeletal Disorders,* 13.

DALBETH, N., HOUSE, M. E., HORNE, A. & TAYLOR, W. J. 2013. Reduced creatinine clearance is associated with early development of subcutaneous tophi in people with gout. *BMC Musculoskeletal Disorders,* 14.

DALBETH, N., KUMAR, S., STAMP, L. & GOW, P. 2006. Dose adjustment of allopurinol according to creatinine clearance does not provide adequate control of hyperuricemia in patients with gout. *Journal of Rheumatology,* 33**,** 1646-1651.

DALBETH, N., MERRIMAN, T. R. & STAMP, L. K. 2016. Gout. *The Lancet,* 388**,** 2039-2052.

DALBETH, N., PHIPPS-GREEN, A., FRAMPTON, C., NEOGI, T., TAYLOR, W. J. & MERRIMAN, T. R. 2018. Relationship between serum urate concentration and clinically evident incident gout: an individual participant data analysis. *Annals of the Rheumatic Diseases,* 77**,** 1048.

DANAEI, G., RODRÍGUEZ, L. A. G., CANTERO, O. F., LOGAN, R. & HERNÁN, M. A. 2013. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical methods in medical research,* 22**,** 70-96.

DANIEL, R. M., COUSENS, S. N., DE STAVOLA, B. L., KENWARD, M. G. & STERNE, J. A. C. 2013. Methods for dealing with time-dependent confounding. *Statistics in Medicine,* 32**,** 1584-1618.

DANIEL, R. M., DE STAVOLA, B. L. & COUSENS, S. N. 2011. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *The Stata Journal,* 11**,** 479-517.

DE ABAJO, F. J., GIL, M. J., RODRIGUEZ, A., GARCIA-POZA, P., ALVAREZ, A., BRYANT, V. & GARCIA-RODRIGUEZ, L. A. 2015. Allopurinol use and risk of non-fatal acute myocardial infarction. *Heart,* 101**,** 679-85.

DE VERA, M. A., MARCOTTE, G., RAI, S., GALO, J. S. & BHOLE, V. 2014. Medication adherence in gout: a systematic review. *Arthritis Care Res (Hoboken),* 66**,** 1551-9.

DEHEJIA, R. H. & WAHBA, S. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association,* 94**,** 1053-62.

DEHLIN, M., EKSTRÖM, E. H., PETZOLD, M., STRÖMBERG, U., TELG, G. & JACOBSSON, L. T. H. 2017. Factors associated with initiation and persistence of urate-lowering therapy. *Arthritis research & therapy,* 19**,** 6-6.

DEHLIN, M. & JACOBSSON, L. T. H. 2018. Trends in Gout Hospitalization in Sweden. *The Journal of Rheumatology,* 45**,** 145.

DELANEY, J. A. C., DASKALOPOULOU MD, P. S. S. & SUISSA, S. 2009. Traditional versus marginal structural models to estimate the effectiveness of β-blocker use on mortality after myocardial infarction. *Pharmacoepidemiology and Drug Safety,* 18**,** 1-6.

DESAI, R. J., FRANKLIN, J. M., SPOENDLIN-ALLEN, J., SOLOMON, D. H., DANAEI, G. & KIM, S. C. 2018. An evaluation of longitudinal changes in serum uric acid levels and associated risk of cardio-metabolic events and renal function decline in gout. *PLoS One,* 13**,** e0193622.

DEYO, R. A., CHERKIN, D. C. & CIOL, M. A. 1992. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol,* 45**,** 613-9.

DOHERTY, M., JANSEN, T. L., NUKI, G., PASCUAL, E., PEREZ-RUIZ, F., PUNZI, L., SO, A. K. & BARDIN, T. 2012. Gout: why is this curable disease so seldom cured? *Annals of the Rheumatic Diseases,* 71**,** 1765.

DOHERTY, M., JENKINS, W., RICHARDSON, H., SARMANOVA, A., ABHISHEK, A., ASHTON, D., BARCLAY, C., DOHERTY, S., DULEY, L., HATTON, R., REES, F., STEVENSON, M. & ZHANG,

W. 2018. Efficacy and cost-effectiveness of nurse-led care involving education and engagement of patients and a treat-to-target urate-lowering strategy versus usual care for gout: a randomised controlled trial. *Lancet (London, England),* 392**,** 1403-1412.

DORAN, T., KONTOPANTELIS, E., VALDERAS, J. M., CAMPBELL, S., ROLAND, M., SALISBURY, C. & REEVES, D. 2011. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ,* 342.

DORIA, A., GALECKI, A. T., SPINO, C., POP-BUSUI, R., CHERNEY, D. Z., LINGVAY, I., PARSA, A., ROSSING, P., SIGAL, R. J., AFKARIAN, M., ARONSON, R., CARAMORI, M. L., CRANDALL, J. P., DE BOER, I. H., ELLIOTT, T. G., GOLDFINE, A. B., HAW, J. S., HIRSCH, I. B., KARGER, A. B., MAAHS, D. M., MCGILL, J. B., MOLITCH, M. E., PERKINS, B. A., POLSKY, S., PRAGNELL, M., ROBINER, W. N., ROSAS, S. E., SENIOR, P., TUTTLE, K. R., UMPIERREZ, G. E., WALLIA, A., WEINSTOCK, R. S., WU, C. & MAUER, M. 2020. Serum Urate Lowering with Allopurinol and Kidney Function in Type 1 Diabetes. *New England Journal of Medicine,* 382**,** 2493-2503.

DUBREUIL, M., ZHU, Y., ZHANG, Y., SEEGER, J. D., LU, N., RHO, Y. H. & CHOI, H. K. 2015. Allopurinol initiation and all-cause mortality in the general population. *Annals of the rheumatic diseases,* 74**,** 1368-1372.

ECKEL, R. H., GRUNDY, S. M. & ZIMMET, P. Z. 2005. The metabolic syndrome. *Lancet,* 365**,** 1415-28.

ELLIOT, A. J., CROSS, K. W. & FLEMING, D. M. 2009. Seasonality and trends in the incidence and prevalence of gout in England and Wales 1994-2007. *Ann Rheum Dis,* 68**,** 1728-33.

EMMERSON, B. T., GORDON, R. B., CROSS, M. & THOMSON, D. B. 1987. Plasma oxipurinol concentrations during allopurinol therapy. *British Journal of Rheumatology,* 26**,** 445-449.

EVANS, P. L., PRIOR, J. A., BELCHER, J., MALLEN, C. D., HAY, C. A. & RODDY, E. 2018. Obesity, hypertension and diuretic use as risk factors for incident gout: a systematic review and meta-analysis of cohort studies. *Arthritis research & therapy,* 20**,** 136-136.

FAIRHURST, C., WATT, I., MARTIN, F., BLAND, M. & BRACKENBURY, W. J. 2014. Exposure to sodium channel-inhibiting drugs and cancer survival: protocol for a cohort study using the QResearch primary care database. *BMJ Open,* 4**,** e006604.

FARMER, R. E., FORD, D., MATHUR, R., CHATURVEDI, N., KAPLAN, R., SMEETH, L. & BHASKARAN, K. 2019. Metformin use and risk of cancer in patients with type 2 diabetes: a cohort study of primary care records using inverse probability weighting of marginal structural models. *International Journal of Epidemiology,* 48**,** 527-537.

FESSEL, W. J. 1979. Renal outcomes of gout and hyperuricemia. *American Journal of Medicine,* 67**,** 74-82.

FEWELL, Z., HERNÁN, M. A., WOLFE, F., TILLING, K., CHOI, H. & STERNE, J. A. C. 2004. Controlling for Time-dependent Confounding using Marginal Structural Models. *The Stata Journal,* 4**,** 402-420.

FISHER, L. D. & LIN, D. Y. 1999. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health,* 20**,** 145-157.

FITZGERALD, J. D., DALBETH, N., MIKULS, T., BRIGNARDELLO-PETERSEN, R., GUYATT, G., ABELES, A. M., GELBER, A. C., HARROLD, L. R., KHANNA, D., KING, C., LEVY, G., LIBBEY, C., MOUNT, D., PILLINGER, M. H., ROSENTHAL, A., SINGH, J. A., SIMS, J. E., SMITH, B. J., WENGER, N. S., BAE, S. S., DANVE, A., KHANNA, P. P., KIM, S. C., LENERT, A., POON, S., QASIM, A., SEHRA, S. T., SHARMA, T. S. K., TOPROVER, M., TURGUNBAEV, M., ZENG, L., ZHANG, M. A., TURNER, A. S. & NEOGI, T. 2020. 2020 American College of

Rheumatology Guideline for the Management of Gout. *Arthritis Care Res (Hoboken),* 72**,** 744-760.

FLEEMAN, N., PILKINGTON, G., DUNDAR, Y., DWAN, K., BOLAND, A., DICKSON, R., ANIJEET, H., KENNEDY, T. & PYATT, J. 2014. Allopurinol for the treatment of chronic kidney disease: a systematic review. *Health Technol Assess,* 18**,** 1-77, v-vi.

FOODY, J., TURPIN, R. S., TIDWELL, B. A., LAWRENCE, D. & SCHULMAN, K. L. 2017. Major Cardiovascular Events in Patients with Gout and Associated Cardiovascular Disease or Heart Failure and Chronic Kidney Disease Initiating a Xanthine Oxidase Inhibitor. *Am Health Drug Benefits,* 10**,** 393-401.

FRIEDEN, T. R. 2017. Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. *N Engl J Med,* 377**,** 465-475.

GAFFO, A. L., EDWARDS, N. L. & SAAG, K. G. 2009. Gout. Hyperuricemia and cardiovascular disease: how strong is the evidence for a causal link? *Arthritis Research & Therapy,* 11**,** 240.

GALLAGHER, A. M., DEDMAN, D., PADMANABHAN, S., LEUFKENS, H. G. M. & DE VRIES, F. 2019. 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. *Pharmacoepidemiology and drug safety,* 28**,** 563-569.

GAMBLE, J. M., CHIBRIKOV, E., TWELLS, L. K., MIDODZI, W. K., YOUNG, S. W., MACDONALD, D. & MAJUMDAR, S. R. 2017. Association of insulin dosage with mortality or major adverse cardiovascular events: a retrospective cohort study. *Lancet Diabetes Endocrinol,* 5**,** 43-52.

GARDNER, M. J., POWER, C., BARKER, D. J. & PADDAY, R. 1982. The prevalence of gout in three English towns. *Int J Epidemiol,* 11**,** 71-5.

GRAFFEO, N., LATOUCHE, A., GESKUS, R. B. & CHEVRET, S. 2018. Modeling time-varying exposure using inverse probability of treatment weights. *Biom J,* 60**,** 323-332.

GRAHAM, G. G., KANNANGARA, D. R. W., STOCKER, S. L., PORTEK, I., PILE, K. D., INDRARATNA, P. L., DATTA, I., WILLIAMS, K. M. & DAY, R. O. 2013. Understanding the dose-response relationship of allopurinol: predicting the optimal dosage. *British Journal of Clinical Pharmacology,* 76**,** 932-938.

GRAMBSCH, P. M. & THERNEAU, T. M. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika,* 81**,** 515-526.

GRANGER, E., SERGEANT, J. C. & LUNT, M. 2019. Avoiding pitfalls when combining multiple imputation and propensity scores. *Statistics in Medicine,* 38**,** 5120-5132.

GREENLAND, S. & MORGENSTERN, H. 1990. Matching and efficiency in cohort studies. *Am J Epidemiol,* 131**,** 151-9.

HAK, A. E., CURHAN, G. C., GRODSTEIN, F. & CHOI, H. K. 2010. Menopause, postmenopausal hormone use and risk of incident gout. *Ann Rheum Dis,* 69**,** 1305-9.

HALPERN, R., FULDEORE, M. J., MODY, R. R., PATEL, P. A. & MIKULS, T. R. 2009. The effect of serum urate on gout flares and their associated costs: An administrative claims analysis. *Journal of Clinical Rheumatology,* 15**,** 3-7.

HARDER, V. S., STUART, E. A. & ANTHONY, J. C. 2010. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Methods,* 15**,** 234-49.

HATOUM, H., KHANNA, D., LIN, S. J., AKHRAS, K. S., SHIOZAWA, A. & KHANNA, P. 2014. Achieving serum urate goal: a comparative effectiveness study between allopurinol and febuxostat. *Postgrad Med,* 126**,** 65-75.

HAWKINS, N. M., SCHOLES, S., BAJEKAL, M., LOVE, H., O'FLAHERTY, M., RAINE, R. & CAPEWELL, S. 2013. The UK National Health Service: delivering equitable treatment across the spectrum of coronary disease. *Circ Cardiovasc Qual Outcomes,* 6**,** 208-16.

HAY, C. A., PRIOR, J. A., BELCHER, J., MALLEN, C. D. & RODDY, E. 2020. Mortality in patients with gout treated with allopurinol: a systematic review and meta-analysis. *Arthritis Care & Research,* n/a.

HEDEKER, D., SIDDIQUI, O. & HU, F. B. 2000. Random-effects regression analysis of correlated grouped-time survival data. *Stat Methods Med Res,* 9**,** 161-79.

HERMAN, J. B. & GOLDBOURT, U. 1982. Uric acid and diabetes: observations in a population study. *Lancet,* 2**,** 240-3.

HERNAN, M. A., BRUMBACK, B. & ROBINS, J. M. 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology,* 11**,** 561-70.

HERNÁN, M. A. & HERNÁNDEZ-DÍAZ, S. 2012. Beyond the intention-to-treat in comparative effectiveness research. *Clinical trials (London, England),* 9**,** 48-55.

HERNÁN, M. A. & ROBINS, J. M. 2020. *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC.

HERRETT, E., GALLAGHER, A. M., BHASKARAN, K., FORBES, H., MATHUR, R., VAN STAA, T. & SMEETH, L. 2015. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol,* 44**,** 827-36.

HERRETT, E., THOMAS, S. L., SCHOONEN, W. M., SMEETH, L. & HALL, A. J. 2010. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol,* 69**,** 4-14.

HO, D. E., IMAI, K., KING, G. & STUART, E. A. 2017. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis,* 15**,** 199-236.

HORSFIELD, P. 2004. *QOF Business Rules v5* [Online]. Available: https://clinicalcodes.rss.mhs.man.ac.uk/medcodes/article/2/ [Accessed 23/05/16 2016].

HSIANG, J. C., WONG, G. L.-H., TSE, Y.-K., WONG, V. W.-S., YIP, T. C.-F. & CHAN, H. L.-Y. 2015. Statin and the risk of hepatocellular carcinoma and death in a hospital-based hepatitis B-infected population: A propensity score landmark analysis. *Journal of Hepatology,* 63**,** 1190-1197.

HUI, M., CARR, A., CAMERON, S., DAVENPORT, G., DOHERTY, M., FORRESTER, H., JENKINS, W., JORDAN, K. M., MALLEN, C. D., MCDONALD, T. M., NUKI, G., PYWELL, A., ZHANG, W., RODDY, E., BRITISH SOCIETY FOR RHEUMATOLOGY STANDARDS, A. & GUIDELINES WORKING, G. 2017. The British Society for Rheumatology Guideline for the Management of Gout. *Rheumatology (Oxford),* 56**,** e1-e20.

HULLSIEK, K. H. & LOUIS, T. A. 2002. Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics,* 3**,** 179-193.

HUTTON, I., GAMBLE, G., GOW, P. & DALBETH, N. 2009. Factors associated with recurrent hospital admissions for gout: a case-control study. *J Clin Rheumatol,* 15**,** 271-4.

IMAI, K. 2008. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Statist. Soc. A,* 171**,** 481-502.

IMAI, K. & VAN DYK, D. A. 2004. Causal inference with general treatment regimes: generalizing the propensity score. *Journal of the American Statistical Association,* 99**,** 854-866.

JACKSON, J. W. 2016. Diagnostics for Confounding of Time-varying and Other Joint Exposures. *Epidemiology (Cambridge, Mass.),* 27**,** 859-869.

JANSSENS, H. J., JANSSEN, M., VAN DE LISDONK, E. H., VAN RIEL, P. L. & VAN WEEL, C. 2008. Use of oral prednisolone or naproxen for the treatment of gout arthritis: a double-blind, randomised equivalence trial. *Lancet,* 371**,** 1854-60.

JICK, H., JICK, S. S. & DERBY, L. E. 1991. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ,* 302**,** 766-8.

JORDAN, D. M., CHOI, H. K., VERBANCK, M., TOPLESS, R., WON, H.-H., NADKARNI, G., MERRIMAN, T. R. & DO, R. 2019. No causal effects of serum urate levels on the risk of chronic kidney disease: A Mendelian randomization study. *PLoS medicine,* 16**,** e1002725-e1002725.

JORDAN, K., CLARKE, A. M., SYMMONS, D. P. M., FLEMING, D., PORCHERET, M., KADAM, U. T. & CROFT, P. 2007a. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *The British journal of general practice : the journal of the Royal College of General Practitioners,* 57**,** 7-14.

JORDAN, K. M., CAMERON, J. S., SNAITH, M., ZHANG, W., DOHERTY, M., SECKL, J., HINGORANI, A., JAQUES, R., NUKI, G., BRITISH SOCIETY FOR, R., BRITISH HEALTH PROFESSIONALS IN RHEUMATOLOGY STANDARDS, G. & AUDIT WORKING, G. 2007b. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of gout. *Rheumatology (Oxford),* 46**,** 1372-4.

KANG, E. H., CHOI, H. K., SHIN, A., LEE, Y. J., LEE, E. B., SONG, Y. W. & KIM, S. C. 2019. Comparative cardiovascular risk of allopurinol versus febuxostat in patients with gout: a nation-wide cohort study. *Rheumatology,* 58**,** 2122-2129.

KEOGH, R. H., DANIEL, R. M., VANDERWEELE, T. J. & VANSTEELANDT, S. 2018. Analysis of Longitudinal Studies With Repeated Outcome Measures: Adjusting for Time-Dependent Confounding Using Conventional Methods. *American Journal of Epidemiology,* 187**,** 1085-1092.

KHAN, N. F., PERERA, R., HARPER, S. & ROSE, P. W. 2010. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Family Practice,* 11**,** 1-1.

KHANNA, D., FITZGERALD, J. D., KHANNA, P. P., BAE, S., SINGH, M. K., NEOGI, T., PILLINGER, M. H., MERILL, J., LEE, S., PRAKASH, S., KALDAS, M., GOGIA, M., PEREZ-RUIZ, F., TAYLOR, W., LIOTE, F., CHOI, H., SINGH, J. A., DALBETH, N., KAPLAN, S., NIYYAR, V., JONES, D., YAROWS, S. A., ROESSLER, B., KERR, G., KING, C., LEVY, G., FURST, D. E., EDWARDS, N. L., MANDELL, B., SCHUMACHER, H. R., ROBBINS, M., WENGER, N., TERKELTAUB, R. & AMERICAN COLLEGE OF RHEUMATOLOGY 2012a. 2012 American College of Rheumatology guidelines for management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. *Arthritis Care Res (Hoboken),* 64**,** 1431-46.

KHANNA, D., KHANNA, P. P., FITZGERALD, J. D., SINGH, M. K., BAE, S., NEOGI, T., PILLINGER, M. H., MERILL, J., LEE, S., PRAKASH, S., KALDAS, M., GOGIA, M., PEREZ-RUIZ, F., TAYLOR, W., LIOTE, F., CHOI, H., SINGH, J. A., DALBETH, N., KAPLAN, S., NIYYAR, V., JONES, D., YAROWS, S. A., ROESSLER, B., KERR, G., KING, C., LEVY, G., FURST, D. E., EDWARDS, N. L., MANDELL, B., SCHUMACHER, H. R., ROBBINS, M., WENGER, N., TERKELTAUB, R. & AMERICAN COLLEGE OF, R. 2012b. 2012 American College of Rheumatology guidelines for management of gout. Part 2: therapy and antiinflammatory prophylaxis of acute gouty arthritis. *Arthritis Care Res (Hoboken),* 64**,** 1447-61.

KHANNA, P. P., GLADUE, H. S., SINGH, M. K., FITZGERALD, J. D., BAE, S., PRAKASH, S., KALDAS, M., GOGIA, M., BERROCAL, V., TOWNSEND, W., TERKELTAUB, R. & KHANNA, D. 2014.

Treatment of acute gout: A systematic review. *Seminars in Arthritis and Rheumatism,* 44**,** 31-38.

KHANNA, P. P., NUKI, G., BARDIN, T., TAUSCHE, A.-K., FORSYTHE, A., GOREN, A., VIETRI, J. & KHANNA, D. 2012c. Tophi and frequent gout flares are associated with impairments to quality of life, productivity, and increased healthcare resource use: Results from a cross-sectional survey. *Health and quality of life outcomes,* 10**,** 117-117.

KIENHORST, L. B. E., JANSSENS, H. J. E. M., FRANSEN, J. & JANSSEN, M. 2014. The validation of a diagnostic rule for gout without joint fluid analysis: a prospective study. *Rheumatology,* 54**,** 609-614.

KIM, S. C., NEOGI, T., KANG, E. H., LIU, J., DESAI, R. J., ZHANG, M. & SOLOMON, D. H. 2018. Cardiovascular Risks of Probenecid Versus Allopurinol in Older Patients With Gout. *J Am Coll Cardiol,* 71**,** 994-1004.

KIM, S. C., NEWCOMB, C., MARGOLIS, D., ROY, J. & HENNESSY, S. 2013a. Severe cutaneous reactions requiring hospitalization in allopurinol initiators: a population-based cohort study. *Arthritis Care Res (Hoboken),* 65**,** 578-84.

KIM, S. C., SCHMIDT, B. M., FRANKLIN, J. M., LIU, J., SOLOMON, D. H. & SCHNEEWEISS, S. 2013b. Clinical and health care use characteristics of patients newly starting allopurinol, febuxostat, and colchicine for the treatment of gout. *Arthritis Care Res (Hoboken),* 65**,** 2008-14.

KIMURA, K., HOSOYA, T., UCHIDA, S., INABA, M., MAKINO, H., MARUYAMA, S., ITO, S., YAMAMOTO, T., TOMINO, Y., OHNO, I., SHIBAGAKI, Y., IIMURO, S., IMAI, N., KUWABARA, M., HAYAKAWA, H., OHTSU, H. & OHASHI, Y. 2018. Febuxostat Therapy for Patients With Stage 3 CKD and Asymptomatic Hyperuricemia: A Randomized Trial. *Am J Kidney Dis,* 72**,** 798-810.

KNOL, M. J., JANSSEN, K. J. M., DONDERS, A. R. T., EGBERTS, A. C. G., HEERDINK, E. R., GROBBEE, D. E., MOONS, K. G. M. & GEERLINGS, M. I. 2010. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *Journal of Clinical Epidemiology,* 63**,** 728-736.

KOK, V. C., HORNG, J.-T., CHANG, W.-S., HONG, Y.-F. & CHANG, T.-H. 2014. Allopurinol therapy in gout patients does not associate with beneficial cardiovascular outcomes: a population-based matched-cohort study. *PloS one,* 9**,** e99102.

KONTOPANTELIS, E., OLIER, I., PLANNER, C., REEVES, D., ASHCROFT, D. M., GASK, L., DORAN, T. & REILLY, S. 2015a. Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the Clinical Practice Research Datalink. *BMJ Open,* 5.

KONTOPANTELIS, E., PARISI, R., SPRINGATE, D. A. & REEVES, D. 2017. Longitudinal multiple imputation approaches for body mass index or other variables with very low individual-level variability: the mibmi command in Stata. *BMC Res Notes,* 10**,** 41.

KONTOPANTELIS, E., REEVES, D., VALDERAS, J. M., CAMPBELL, S. & DORAN, T. 2012. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Quality & Safety*.

KONTOPANTELIS, E., SPRINGATE, D., REEVES, D., ASHCROFT, D., RUTTER, M., BUCHAN, I. & DORAN, T. 2015b. Glucose, blood pressure and cholesterol levels and their relationships to clinical outcomes in type 2 diabetes: a retrospective cohort study. *Diabetologia,* 58**,** 505-518.

KONTOPANTELIS, E., SPRINGATE, D., REEVES, D., ASHCROFT, D. M., VALDERAS, J. M. & DORAN, T. 2014. Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. *BMJ,* 348.

KONTOPANTELIS, E., STEVENS, R. J., HELMS, P. J., EDWARDS, D., DORAN, T. & ASHCROFT, D. M. 2018. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open,* 8**,** e020738.

KÖTTGEN, A., ALBRECHT, E., TEUMER, A., VITART, V., KRUMSIEK, J., HUNDERTMARK, C., PISTIS, G., RUGGIERO, D., O'SEAGHDHA, C. M., HALLER, T., YANG, Q., TANAKA, T., JOHNSON, A. D., KUTALIK, Z., SMITH, A. V., SHI, J., STRUCHALIN, M., MIDDELBERG, R. P. S., BROWN, M. J., GAFFO, A. L., PIRASTU, N., LI, G., HAYWARD, C., ZEMUNIK, T., HUFFMAN, J., YENGO, L., ZHAO, J. H., DEMIRKAN, A., FEITOSA, M. F., LIU, X., MALERBA, G., LOPEZ, L. M., VAN DER HARST, P., LI, X., KLEBER, M. E., HICKS, A. A., NOLTE, I. M., JOHANSSON, A., MURGIA, F., WILD, S. H., BAKKER, S. J. L., PEDEN, J. F., DEHGHAN, A., STERI, M., TENESA, A., LAGOU, V., SALO, P., MANGINO, M., ROSE, L. M., LEHTIMÄKI, T., WOODWARD, O. M., OKADA, Y., TIN, A., MÜLLER, C., OLDMEADOW, C., PUTKU, M., CZAMARA, D., KRAFT, P., FROGHERI, L., THUN, G. A., GROTEVENDT, A., GISLASON, G. K., HARRIS, T. B., LAUNER, L. J., MCARDLE, P., SHULDINER, A. R., BOERWINKLE, E., CORESH, J., SCHMIDT, H., SCHALLERT, M., MARTIN, N. G., MONTGOMERY, G. W., KUBO, M., NAKAMURA, Y., TANAKA, T., MUNROE, P. B., SAMANI, N. J., JACOBS, D. R., JR., LIU, K., D'ADAMO, P., ULIVI, S., ROTTER, J. I., PSATY, B. M., VOLLENWEIDER, P., WAEBER, G., CAMPBELL, S., DEVUYST, O., NAVARRO, P., KOLCIC, I., HASTIE, N., BALKAU, B., FROGUEL, P., ESKO, T., SALUMETS, A., KHAW, K. T., LANGENBERG, C., WAREHAM, N. J., ISAACS, A., KRAJA, A., ZHANG, Q., et al. 2013. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nature genetics,* 45**,** 145-154.

KRISHNAN, E., BAKER, J. F., FURST, D. E. & SCHUMACHER, H. R. 2006. Gout and the risk of acute myocardial infarction. *Arthritis & Rheumatism,* 54**,** 2688-2696.

KUO, C.-F., CHOU, I.-J., SEE, L.-C., CHEN, J.-S., YU, K.-H., LUO, S.-F., HSIEH, A.-H., ZHANG, W. & DOHERTY, M. 2018. Urate-lowering treatment and risk of total joint replacement in patients with gout. *Rheumatology,* 57**,** 2129-2139.

KUO, C.-F., GRAINGE, M. J., MALLEN, C., ZHANG, W. & DOHERTY, M. 2016a. Impact of gout on the risk of atrial fibrillation. *Rheumatology (Oxford, England),* 55**,** 721-728.

KUO, C.-F., YU, K.-H., SEE, L.-C., CHOU, I. J., TSENG, W.-Y., CHANG, H.-C., SHEN, Y.-M. & LUO, S.-F. 2011. Elevated risk of mortality among gout patients: A comparison with the National Population in Taiwan. *Joint Bone Spine,* 78**,** 577-580.

KUO, C., GRAINGE, M. J., MALLEN, C., ZHANG, W. & DOHERTY, M. 2014. Eligibility for and prescription of urate-lowering treatment in patients with incident gout in england. *JAMA,* 312**,** 2684-2686.

KUO, C. F., GRAINGE, M. J., MALLEN, C., ZHANG, W. & DOHERTY, M. 2015a. Effect of allopurinol on all-cause mortality in adults with incident gout: propensity score-matched landmark analysis. *Rheumatology (Oxford),* 54**,** 2145-50.

KUO, C. F., GRAINGE, M. J., MALLEN, C., ZHANG, W. & DOHERTY, M. 2015b. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann Rheum Dis,* 74**,** 661-7.

KUO, C. F., GRAINGE, M. J., MALLEN, C., ZHANG, W. & DOHERTY, M. 2016b. Comorbidities in patients with gout prior to and following diagnosis: case-control study. *Ann Rheum Dis,* 75**,** 210-7.

KUSS, O., LEGLER, T. & BÖRGERMANN, J. 2011. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *Journal of Clinical Epidemiology,* 64**,** 1076-1084.

KYLE, R. P., MOODIE, E. E. M., KLEIN, M. B. & ABRAHAMOWICZ, M. 2019. Evaluating Flexible Modeling of Continuous Covariates in Inverse-Weighted Estimators. *American journal of epidemiology,* 188**,** 1181-1191.

LAVIKAINEN, P., HELIN-SALMIVAARA, A., EEROLA, M., FANG, G., HARTIKAINEN, J., HUUPPONEN, R. & KORHONEN, M. J. 2016. Statin adherence and risk of acute cardiovascular events among women: a cohort study accounting for time-dependent confounding affected by previous adherence. *BMJ Open,* 6**,** e011306.

LAWRENSON, R., WILLIAMS, T. & FARMER, R. 1999. Clinical information for research; the use of general practice databases. *J Public Health Med,* 21**,** 299-304.

LEE, J. S., WON, J., KWON, O. C., LEE, S. S., OH, J. S., KIM, Y.-G., LEE, C.-K., YOO, B. & HONG, S. 2019. Hepatic Safety of Febuxostat Compared with Allopurinol in Gout Patients with Fatty Liver Disease. *The Journal of Rheumatology,* 46**,** 527.

LEFEBVRE, G., DELANEY, J. A. & PLATT, R. W. 2008. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Stat Med,* 27**,** 3629-42.

LEITE, W. 2016. *Practical Propensity Score Methods Using R*.

LEON, A. C. 2011a. Evaluation of psychiatric interventions in an observational study: issues in design and analysis. *Dialogues Clin Neurosci,* 13**,** 191-8.

LEON, A. C. 2011b. Propensity score stratification for observational comparison of repeated binary outcomes. *Statistics and its interface*.

LEON, A. C., DEMIRTAS, H., LI, C. & HEDEKER, D. 2012a. Two propensity score-based strategies for a three-decade observational study: investigating psychotropic medications and suicide risk. *Stat Med,* 31**,** 3255-60.

LEON, A. C. & HEDEKER, D. 2005. A mixed-effects quintile-stratified propensity adjustment for effectiveness analyses of ordered categorical doses. *Stat Med,* 24**,** 647-58.

LEON, A. C. & HEDEKER, D. 2007a. A comparison of mixed-effects quantile stratification propensity adjustment strategies for longitudinal treatment effectiveness analyses of continuous outcomes. *Stat Med,* 26**,** 2650-65.

LEON, A. C. & HEDEKER, D. 2007b. Quantile Stratification Based on a Misspecified Propensity Score in Longitudinal Treatment Effectiveness Analyses of Ordinal Doses. *Comput Stat Data Anal,* 51**,** 6114-6122.

LEON, A. C., MUELLER, T. I., SOLOMON, D. A. & KELLER, M. B. 2001. A dynamic adaptation of the propensity score adjustment for effectiveness analyses of ordinal doses of treatment. *Stat Med,* 20**,** 1487-98.

LEON, A. C., SOLOMON, D. A., LI, C., FIEDOROWICZ, J. G., CORYELL, W. H., ENDICOTT, J. & KELLER, M. B. 2011. Antidepressants and risks of suicide and suicide attempts: a 27-year observational study. *J Clin Psychiatry,* 72**,** 580-6.

LEON, A. C., SOLOMON, D. A., LI, C., FIEDOROWICZ, J. G., CORYELL, W. H., ENDICOTT, J. & KELLER, M. B. 2012b. Antiepileptic drugs for bipolar disorder and the risk of suicidal behavior: a 30-year observational study. *The American journal of psychiatry,* 169**,** 285-291.

LEON, A. C., SOLOMON, D. A., MUELLER, T. I., ENDICOTT, J., RICE, J. P., MASER, J. D., CORYELL, W. & KELLER, M. B. 2003. A 20-year longitudinal observational study of somatic antidepressant treatment effectiveness. *Am J Psychiatry,* 160**,** 727-33.

LÉVESQUE, L. E., HANLEY, J. A., KEZOUH, A. & SUISSA, S. 2010. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ,* 340.

LEYRAT, C., SEAMAN, S. R., WHITE, I. R., DOUGLAS, I., SMEETH, L., KIM, J., RESCHE-RIGON, M., CARPENTER, J. R. & WILLIAMSON, E. J. 2019. Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Statistical Methods in Medical Research,* 28**,** 3-19.

LI, F., ZASLAVSKY, A. M. & LANDRUM, M. B. 2013. Propensity score weighting with multilevel data. *Stat Med,* 32**,** 3373-87.

LI, L., YANG, C., ZHAO, Y., ZENG, X., LIU, F. & FU, P. 2014. Is hyperuricemia an independent risk factor for new-onset chronic kidney disease?: A systematic review and meta-analysis based on observational cohort studies. *BMC nephrology,* 15**,** 122-122.

LI, R., YU, K. & LI, C. 2018. Dietary factors and risk of gout and hyperuricemia: a meta-analysis and systematic review. *Asia Pac J Clin Nutr,* 27**,** 1344-1356.

LI, Y. P., PROPERT, K. J. & ROSENBAUM, P. R. 2001. Balanced Risk Set Matching. *Journal of the American Statistical Association,* 96**,** 870-882.

LIM, A. Y. N., SHEN, L., TAN, C. H., LATEEF, A., LAU, T. C. & TENG, G. G. 2012. Achieving treat to target in gout: A clinical practice improvement project. *Scandinavian Journal of Rheumatology,* 41**,** 450-457.

LIN, D. Y., PSATY, B. M. & KRONMAL, R. A. 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics,* 54**,** 948-63.

LIPKOVICH, I., MALLINCKRODT, C. H. & FARIES, D. E. 2012. The challenges of evaluating dose response in flexible-dose trials using marginal structural models. *Pharm Stat,* 11**,** 485-93.

LOEB, J. N. 1972. The influence of temperature on the solubility of monosodium urate. *Arthritis & Rheumatism,* 15**,** 189-192.

LONJON, G., BOUTRON, I., TRINQUART, L., AHMAD, N., AIM, F., NIZARD, R. & RAVAUD, P. 2014. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg,* 259**,** 18-25.

LOTTMANN, K., CHEN, X. & SCHÄDLICH, P. K. 2012. Association between gout and all-cause as well as cardiovascular mortality: a systematic review. *Current rheumatology reports,* 14**,** 195-203.

LU, B. 2005. Propensity Score Matching with Time-Dependent Covariates. *Biometrics,* 61**,** 721-728.

LUNCEFORD, J. K. & DAVIDIAN, M. 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med,* 23**,** 2937-60.

MA, C. A. & LEUNG, Y. Y. 2017. Exploring the Link between Uric Acid and Osteoarthritis. *Frontiers in medicine,* 4**,** 225-225.

MACDONALD, T. M., FORD, I., NUKI, G., MACKENZIE, I. S., DE CATERINA, R., FINDLAY, E., HALLAS, J., HAWKEY, C. J., RALSTON, S., WALTERS, M., WEBSTER, J., MCMURRAY, J., PEREZ RUIZ, F. & JENNINGS, C. G. 2014. Protocol of the Febuxostat versus Allopurinol Streamlined Trial (FAST): a large prospective, randomised, open, blinded endpoint study comparing the cardiovascular safety of allopurinol and febuxostat in the management of symptomatic hyperuricaemia. *BMJ Open,* 4**,** e005354.

MACKENZIE, I. S., FORD, I., NUKI, G., HALLAS, J., HAWKEY, C. J., WEBSTER, J., RALSTON, S. H., WALTERS, M., ROBERTSON, M., DE CATERINA, R., FINDLAY, E., PEREZ-RUIZ, F., MCMURRAY, J. J. V. & MACDONALD, T. M. 2020. Long-term cardiovascular safety of

febuxostat compared with allopurinol in patients with gout (FAST): a multicentre, prospective, randomised, open-label, non-inferiority trial. *Lancet,* 396**,** 1745-1757.

MACKENZIE, I. S., FORD, I., WALKER, A., HAWKEY, C., BEGG, A., AVERY, A., TAGGAR, J., WEI, L., STRUTHERS, A. D. & MACDONALD, T. M. 2016. Multicentre, prospective, randomised, open-label, blinded end point trial of the efficacy of allopurinol therapy in improving cardiovascular outcomes in patients with ischaemic heart disease: protocol of the ALL-HEART study. *BMJ Open,* 6**,** e013774.

MAJOR, T. J., DALBETH, N., STAHL, E. A. & MERRIMAN, T. R. 2018. An update on the genetics of hyperuricaemia and gout. *Nat Rev Rheumatol,* 14**,** 341-353.

MAK, A., HO, R. C., TAN, J. Y., TENG, G. G., LAHIRI, M., LATEEF, A., VASOO, S., BOEY, M. L., KOH, D. R. & FENG, P. H. 2009. Atherogenic serum lipid profile is an independent predictor for gouty flares in patients with gouty arthropathy. *Rheumatology (Oxford),* 48**,** 262-5.

MALLA, L., PERERA-SALAZAR, R., MCFADDEN, E., OGERO, M., STEPNIEWSKA, K. & ENGLISH, M. 2018. Handling missing data in propensity score estimation in comparative effectiveness evaluations: a systematic review. *Journal of Comparative Effectiveness Research,* 7**,** 271-279.

MAN, C. Y., CHEUNG, I. T., CAMERON, P. A. & RAINER, T. H. 2007. Comparison of oral prednisolone/paracetamol and oral indomethacin/paracetamol combination therapy in the treatment of acute goutlike arthritis: a double-blind, randomized, controlled trial. *Ann Emerg Med,* 49**,** 670-7.

MANTEL, N. 1963. Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel-Haenszel Procedure. *Journal of the American Statistical Association,* 58**,** 690-700.

MANTEL, N. & HAENSZEL, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst,* 22**,** 719-48.

MARCHINI, G. S., SARKISSIAN, C., TIAN, D. V., GEBRESELASSIE, S. & MONGA, M. 2013. Gout, Stone Composition and Urinary Stone Risk: A Matched Case Comparative Study. *Journal of Urology,* 189**,** 1334-1339.

MARSTON, L., CARPENTER, J. R., WALTERS, K. R., MORRIS, R. W., NAZARETH, I. & PETERSEN, I. 2010. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf,* 19**,** 618-26.

MATHUR, M. B., DING, P., RIDDELL, C. A. & VANDERWEELE, T. J. 2018. Web Site and R Package for Computing E-values. *Epidemiology,* 29.

MCCANDLESS, L. C., GUSTAFSON, P. & AUSTIN, P. C. 2009. Bayesian propensity score analysis for observational data. *Statistics in Medicine,* 28**,** 94-112.

MCQUEEN, F. M., DOYLE, A., REEVES, Q., GAO, A., TSAI, A., GAMBLE, G. D., CURTEIS, B., WILLIAMS, M. & DALBETH, N. 2014. Bone erosions in patients with chronic gouty arthropathy are associated with tophi but not bone oedema or synovitis: new insights from a 3 T MRI study. *Rheumatology (Oxford),* 53**,** 95-103.

MEEK, I. L., VONKEMAN, H. E. & VAN DE LAAR, M. 2014. Hyperuricaemia: a marker of increased cardiovascular risk in rheumatic patients: analysis of the ACT-CVD cohort. *Bmc Musculoskeletal Disorders,* 15.

MEIER, C. R. & JICK, H. 1997. Omeprazole, other antiulcer drugs and newly diagnosed gout. *Br J Clin Pharmacol,* 44**,** 175-8.

MOJAVERIAN, N., MOODIE, E. E., BLIU, A. & KLEIN, M. B. 2015. The Impact of Sparse Follow-up on Marginal Structural Models for Time-to-Event Data. *Am J Epidemiol,* 182**,** 1047-55.

NEOGI, T., CHEN, C., NIU, J., CHAISSON, C., HUNTER, D. J. & ZHANG, Y. 2014. Alcohol quantity and type on risk of recurrent gout attacks: an internet-based case-crossover study. *Am J Med,* 127**,** 311-8.

NEOGI, T., HUNTER, D. J., CHAISSON, C. E., ALLENSWORTH-DAVIES, D. & ZHANG, Y. 2006. Frequency and predictors of inappropriate management of recurrent gout attacks in a longitudinal study. *J Rheumatol,* 33**,** 104-9.

NEUGEBAUER, R., VAN DER LAAN, M. J., JOFFE, M. M. & TAGER, I. B. 2007. Causal inference in longitudinal studies with history-restricted marginal structural models. *Electronic journal of statistics,* 1**,** 119-154.

NEUHÄUSER, M., THIELMANN, M. & RUXTON, G. D. 2018. The number of strata in propensity score stratification for a binary outcome. *Archives of medical science : AMS,* 14**,** 695-700.

NEYMAN, J. 1923. On the application of probability theory to agricultural experiments: Essay on principles, section 9. *Statistical Science,* 5**,** 465-472.

NGWA, J. S., CABRAL, H. J., CHENG, D. M., PENCINA, M. J., GAGNON, D. R., LAVALLEY, M. P. & CUPPLES, L. A. 2016. A comparison of time dependent Cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the Framingham Heart Study. *BMC Med Res Methodol,* 16**,** 148.

NICE. 2018a. *Gout - What are the complications* [Online]. National Institue of Health and Care Excellence. Available: https://cks.nice.org.uk/gout#!backgroundSub:3 [Accessed 25/03/2020].

NICE. 2018b. *Lesinurad for treating chronic hyperuricaemia in people with gout* [Online]. National Institue of Health and Care Excellence. Available: https://www.nice.org.uk/guidance/ta506/chapter/2-Information-about-lesinurad [Accessed 27/01/2020].

NICE. 2019. *NSAIDs - prescribing issues* [Online]. The National Institute for Health and Care Excellence. Available: https://cks.nice.org.uk/nsaids-prescribing-issues#!scenario [Accessed 22/01/2020].

NICHOLLS, A., SNAITH, M. L. & SCOTT, J. T. 1973. Effect of oestrogen therapy on plasma and urinary levels of uric acid. *British medical journal,* 1**,** 449-451.

PANDYA, B. J., RIEDEL, A. A., SWINDLE, J. P., BECKER, L. K., HARIRI, A., DABBOUS, O. & KRISHNAN, E. 2011. Relationship between physician specialty and allopurinol prescribing patterns: A study of patients with gout in managed care settings. *Current Medical Research and Opinion,* 27**,** 737-744.

PARISI, R., RUTTER, M. K., LUNT, M., YOUNG, H. S., SYMMONS, D. P., GRIFFITHS, C. E. & ASHCROFT, D. M. 2015. Psoriasis and the Risk of Major Cardiovascular Events: Cohort Study Using the Clinical Practice Research Datalink. *J Invest Dermatol,* 135**,** 2189-97.

PASCUAL, E. & PERDIGUERO, M. 2006. Gout, diuretics and the kidney. *Annals of the rheumatic diseases,* 65**,** 981-982.

PATORNO, E., GROTTO, A., BELLOCCO, R. & SCHNEEWEISS, S. 2013. Propensity score methodology for confounding control in health care utlization databases. *Epidemiology Biostatistics and Public Health,* 10**,** e89401 - e89401-16.

PATRICK, A. R., SCHNEEWEISS, S., BROOKHART, M. A., GLYNN, R. J., ROTHMAN, K. J., AVORN, J. & STÜRMER, T. 2011. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiology and drug safety,* 20**,** 551-559.

PCC. 2012. *QOF Business Rules V24.0* [Online]. Available: http://www.pcc-cic.org.uk/article/qof-business-rules-v240 [Accessed 23/05/16 2016].

PEARCE, N. 2016. Analysis of matched case-control studies. *BMJ,* 352**,** i969.

PEREZ-RUIZ, F., ALONSO-RUIZ, A., CALABOZO, M., HERRERO-BEITES, A., GARCIA-ERAUSKIN, G. & RUIZ-LUCEA, E. 1998. Efficacy of allopurinol and benzbromarone for the control of hyperuricaemia. A pathogenic approach to the treatment of primary chronic gout. *Annals of the Rheumatic Diseases,* 57**,** 545-549.

PEREZ-RUIZ, F., CALABOZO, M., PIJOAN, J., HERRERO-BEITES, A. M. & RUIBAL, A. 2002. Effect of urate-lowering therapy on the velocity of size reduction of tophi in chronic gout. *Arthritis & Rheumatism-Arthritis Care & Research,* 47**,** 356-360.

PEREZ-RUIZ, F., HERNANDEZ-BALDIZON, S., HERRERO-BEITES, A. M. & GONZALEZ-GAY, M. A. 2010. Risk factors associated with renal lithiasis during uricosuric treatment of hyperuricemia in patients with gout. *Arthritis Care and Research,* 62**,** 1299-1305.

PLATT, R. W., BROOKHART, M. A., COLE, S. R., WESTREICH, D. & SCHISTERMAN, E. F. 2013. An information criterion for marginal structural models. *Statistics in Medicine,* 32**,** 1383-1393.

PLATT, R. W., DELANEY, J. A. C. & SUISSA, S. 2012. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *European Journal of Epidemiology,* 27**,** 77-83.

PRIOR, J. A., OGOLLAH, R., MULLER, S., CHANDRATRE, P., RODDY, E. & MALLEN, C. D. 2015. Gout, anxiety, and depression in primary care: a matched retrospective cohort study. *Scandinavian Journal of Rheumatology,* 44**,** 257-258.

RABE-HESKETH, S. & SKRONDAL, A. 2012. *Multilevel and Longitudinal Modeling Using Stata*, Stata Press.

RAIT, G., WALTERS, K., GRIFFIN, M., BUSZEWICZ, M., PETERSEN, I. & NAZARETH, I. 2009. Recent trends in the incidence of recorded depression in primary care. *Br J Psychiatry,* 195**,** 520-4.

REES, F., JENKINS, W. & DOHERTY, M. 2013. Patients with gout adhere to curative treatment if informed appropriately: proof-of-concept observational study. *Annals of the Rheumatic Diseases,* 72**,** 826.

REEVES, D., SPRINGATE, D. A., ASHCROFT, D. M., RYAN, R., DORAN, T., MORRIS, R., OLIER, I. & KONTOPANTELIS, E. 2014. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case-control analysis. *BMJ Open,* 4**,** e004952.

REILLY, S., OLIER, I., PLANNER, C., DORAN, T., REEVES, D., ASHCROFT, D. M., GASK, L. & KONTOPANTELIS, E. 2015. Inequalities in physical comorbidity: a longitudinal comparative cohort study of people with severe mental illness in the UK. *BMJ Open,* 5.

RICHETTE, P., DOHERTY, M., PASCUAL, E., BARSKOVA, V., BECCE, F., CASTAÑEDA-SANABRIA, J., COYFISH, M., GUILLO, S., JANSEN, T. L., JANSSENS, H., LIOTÉ, F., MALLEN, C., NUKI, G., PEREZ-RUIZ, F., PIMENTAO, J., PUNZI, L., PYWELL, T., SO, A., TAUSCHE, A. K., UHLIG, T., ZAVADA, J., ZHANG, W., TUBACH, F. & BARDIN, T. 2017. 2016 updated EULAR evidence-based recommendations for the management of gout. *Annals of the Rheumatic Diseases,* 76**,** 29.

RICHETTE, P., DOHERTY, M., PASCUAL, E., BARSKOVA, V., BECCE, F., CASTANEDA, J., COYFISH, M., GUILLO, S., JANSEN, T., JANSSENS, H., LIOTÉ, F., MALLEN, C. D., NUKI, G., PEREZ-RUIZ, F., PIMENTAO, J., PUNZI, L., PYWELL, A., SO, A. K., TAUSCHE, A.-K., UHLIG, T., ZAVADA, J., ZHANG, W., TUBACH, F. & BARDIN, T. 2020. 2018 updated European League Against Rheumatism evidence-based recommendations for the diagnosis of gout. *Annals of the Rheumatic Diseases,* 79**,** 31.

ROBINS, J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Math Modelling,* 7**,** 1393-1512.

ROBINS, J. M. Marginal Structural Models versus Structural nested Models as Tools for Causal inference. 2000 New York, NY. Springer New York, 95-133.

ROBINS, J. M., BLEVINS, D., RITTER, G. & WULFSOHN, M. 1992. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology,* 3**,** 319-36.

ROBINS, J. M., GREENLAND, S. & HU, F.-C. 1999. Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome. *Journal of the American Statistical Association,* 94**,** 687-700.

ROBINS, J. M., HERNÁN, M. Á. & BRUMBACK, B. 2000. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology,* 11**,** 550-560.

RODDY, E. 2011. Revisiting the pathogenesis of podagra: why does gout target the foot? *Journal of Foot and Ankle Research,* 4**,** 13.

RODDY, E. & CHOI, H. K. 2014. Epidemiology of gout. *Rheum Dis Clin North Am.,* 40**,** 155-75.

RODDY, E., CLARKSON, K., BLAGOJEVIC-BUCKNALL, M., MEHTA, R., OPPONG, R., AVERY, A., HAY, E. M., HENEGHAN, C., HARTSHORNE, L., HOOPER, J., HUGHES, G., JOWETT, S., LEWIS, M., LITTLE, P., MCCARTNEY, K., MAHTANI, K. R., NUNAN, D., SANTER, M., WILLIAMS, S. & MALLEN, C. D. 2019. Open-label randomised pragmatic trial (CONTACT) comparing naproxen and low-dose colchicine for the treatment of gout flares in primary care. *Annals of the Rheumatic Diseases***,** annrheumdis-2019-216154.

RODDY, E., MALLEN, C. D. & DOHERTY, M. 2013. Gout. *BMJ,* 347**,** f5648.

RODDY, E., MALLEN, C. D., HIDER, S. L. & JORDAN, K. P. 2010. Prescription and comorbidity screening following consultation for acute gout in primary care. *Rheumatology (Oxford),* 49**,** 105-11.

RODDY, E., ZHANG, W. & DOHERTY, M. 2007a. Are joints affected by gout also affected by osteoarthritis? *Annals of the rheumatic diseases,* 66**,** 1374-1377.

RODDY, E., ZHANG, W. & DOHERTY, M. 2007b. Concordance of the management of chronic gout in a UK primary-care population with the EULAR gout recommendations. *Ann Rheum Dis,* 66**,** 1311-5.

ROSENBAUM, P. R. 1987. Model-Based Direct Adjustment. *Journal of the American Statistical Association,* 82**,** 387-394.

ROSENBAUM, P. R. & RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70**,** 41-55.

ROSENBAUM, P. R. & RUBIN, D. B. 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association,* 79**,** 516-524.

ROTHENBACHER, D., PRIMATESTA, P., FERREIRA, A., CEA-SORIANO, L. & RODRIGUEZ, L. A. 2011. Frequency and risk factors of gout flares in a large population-based cohort of incident gout. *Rheumatology (Oxford),* 50**,** 973-81.

ROUGHLEY, M., SULTAN, A. A., CLARSON, L., MULLER, S., WHITTLE, R., BELCHER, J., MALLEN, C. D. & RODDY, E. 2018. Risk of chronic kidney disease in patients with gout and the impact of urate lowering therapy: a population-based cohort study. *Arthritis Research & Therapy,* 20**,** 243.

ROUGHLEY, M. J., BELCHER, J., MALLEN, C. D. & RODDY, E. 2015. Gout and risk of chronic kidney disease and nephrolithiasis: meta-analysis of observational studies. *Arthritis Res Ther,* 17**,** 90.

ROYSTON, P., AMBLER, G. & SAUERBREI, W. 1999. The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology,* 28**,** 964-974.

RUBIN, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology,* 66**,** 688-701.

RUBIN, D. B. 1980. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association,* 75**,** 591-593.

RUBIN, D. B. 1986. Comment: What ifs have causal answers. *Journal of the American Statistical Association,* 81**,** 961-962.

RUBIN, D. B. & THOMAS, N. 1996. Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics,* 52**,** 249-264.

RUDOLPH, K. E., COLSON, K. E., STUART, E. A. & AHERN, J. 2016. Optimally combining propensity score subclasses. *Statistics in medicine,* 35**,** 4937-4947.

RUTHERFORD, M. J., CROWTHER, M. J. & LAMBERT, P. C. 2015. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation,* 85**,** 777-793.

SANSON-FISHER, R. W., BONEVSKI, B., GREEN, L. W. & D'ESTE, C. 2007. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med,* 33**,** 155-61.

SCHEEPERS, L., BURDEN, A. M., ARTS, I. C. W., SPAETGENS, B., SOUVEREIN, P., DE VRIES, F. & BOONEN, A. 2018. Medication adherence among gout patients initiated allopurinol: a retrospective cohort study in the Clinical Practice Research Datalink (CPRD). *Rheumatology (Oxford),* 57**,** 1641-1650.

SCHNEEWEISS, S. & AVORN, J. 2005. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol,* 58**,** 323-37.

SCHUMACHER, H. R., JR., BECKER, M. A., WORTMANN, R. L., MACDONALD, P. A., HUNT, B., STREIT, J., LADEMACHER, C. & JOSEPH-RIDGE, N. 2008. Effects of febuxostat versus allopurinol and placebo in reducing serum urate in subjects with hyperuricemia and gout: a 28-week, phase III, randomized, double-blind, parallel-group trial. *Arthritis Rheum,* 59**,** 1540-8.

SEDGWICK, P. 2014. How to read a Kaplan-Meier survival plot. *BMJ : British Medical Journal,* 349**,** g5608.

SEEGMILLER, J. E. 1965. The acute attack of gouty arthritis. *Arthritis & Rheumatism,* 8**,** 714-725.

SETH, R., KYDD, A., BUCHBINDER, R., BOMBARDIER, C. & EDWARDS, C. 2014. Allopurinol for chronic gout. *Cochrane Database Syst Rev.,* 10.

SHAH, B. R., LAUPACIS, A., HUX, J. E. & AUSTIN, P. C. 2005. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol,* 58**,** 550-9.

SHIH, H. J., KAO, M. C., TSAI, P. S., FAN, Y. C. & HUANG, C. J. 2017. Long-term allopurinol use decreases the risk of prostate cancer in patients with gout: a population-based study. *Prostate Cancer and Prostatic Diseases,* 20**,** 328-333.

SINGH, J. A. & CLEVELAND, J. 2018a. Allopurinol and the risk of incident peripheral arterial disease in the elderly: a US Medicare claims data study. *Rheumatology (Oxford),* 57**,** 451-461.

SINGH, J. A. & CLEVELAND, J. D. 2018b. Comparative effectiveness of allopurinol versus febuxostat for preventing incident dementia in older adults: a propensity-matched analysis. *Arthritis Research & Therapy,* 20**,** 167.

SINGH, J. A., RAMACHANDARAN, R., YU, S. & CURTIS, J. R. 2017. Allopurinol use and the risk of acute cardiovascular events in patients with gout and diabetes. *BMC cardiovascular disorders,* 17**,** 76-76.

SINGH, J. A. & YU, S. 2016. Allopurinol reduces the risk of myocardial infarction (MI) in the elderly: a study of Medicare claims. *Arthritis research & therapy,* 18**,** 209-209.

SMITH, A. R. & SCHAUBEL, D. E. 2015. Time-dependent prognostic score matching for recurrent event analysis to evaluate a treatment assigned during follow-up. *Biometrics,* 71**,** 950-9.

SPRINGATE, D. A., ASHCROFT, D. M., KONTOPANTELIS, E., DORAN, T., RYAN, R. & REEVES, D. 2015. Can analyses of electronic patient records be independently and externally validated? Study 2--the effect of beta-adrenoceptor blocker therapy on cancer survival: a retrospective cohort study. *BMJ Open,* 5**,** e007299.

SPRINGATE, D. A., KONTOPANTELIS, E., ASHCROFT, D. M., OLIER, I., PARISI, R., CHAMAPIWA, E. & REEVES, D. 2014. ClinicalCodes: An Online Clinical Codes Repository to Improve the Validity and Reproducibility of Research Using Electronic Medical Records. *PLoS ONE,* 9**,** e99825.

STACK, A. G., JOHNSON, M. E., BLAK, B., KLEIN, A., CARPENTER, L., MORLOCK, R., MAGUIRE, A. R. & PARSONS, V. L. 2019. Gout and the risk of advanced chronic kidney disease in the UK health system: a national cohort study. *BMJ Open,* 9**,** e031550.

STAMP, L., GOW, P., SHARPLES, K. & RAILL, B. 2000. The optimal use of allopurinol: an audit of allopurinol use in South Auckland. *Australian and New Zealand journal of medicine,* 30**,** 567-72.

STAMP, L. K., BARCLAY, M. L., O'DONNELL, J. L., ZHANG, M., DRAKE, J., FRAMPTON, C. & CHAPMAN, P. T. 2011a. Relationship Between Serum Urate and Plasma Oxypurinol in the Management of Gout: Determination of Minimum Plasma Oxypurinol Concentration to Achieve a Target Serum Urate Level. *Clinical Pharmacology & Therapeutics,* 90**,** 392-398.

STAMP, L. K., O'DONNELL, J. L., ZHANG, M., JAMES, J., FRAMPTON, C., BARCLAY, M. L. & CHAPMAN, P. T. 2011b. Using allopurinol above the dose based on creatinine clearance is effective and safe in patients with chronic gout, including those with renal impairment. *Arthritis and Rheumatism,* 63**,** 412-421.

STAMP, L. K., TAYLOR, W. J., JONES, P. B., DOCKERTY, J. L., DRAKE, J., FRAMPTON, C. & DALBETH, N. 2012. Starting dose is a risk factor for allopurinol hypersensitivity syndrome: a proposed safe starting dose of allopurinol. *Arthritis Rheum,* 64**,** 2529-36.

STAMP, L. K., WELLS, J. E., PITAMA, S., FAATOESE, A., DOUGHTY, R. N., WHALLEY, G., RICHARDS, A. M. & CAMERON, V. A. 2013. Hyperuricaemia and gout in New Zealand rural and urban Maori and non-Maori communities. *Internal Medicine Journal,* 43**,** 678-684.

STEWART, D., HAN, L., DORAN, T. & MCCAMBRIDGE, J. 2017. Alcohol consumption and all-cause mortality: an analysis of general practice database records for patients with long-term conditions. *Journal of Epidemiology and Community Health,* 71**,** 729.

STOCKS, S. J., KONTOPANTELIS, E., AKBAROV, A., RODGERS, S., AVERY, A. J. & ASHCROFT, D. M. 2015. Examining variations in prescribing safety in UK general practice: cross sectional study using the Clinical Practice Research Datalink. *BMJ,* 351.

STREETER, A. J., LIN, N. X., CRATHORNE, L., HAASOVA, M., HYDE, C., MELZER, D. & HENLEY, W. E. 2017. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *Journal of clinical epidemiology,* 87**,** 23-34.

STUART, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science,* 25**,** 1-21.

STUART, E. A., DUGOFF, E., ABRAMS, M., SALKEVER, D. & STEINWACHS, D. 2013a. Estimating causal effects in observational studies using Electronic Health Data: Challenges and (some) solutions. *EGEMS (Wash DC),* 1.

STUART, E. A., LEE, B. K. & LEACY, F. P. 2013b. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of clinical epidemiology,* 66**,** S84-S90.e1.

STUART, E. A. & RUBIN, D. B. 2008. Best Practices in Quasi–Experimental Designs: Matching Methods for Causal Inference. *Best Practices in Quantitative Methods.* SAGE Publications Ltd.

STURROCK, R. D. 2000. Gout. Easy to misdiagnose. *BMJ (Clinical research ed.),* 320**,** 132-133.

SULTAN, A. A., MULLER, S., WHITTLE, R., RODDY, E., MALLEN, C. & CLARSON, L. 2019. Venous thromboembolism in patients with gout and the impact of hospital admission, disease duration and urate-lowering therapy. *CMAJ,* 191**,** E597-E603.

SULTAN, A. A., WHITTLE, R., MULLER, S., RODDY, E., MALLEN, C. D., BUCKNALL, M., HELLIWELL, T., HIDER, S. & PASKINS, Z. 2018. Risk of fragility fracture among patients with gout and the effect of urate-lowering therapy. *Cmaj,* 190**,** E581-e587.

TALBOT, D., ATHERTON, J., ROSSI, A. M., BACON, S. L. & LEFEBVRE, G. 2015. A cautionary note concerning the use of stabilized weights in marginal structural models. *Stat Med,* 34**,** 812-23.

TAYLOR, K. 2013. *Uric acid concentration conversion* [Online]. GoutPal Gout Help. Available: http://www.goutpal.com/449/uric-acid-concentration/ [Accessed 01/02/2016].

TER MAATEN, J. C., VOORBURG, A., HEINE, R. J., TER WEE, P. M., DONKER, A. J. & GANS, R. O. 1997. Renal handling of urate and sodium during acute physiological hyperinsulinaemia in healthy subjects. *Clin Sci (Lond),* 92**,** 51-8.

TERKELTAUB, R. A., FURST, D. E., BENNETT, K., KOOK, K. A., CROCKETT, R. S. & DAVIS, M. W. 2010. High versus low dosing of oral colchicine for early acute gout flare: Twenty-four-hour outcome of the first multicenter, randomized, double-blind, placebo-controlled, parallel-group, dose-comparison colchicine study. *Arthritis Rheum,* 62**,** 1060-8.

THANASSOULIS, G., BROPHY, J. M., RICHARD, H. & PILOTE, L. 2010. Gout, allopurinol use, and heart failure outcomes. *Arch Intern Med,* 170**,** 1358-64.

TRIFIRO, G., MORABITO, P., CAVAGNA, L., FERRAJOLO, C., PECCHIOLI, S., SIMONETTI, M., BIANCHINI, E., MEDEA, G., CRICELLI, C., CAPUTI, A. P. & MAZZAGLIA, G. 2013. Epidemiology of gout and hyperuricaemia in Italy during the years 2005-2009: a nationwide population-based study. *Ann Rheum Dis,* 72**,** 694-700.

TZENG, H.-E., LIN, C.-C., WANG, I. K., HUANG, P.-H. & TSAI, C.-H. 2016. Gout increases risk of fracture: A nationwide population-based cohort study. *Medicine,* 95**,** e4669-e4669.

UDDIN, M. J., GROENWOLD, R. H. H., ALI, M. S., DE BOER, A., ROES, K. C. B., CHOWDHURY, M. A. B. & KLUNGEL, O. H. 2016. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *International Journal of Clinical Pharmacy,* 38**,** 714-723.

UNDERWOOD, M. 2006. Diagnosis and management of gout. *BMJ (Clinical research ed.),* 332**,** 1315-1319.

VAN DER WAL, W. M. & GESKUS, R. B. 2011. ipw: An R Package for Inverse Probability Weighting. *2011,* 43**,** 23.

VANDECANDELAERE, M., VANSTEELANDT, S., DE FRAINE, B. & VAN DAMME, J. 2016. Time-Varying Treatments in Observational Studies: Marginal Structural Models of the Effects of Early Grade Retention on Math Achievement. *Multivariate Behav Res,* 51**,** 843-864.

VANDERWEELE, T. J. & DING, P. 2017. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med,* 167**,** 268-274.

VARGAS-SANTOS, A. B., NEOGI, T., DA ROCHA CASTELAR-PINHEIRO, G., KAPETANOVIC, M. C. & TURKIEWICZ, A. 2019. Cause-Specific Mortality in Gout: Novel Findings of Elevated Risk of Non-Cardiovascular-Related Deaths. *Arthritis Rheumatol,* 71**,** 1935-1942.

VARGAS-SANTOS, A. B., PELOQUIN, C. E., ZHANG, Y. & NEOGI, T. 2018. Association of Chronic Kidney Disease With Allopurinol Use in Gout Treatment. *JAMA Intern Med,* 178**,** 1526-1533.

VAZQUEZ-MELLADO, J., MORALES, E. M., PACHECO-TENA, C. & BURGOS-VARGAS, R. 2001. Relation between adverse events associated with allopurinol and renal function in patients with gout. *Annals of the rheumatic diseases,* 60**,** 981-3.

VITTINGHOFF, E. & MCCULLOCH, C. E. 2006. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology,* 165**,** 710-718.

VOURLI, G. & TOULOUMI, G. 2015. Performance of the marginal structural models under various scenarios of incomplete marker's values: a simulation study. *Biom J,* 57**,** 254-70.

WALTERS, K., RAIT, G., GRIFFIN, M., BUSZEWICZ, M. & NAZARETH, I. 2012. Recent Trends in the Incidence of Anxiety Diagnoses and Symptoms in Primary Care. *PLoS ONE,* 7**,** e41670.

WEI, L., MACKENZIE, I. S., CHEN, Y., STRUTHERS, A. D. & MACDONALD, T. M. 2011. Impact of allopurinol use on urate concentration and cardiovascular outcome. *Br J Clin Pharmacol,* 71**,** 600-7.

WEITZEN, S., LAPANE, K. L., TOLEDANO, A. Y., HUME, A. L. & MOR, V. 2004. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf,* 13**,** 841-53.

WEITZEN, S., LAPANE, K. L., TOLEDANO, A. Y., HUME, A. L. & MOR, V. 2005. Weaknesses of goodness-of-fit tests for evaluating propensity score models: the case of the omitted confounder. *Pharmacoepidemiol Drug Saf,* 14**,** 227-38.

WELCH, C. A., PETERSEN, I., BARTLETT, J. W., WHITE, I. R., MARSTON, L., MORRIS, R. W., NAZARETH, I., WALTERS, K. & CARPENTER, J. 2014. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Stat Med,* 33**,** 3725-37.

WESTREICH, D., COLE, S. R., FUNK, M. J., BROOKHART, M. A. & STURMER, T. 2011. The role of the c-statistic in variable selection for propensity score models. *Pharmacoepidemiol Drug Saf,* 20**,** 317-20.

WHITE, W. B., SAAG, K. G., BECKER, M. A., BORER, J. S., GORELICK, P. B., WHELTON, A., HUNT, B., CASTILLO, M. & GUNAWARDHANA, L. 2018. Cardiovascular Safety of Febuxostat or Allopurinol in Patients with Gout. *New England Journal of Medicine,* 378**,** 1200-1210.

WU, M., LIU, F. J., CHEN, J., CHEN, L., WEI, C., HU, Z. M., HAN, Y., LU, J. X., JIANG, L. X. & CHEN, H. B. 2019. Prevalence and Factors Associated With Bone Erosion in Patients With Gout. *Arthritis Care & Research,* 71**,** 1653-1659.

XIAO, Y., ABRAHAMOWICZ, M., MOODIE, E. E. M., WEBER, R. & YOUNG, J. 2014. Flexible Marginal Structural Models for Estimating the Cumulative Effect of a Time-Dependent Treatment on the Hazard: Reassessing the Cardiovascular Risks of Didanosine

Treatment in the Swiss HIV Cohort Study. *Journal of the American Statistical Association,* 109**,** 455-464.

XIAO, Y., ABRAHAMOWICZ, M. & MOODIE ERICA, E. M. 2010. Accuracy of Conventional and Marginal Structural Cox Model Estimators: A Simulation Study. *The International Journal of Biostatistics.*

XIAO, Y., MOODIE ERICA, E. M. & ABRAHAMOWICZ, M. 2013. Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models. *Epidemiologic Methods.*

XUE, F., GOLI, V., PETRARO, P., MCMULLAN, T., SPRAFKA, J. M. & TCHETGEN TCHETGEN, E. J. 2017. Marginal structural model to evaluate the association between cumulative osteoporosis medication and infection using claims data. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA,* 28**,** 2893-2901.

YANG, S., EATON, C. B., LU, J. & LAPANE, K. L. 2014. Application of marginal structural models in pharmacoepidemiologic studies: a systematic review. *Pharmacoepidemiol Drug Saf,* 23**,** 560-71.

YANG, S., EATON, C. B., MCALINDON, T. E. & LAPANE, K. L. 2015a. Effects of glucosamine and chondroitin supplementation on knee osteoarthritis: an analysis with marginal structural models. *Arthritis Rheumatol,* 67**,** 714-23.

YANG, S., LU, J., EATON, C. B., HARPE, S. & LAPANE, K. L. 2015b. The Choice of Analytical Strategies in Inverse-Probability-of-Treatment-Weighted Analysis: A Simulation Study. *Am J Epidemiol,* 182**,** 520-7.

YU, T. F. & GUTMAN, A. B. 1961. Efficacy of colchicine prophylaxis in gout. Prevention of recurrent gouty arthritis over a mean period of five years in 208 gouty subjects. *Ann Intern Med,* 55**,** 179-92.

ZEGER, S. L., LIANG, K.-Y. & ALBERT, P. S. 1988. Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics,* 44**,** 1049-1060.

ZHANG, M., SOLOMON, D. H., DESAI, R. J., KANG, E. H., LIU, J., NEOGI, T. & KIM, S. C. 2018. Assessment of Cardiovascular Risk in Older Patients With Gout Initiating Febuxostat Versus Allopurinol: Population-Based Cohort Study. *Circulation,* 138**,** 1116-1126.

ZHANG, W., DOHERTY, M., BARDIN, T., PASCUAL, E., BARSKOVA, V., CONAGHAN, P., GERSTER, J., JACOBS, J., LEEB, B., LIOTE, F., MCCARTHY, G., NETTER, P., NUKI, G., PEREZ-RUIZ, F., PIGNONE, A., PIMENTAO, J., PUNZI, L., RODDY, E., UHLIG, T., ZIMMERMANN-GORSKA, I. & EULAR STANDING COMMITTEE FOR INTERNATIONAL CLINICAL STUDIES INCLUDING THERAPEUTICS 2006a. EULAR evidence based recommendations for gout. Part II: Management. Report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis,* 65**,** 1312-24.

ZHANG, W., DOHERTY, M., PASCUAL, E., BARDIN, T., BARSKOVA, V., CONAGHAN, P., GERSTER, J., JACOBS, J., LEEB, B., LIOTE, F., MCCARTHY, G., NETTER, P., NUKI, G., PEREZ-RUIZ, F., PIGNONE, A., PIMENTAO, J., PUNZI, L., RODDY, E., UHLIG, T. & ZIMMERMANN-GORSKA, I. 2006b. EULAR evidence based recommendations for gout. Part I: Diagnosis. Report of a task force of the standing committee for international clinical studies including therapeutics (ESCISIT). *Ann Rheum Dis.,* 65**,** 1301-11.

ZHANG, Y., NEOGI, T., CHEN, C., CHAISSON, C., HUNTER, D. J. & CHOI, H. 2014. Low-dose aspirin use and recurrent gout attacks. *Annals of the rheumatic diseases,* 73**,** 385-390.

ZHANG, Y., NEOGI, T., CHEN, C., CHAISSON, C., HUNTER, D. J. & CHOI, H. K. 2012. Cherry consumption and decreased risk of recurrent gout attacks. *Arthritis Rheum,* 64**,** 4004-11.

13 Appendices

Appendix A Literature review protocol

Arthritis Research UK Primary Care Centre
Systematic Review Protocol & Support Template

This template is primarily intended to help you plan your review in a systematic way. A copy of this completed form will be available via the intranet to help others carrying out reviews in the future and to avoid duplicating work already undertaken in the Centre. Keeping a record of all the reviews will also assist in planning the work of the Centre and ensuring adequate methodological support. Not all the information will be relevant to every review. However, items can be adapted to fit the type of review that is being undertaken.

Please complete the form in as much detail as possible for your review and email to Jo Jordan, j.jordan@cphc.keele.ac.uk

Title of the review	Statistical methods used to control for confounding in treatment effect estimation of allopurinol in gout: a literature review of observational studies
First reviewer	Trishna Rathod
Team of reviewers	N/A
Supervisor/Project PI	Milisa Bucknall & Ed Roddy
Clinical Portfolio Group	Inflammatory arthritis
Project title (if different from review title)	N/A

Support – please state if advice/training or personnel required at each stage	
SR overview	01/04/14 – Overall advice on the literature review protocol given by Jo Jordan
Protocol development	N/A
Literature searching	N/A
Quality appraisal	N/A
Data Extraction	N/A
Synthesis	N/A
Writing up	N/A

1. Background to review

Brief introduction to the subject of the review, including rationale for undertaking the review and overall aim

Gout is the most prevalent type of inflammatory arthritis affecting 2.5% of the UK population. It is a chronic and progressive disease caused by elevated levels of serum urate in the blood leading to deposits of monosodium urate crystals in and around the joints. Gout is characterised as acute flares of severe pain and swelling at the affected joint site. Although there is no cure, current treatment aims to treat inflammation of the acute gout flare followed by urate-lowering therapy to prevent recurrent flares.

The EULAR, BSR, and ACR have all published guidelines on the management of gout. They advocate allopurinol as the first line urate-lowering drug of choice to treat chronic gout. Although allopurinol has been available since 1963, no randomised controlled trials (RCTs) have been conducted to establish its long-term efficacy. In recent RCTs determining the efficacy of newer urate-lowering drugs, allopurinol was used as the comparator group. The effect of allopurinol (vs. non- use) has been estimated from cohort studies. Cohort studies have established that allopurinol lowers serum urate level, reduces the frequency of gout flares, and is well-tolerated. Consequently, although allopurinol is recommended as the first line urate-lowering therapy, further evidence is needed of its long-term effect on a range of gout outcomes.

To infer a causal effect of allopurinol, the ignorable treatment assignment assumption needs to be satisfied i.e., conditional on participant characteristics (covariates), the assignment of study participants to binary treatment (allopurinol vs non-allopurinol) is independent of the outcome of non-allopurinol treatment and the outcome of the allopurinol treatment. In RCTs, this assumption holds as randomisation minimises differences of measured and unmeasured covariates between treatment groups. However, this assumption is not satisfied in observational studies as treatment assignment is dependent on measured covariates resulting in differences between the treatment groups, introducing confounding hence the causal effect is biased.

Statistical methods in varying degrees of complexity have been used to control for confounding from measured covariates. Traditional methods of multivariable regression models, matching and stratification are commonly used to control for confounding. However, such methods are limited to the number to covariates that can be controlled for. Alternatively, propensity score (PS) methodology is not constricted by this and aims to estimate the treatment effect by accounting for covariates that predict treatment assignment thus balancing the observed covariates between the treatment groups. However, a pitfall of PS methodology (and traditional methods) is that not all confounders are known or measured. Instrumental variables, frailty and Bayesian models can account for the heterogeneity from unmeasured covariates.

This literature review aims to identify the statistical methodologies used to model the effect of allopurinol in gout outcomes from observational studies.

2. Specific objectives

1) To describe the range of study designs and statistical methods used to control for confounding in modelling the effect of allopurinol on gout outcomes.
2) To compare and contrast the suitability and the limitations of the statistical methods identified.

3. Criteria for including studies in the review

If the PICOS format does not fit the research question of interest, please split up the question into separate concepts and put one under each heading

Population, or participants and conditions of interest	Search terms for gout in title and abstract: 1) MeSH term for gout 2) Gout 3) Gout* 4) Podagra 5) Arthragra 6) Chiragra 7) Toph* 8) 1 or 2 or 3 or 4 or 5 or 6 or 7
Interventions or exposures	Search terms for allopurinol in title and abstract: 9) Allopurinol (Check is mesh term for allopurinol is available) 10) Xanthine oxidase inhibit* 11) Urate-lowering 12) Urate lowering 13) Uric acid-lowering 14) Uric acid lowering 15) Uricostatic 16) Abburic OR Abopur OR Acepurin OR Acifugan OR Acyprin OR Alfadiman OR Algut OR Alinol OR allo* OR Allpargin OR Allupol OR Allura* OR Alluri* OR Aloprim OR Alopur OR Aloral OR Alosfar OR Alpur* OR Aluline OR Aluprol OR Aluron OR Alzoprim OR Anurate OR Apnol OR Apo-Tinole OR Apronol OR Apulonga OR Apurin OR Apurol OR Arnol OR Arsol OR Artrex OR Arturic OR Atisuril OR Aurigen OR Benoxuric OR Be-Uric OR Bionol OR Biuricowas OR Bleminol OR Caplenal OR Capurate OR Cellidrin* OR Chinnol OR Ciploric OR Colpuril OR Comburic OR Cosuric OR Dabroson OR Darzune OR Desatura OR Docallopu OR Duovitan OR "dura AL" OR Elavil OR Embarin OR Epidropal OR Erloric OR Ethipurinol OR Etindrax OR Facilit OR Foligan OR Gealgica OR Gewapurol OR Gichtex OR Gotir OR Hamarin OR Harpagin OR Hexanurat OR Hycemia OR Isoric OR Jenapurinol OR Labopurinol OR Labypurol OR Lanolone OR Licoric OR Llanol OR Lonol OR Lop*ric OR Lopur* OR Loricid OR Lo-Uric OR Lysuron OR Marinol OR Medoric OR Mephanol OR Milurit OR Nilapur OR Novo-Purol OR Oloprim OR Petrazyc OR Ponuric OR Prinol OR Pritanol OR Progout OR Pureduct OR Puricemia OR Puricin OR Puricos OR Puride OR Purigan OR Purinase OR Purinol OR Purispec OR Puristen OR Puritenk OR Pyrazol OR Ranpuric OR Redurate OR Remid OR Reucid OR Rimapurinol OR Rinolic OR Sigapurol OR Sinoric OR Soluric OR Stradumel OR Suspendol OR Synol OR Synpurinol OR Talol OR Tipuric OR Trianol OR Tylonic OR Unizuric OR Uredimin OR Uribenz OR Urica* OR Uricemil OR Uricina OR Uricnol OR Urico* OR Urifugan OR

	Urikoliz OR Urinol OR Uriprim OR Uripurinol OR Uritab OR Urobenyl OR Urogotan OR Uroplus OR Urosi OR Urosin OR Urozyl-SR OR Urtias OR Valeric OR Xandase OR Xanol OR Xanthomax OR Xanturic OR Xanurace OR Xuric-A OR "Z 300" OR Zilopur OR Zurim OR Zygout OR Zylapour OR Zylic OR Zylo* 17) 9 or 10 or 11 or 12 or 13 or 14 or 16 18) 8 and 17
Comparisons or control groups	Any comparator that may include: • Non-steroidal anti-inflammatory drugs • Corticosteroids • Colchicine • Febuxostat • Adrenocorticotrophic hormone • Interleukin-1 inhibitors (anakinra, canakinumab) • Pegloticase • Non-pharmacological intervention • Placebo • Different dose of allopurinol • Allopurinol in combination with another drug • Usual care • Uricosuric drugs (sulfinpyazone, probenecid, benzbromarone, losartan, fenfibrate, atorvastin, lefluomide) • No active intervention Note, these comparators will not be searched for.
Outcomes of interest	Any outcome that may include: • Pain • Joint impairment • Joint inflammation • Joint damage imaging • Serum urate levels • Acute gout flares • Tophus burden • Comorbidities • Activity limitations • Participation restrictions (e.g., employment) • Work disability • Healthcare utilisation • Cos • Patient utility • Patient/Physician global assessment • Patient global assessment of disease scales • Health Related Quality of Life • Acute phase markers • Adverse effects from allopurinol use: skin and subcutaneous tissue disorders such as rash, severe cutaneous adverse reactions (including Stevens-Johnson syndrome, toxic-epidermal necrolysis), and eosinophilia • Mortality • Oxypurinol concentrations • Any other medical conditions Note, these outcomes will not be searched for.
Setting	All settings
Study designs	Observational studies (cohort study, case-cohort study, cross sectional etc)

Statistical methods	Statistical methods used to control for confounding: Propensity scores Regression models G estimation Frailty models Instrumental variables Stratification Matching

4. Criteria for excluding studies not covered in inclusion criteria Any specific populations excluded, date range, language, whether abstracts or full text available, etc
• Abstracts from grey literature (e.g., conferences) • Papers not written in English • Randomised trials • Unpublished material • Case reports of individual patients • Commentaries on articles • Letters to the editor • Paper is on the diagnosis/strategies/ guidelines in the treatment of gout • Prevalence studies • Narratives Studies on children (<18 years)

5. Search methods	
Electronic databases Please list all databases that are to be searched and include the interface (eg NHS, EBSCO, etc) and date ranges searched for each	Interface: NHS evidence https://www.evidence.nhs.uk/about-evidence-services/journals-and-databases CINAHL (1981 onwards) Interface: Ovidsp embASE: Excerpta Medica Database (1980 onwards) Medline: General medical database (1946 onwards) Interface: Web of Science http://www.keele.ac.uk/healthlibrary/find/medicaldatabaseskeele/Science citation index (1964 onwards) Web of Science Interface: Ebsco http://www.keele.ac.uk/healthlibrary/find/medicaldatabaseskeele/ AgeLine (1978 onwards)

Other methods used for identifying relevant research ie contacting experts and reference checking	Checking the reference list of the eligible studies identified.
Journals hand searched If any are to be hand searched, please list which journals and date searched from, including a rationale.	N/A

6. Methods of review	
Details of methods Number of reviewers, how agreements to be reached and disagreements dealt with, etc.	From each database, conduct the search and save the results into an appropriate file format. Import the files into Refworks and save the search results into a folder called '1. All studies'. Remove duplicates and save the remaining studies into a new folder called '2. Studies without duplicates'. The titles would be screened for eligibility. Eligible titles would be saved into a folder called '3. Eligible titles' Abstracts of the eligible titles would be obtained and screened for eligibility. Eligible abstracts would be saved into a folder called '4. Eligible abstracts' The papers of the eligible abstracts would be obtained and screened. Eligible papers would be saved into a folder called '5. Eligible papers'
Quality assessment Tools or checklists used with references or URLs	Hierarchy on the quality of the study: • Systematic reviews of observational studies • Cohort and case control studies • Cross-sectional studies Hierarchy of statistical methods used to adjust for confounding: • Propensity scores and other novel methods (G estimation, instrumental variables, frailty models, Bayesian models) • Traditional methods (Regression models, stratification, matching) and sensitivity/subgroup analysis • No adjustment for confounding (descriptive statistics e.g., chi square tests) Hierarchy of type of outcome/treatment analysed: • Repeated measures (time-varying treatment/multiple outcomes) • Only analysed a single point in time Hierarchy of methods used to deal with missing data: • Multiple imputation • Last observation taken forward/single imputation/missing indicator method • Complete case analysis

Data extraction What information is to be collected on each included study. If databases or forms on Word or Excel are used and how this is recorded and by how many reviewers **Narrative synthesis** Details of what and how synthesis will be done	Themes to be identified: • Study population (gout/ hyperuricaemia patients/severity of disease) • Study design (cohort/cross-sectional/case-control) • How data was collected (medical records/self-report questionnaires) • Sample size • Intervention (drug name, duration, dosage, mode of administration) • Comparator pharmacological intervention (drug name, duration, dosage, frequency, mode of administration) • Comparator of non-pharmacological intervention (description of intervention, duration, frequency) • Primary and secondary outcomes • Length of follow up period • Statistical methods used • Adjusted covariates to reduce confounding • Missing data • Any other novel techniques • Limitations of study This would be recorded in an excel file with the studies listed in a column and the themes across the row.
Meta-analysis Details of what and how analysis and testing will be done. If no meta-analysis is to be conducted, please give reason.	The number of studies identified by this search is expected to be low. Combining results across the studies would be difficult as there are a wide range of gout outcomes and the statistical methods used may vary.
Grading evidence System used, if any, such as GRADE	N/A

7. Presentation of results	
Additional material Summary tables, flowcharts, etc, to be included in the final paper	The PRISMA flow chart would be used showing the total number of studies screened, assessed for eligibility, included in the review with reasons why the ineligible studies were excluded. In tables, describe the: Study characteristics Statistical methods
Outputs from review Papers and target journals, conference presentations, reports, etc	Results from this review is for the PhD.

8. Timeline for review – when do you aim to complete each stage of the review	
Protocol	18th April
Literature searching	21st April – 9th May
Quality appraisal	12th May – 30th May
Data extraction	
Synthesis	
Writing up	2nd June – 31st June

Appendix B Search strategy in MEDLINE

Table B1: MEDLINE search strategy

Item	Search Terms
1	exp Gout/
2	gout.ab. or gout.ti.
3	gout$.ab. or gout$.ti.
4	podagra.ab. or podagra.ti.
5	artagra.ab. or artagra.ti.
6	chiragra.ab. or chiragra.ti.
7	toph$.ab. or,ph$.ti.
8	1 or 2 or 3 or 4 or 5 or 6 or 7
9	allopurinol.ab. or allopurinol.ti.
10	xanthine oxidase inhibit$.ab. or xanthine oxidase inhibit$.ti.
11	"urate-lowering therap$".ab. or "urate-lowering therap$".ti.
12	urate lowering therap$.ab. or urate lowering therap$.ti.
13	"urate-lowering drug$".ab. or "urate-lowering drug$".ti.
14	urate lowering drug$.ab. or urate lowering drug$.ti.
15	"uric acid-lowering therap$".ab. or "uric acid-lowering therap$".ti.
16	uric acid lowering therap$.ab. or uric acid lowering therap$.ti.
17	"uric acid-lowering drug$".ab. or "uric acid-lowering drug$".ti.
18	uric acid lowering drug$.ab. or uric acid lowering drug$.ti.
19	uricostatic.ab. or uricostatic.ti.
20	(AL or Abburic or Abopur or Acepurin or Acifugan or Acyprin or Adenock or Ailural or Ailurial or Alfadiman or Algut or Aligout or Alinol or Allgoric or Allnol or Allo or Allo$ or Allpargin or Allupol or Allura$ or Alluri$ or Aloprim or Alopur or Alopurinol or Aloral or Aloric or Aloriv or Alosfar or Alositol or Alpur$ or Aluline or Aluprol or Alur or Alurid or Aluron or Alzoprim or Anoprolin or Anurate or Anzief or Apnol or Apo-Allopurinol or Apo-Tinole or Apronol or Apulonga or Apurin or Apurol or Arnol or Arsol or Artrex or Arturic or Atisuril or Aurigen).ab.
21	(AL or Abburic or Abopur or Acepurin or Acifugan or Acyprin or Adenock or Ailural or Ailurial or Alfadiman or Algut or Aligout or Alinol or Allgoric or Allnol or Allo or Allo$ or Allpargin or Allupol or Allura$ or Alluri$ or Aloprim or Alopur or Alopurinol or Aloral or Aloric or Aloriv or Alosfar or Alositol or Alpur$ or Aluline or Aluprol or Alur or Alurid or Aluron or Alzoprim or Anoprolin or Anurate or Anzief or Apnol or Apo-Allopurinol or Apo-Tinole or Apronol or Apulonga or Apurin or Apurol or Arnol or Arsol or Artrex or Arturic or Atisuril or Aurigen).ti.

22	(Benoxuric or Be-Uric or Bionol or Biuricowas or Bleminol or Burin or Caplenal or Capurate or Cellidrin or Cellidrin$ or Chinnol or Ciploric or Colpuril or Comburic or Cosuric or Dabrosin or Dabroson or Darzune or Dertrifort or Desatura or Docallopu or Duovitan or Dura or Edorin or Elavil or Embarin or Epidropal or Epuric or Erloric or Ethipurinol or Etindrax or Facilit or Foligan or Gealgica or Geapur or Gewapurol or Gichtex or Gotax or Gotir or Gurik or Hamarin or Harpagin or Hexanurat or Hexanuret or Hycemia or Isoric or Jenapurinol or Ketanrift or Ketobun-A or Labopurinol or Labypurol or Lanolone or Ledopur or Licoric or Linogra or Llanol or Lodiric or Logout-SR or Lonol or Loporic or Lopric or Lopur$ or Loricid or Lo-Uric or Lysuron or Marinol or Medoric or Mephanol or Milurit or Miniplanor).ab.
23	(Benoxuric or Be-Uric or Bionol or Biuricowas or Bleminol or Burin or Caplenal or Capurate or Cellidrin or Cellidrin$ or Chinnol or Ciploric or Colpuril or Comburic or Cosuric or Dabrosin or Dabroson or Darzune or Dertrifort or Desatura or Docallopu or Duovitan or Dura or Edorin or Elavil or Embarin or Epidropal or Epuric or Erloric or Ethipurinol or Etindrax or Facilit or Foligan or Gealgica or Geapur or Gewapurol or Gichtex or Gotax or Gotir or Gurik or Hamarin or Harpagin or Hexanurat or Hexanuret or Hycemia or Isoric or Jenapurinol or Ketanrift or Ketobun-A or Labopurinol or Labypurol or Lanolone or Ledopur or Licoric or Linogra or Llanol or Lodiric or Logout-SR or Lonol or Loporic or Lopric or Lopur$ or Loricid or Lo-Uric or Lysuron or Marinol or Medoric or Mephanol or Milurit or Miniplanor).ti.
24	(Nektrohan or Nilapur or Novo-Purol or Oloprim or Orlu or Petrazyc or Piloric or Ponuric or Prinol or Pritanol or Progout or Pureduct or Puricemia or Puricin or Puricos or Puride or Purigan or Purinase or Purinol or Purispec or Puristen or Puritenk or Pyrazol or Ranpuric or Redurate or Remid or Reucid or Riball or Rimapurinol or Rinolic or Riva-Purinol or Satric or Sigapurol or Sinoric or Soluric or Stradumel or Suspendol or Swiloric or Synol or Synpurinol or Takanarumin or Talol or Tipuric or Trianol or Tylonic or Unizuric or Urbol or Uredimin or Uribenz or Urica or Urica$ or Uricemil or Uricina or Uricnol or Urico$ or Urifugan or Urikoliz or Urinol or Uriprim or Uripurinol or Uritab or Uritas or Urlo or Urobenyl or Urogotan or Urolit or Uroplus or Urosi or Urosin or Urozyl-SR or Urtias or Valeric or Xandase or Xanol or Xanthomax or Xanturat or Xanturic or Xanurace or Xuric-A or Z 300 or Zilopur or Z-Nol or Zurim or Zygout or Zylapour or Zylic or Zylo$ or Zyprinol or Zytol).ab.
25	(Nektrohan or Nilapur or Novo-Purol or Oloprim or Orlu or Petrazyc or Piloric or Ponuric or Prinol or Pritanol or Progout or Pureduct or Puricemia or Puricin or Puricos or Puride or Purigan or Purinase or Purinol or Purispec or Puristen or Puritenk or Pyrazol or Ranpuric or Redurate or Remid or Reucid or Riball or Rimapurinol or Rinolic or Riva-Purinol or Satric or Sigapurol or Sinoric or Soluric or Stradumel or Suspendol or Swiloric or Synol or Synpurinol or Takanarumin or Talol or Tipuric or Trianol or Tylonic or Unizuric or Urbol or Uredimin or Uribenz or Urica or Urica$ or Uricemil or Uricina or Uricnol or Urico$ or Urifugan or Urikoliz or Urinol or Uriprim or Uripurinol or Uritab or Uritas or Urlo or Urobenyl or Urogotan or Urolit or Uroplus or Urosi or Urosin or Urozyl-SR or Urtias or Valeric or Xandase or Xanol or Xanthomax or Xanturat or Xanturic or Xanurace or Xuric-A or Z 300 or Zilopur or Z-Nol or Zurim or Zygout or Zylapour or Zylic or Zylo$ or Zyprinol or Zytol).ti.
26	exp Allopurinol/
27	9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
28	8 and 27

Appendix C Literature review table

Table C1: Summary of study design characteristics and outcomes considered

Article, [Country]	Study design	Matching variables	Setting & data source	Length of follow-up	Sample size for analysis	Outcome
Alvarez-Nemegyei et al. (2005), [Mexico]	Cohort nested case-control cohort study	Not stated	Recruited from a hospital rheumatology clinic. Data collected via clinical assessment, patient interviews or the cohort database	Followed up every six months	90	Renal failure; MSK physical disability
Azevedo et al. (2014), [Brazil]	Cohort study	N/A	Medical record review from a rheumatology clinic	Followed up every three months for 18 months	48	SU level
Cheyoe et al. (2012), [Thailand]	Cross-sectional study	N/A	Medical record review of a hospital	N/A	154	Chronic kidney disease
Choe et al. (2010), [South Korea]	Cross-sectional study	N/A	Recruited from a hospital rheumatology clinic. Data collected via clinical assessment	N/A	68	Serum cystatin C concentration
Crittenden et al. (2012), [USA]	Cross-sectional study	N/A	Medical record review of the New York Harbor Healthcare System from 3 hospitals	N/A	1,288	Myocardial infarction
Dalbeth et al. (2006), [New Zealand]	Cross-sectional study	N/A	Medical record review from rheumatology clinics	N/A	214 - 227	Presence of tophi; SU level; SU level ≤360µmol/L

Dalbeth et al. (2012), [New Zealand]	Cross-sectional study	N/A	Recruited from community advertising and primary and secondary care clinics. Data collected via clinical assessment	N/A	177 - 273	SU level <360µmol/L
Dalbeth et al. (2013), [New Zealand]	Cross-sectional study	N/A	Recruited from community advertising and primary and secondary care clinics. Data collected via clinical assessment	N/A	290	Presence of tophi
Dubreuil et al. (2015), [UK]	PS matched, incident user cohort study	PS conditional on BMI; age; sex; hypertension; cardiovascular disease; diabetes; Charlson comorbidity index; statins; fibrates; ACE inhibitors; ARBs; β blockers; calcium channel blockers; aspirin; NSAIDS; loop diuretics; hydrochlorothiazide; losartan; insulin; SU level; cholesterol; albumin; GFR; no. of primary care visits	EHR review from THIN	Mean 2.9 years	9,590	Time to all-cause mortality
Emmerson et al. (1987), [Australia]	Cross-sectional study	N/A	Recruited from a hospital. Data collected via clinical assessment	N/A	66	Plasma oxypurinol
Fessel (1979), [USA]	Cohort study	N/A	Medical record review from clinics and hospitals	Mean 127 ± 68 months	168	Stone formation

Graham et al. (2013), [Australia]	Cohort study	N/A	Recruited from hospital and rheumatology clinics. Data collected via clinical assessment	Not stated	46	SU level
Halpern et al. (2009), [USA]	Cohort study	N/A	EHR review of an administrative claims database	1 year	3,070	Gout flare
Hatoum et al. (2014), [USA]	Cohort study	N/A	EHR review of the General Electric Electronic Medical Record database	6 and 24 months	7,324 – 10,871	SU level <6mg/dL (<360µmol/L)
Hutton et al. (2009), [New Zealand]	Case-control study	Age; sex; ethnicity	Medical record review from a hospital	5 years	67 - 96	Two or more unplanned hospital admissions for gout management
Kim et al. (2013b), [USA]	Cohort study	N/A	EHR review of the Innovus InVision Data Mart database	Range 30 days to 3 years. Allopurinol cohort, mean follow up 0.5 years. Febuxostat cohort, mean follow up 0.4 years. Colchicine cohort, mean follow up 0.2 years.	35,577	Frequency of gout flares
Kok et al. (2014), [Taiwan]	Matched cohort study	Age; sex; index date; diabetes; hypertension; hyperlipidaemia; atrial fibrillation	EHR review of the Taiwan National Health Insurance Research Database	Maximum 10-year follow-up. Allopurinol cohort, median (IQR) 5.25 years (2.81, 7.69). Non-allopurinol cohort, median (IQR) 5.04 years (2.55, 7.53)	2,483 – 4,966	Cardiovascular outcome requiring hospitalisation; coronary heart disease; cerebrovascular disease (stroke); hypertensive heart disease; heart failure; other cardiovascular disorders

Lim et al. (2012), [Singapore]	Cohort study	N/A	Recruited from a hospital rheumatology clinic. Data collected via clinical assessment	Median (range): 39.8 weeks (3.9, 96.6)	126	Time to reach SU level <360µmol/L; attained SU level <360µmol/L
Mak et al. (2009), [Singapore]	Cohort study	N/A	Recruited from a hospital. Data collected via clinical assessment and medical record review	15 months	100	Frequency of gout flares
Marchini et al. (2013), [USA]	Cross-sectional study	N/A	Medical record review from a hospital	N/A	278	Stone composition (mixed vs. pure) Percentage composition of CaOMH, CaODH, CaPh, uric acid, struvite, cystine. Number of pure stones composed of uric acid, CaOMH, CaPh, cystine.
Meek et al. (2014), [Netherlands]	Cohort study	N/A	Medical record review of the Arthritis Centre Twente CardioVascular Disease database, GP questionnaires, and the Dutch national death registry	Median (IQR) 36 months (30, 41)	172	Cardiovascular event or death
Neogi et al. (2014), [USA]	Case-crossover study	Within patient matching of hazard and control periods	Recruited from community advertisement. Data collected via online self-report questionnaires and medical record review	1 year	724	Gout flare
Pandya et al. (2011), [USA]	Cohort study	N/A	EHR review of a medical claims database	At least one year follow-up. Mean (SD) 32 months (17.1)	1,768	SU level <6mg/dL

Perez-Ruiz et al. (1998), [Spain]	Cohort study	N/A	Recruited from a hospital rheumatology clinic. Data collected via clinical assessment	Mean 12.5 months	86	SU level; percentage reduction of SU level; clearance of creatinine; clearance of urate; urinary excretion of urate; SU level >6mg/dL
Perez-Ruiz et al. (2002), [Spain]	Cohort study	N/A	Recruited from a hospital gout clinic. Data collected via clinical assessment	Not stated	49	SU level; diameter of target tophus; time until tophi resolution; velocity of tophi reduction
Perez-Ruiz et al. (2010), [Spain]	Cohort study	N/A	Recruited from a hospital. Data collected via clinical assessment	At least 12 months	546	Urinary volume; SU level; urinary pH; urinary uric acid level; undissociated urinary uric acid level; 24-hr urine uric acid level; 24-hr urine dissociated uric acid level; clearance of creatinine; clearance of uric acid
Roddy et al. (2007b), [UK]	Cross-sectional study	N/A	Patients recruited from primary care practices. Data collected via self-report questionnaire and clinical assessment	N/A	145	SU level; SU level >360µmol/L; gout flare in preceding year
Rothenbacher et al. (2011), [UK]	Cohort study	N/A	EHR review from THIN	Mean (range) 3.8 years (30 days, 8 years)	23,857	Time to first gout flare; frequency of gout flares
Stamp et al. (2000), [New Zealand]	Cross-sectional study	N/A	Medical record review from a hospital rheumatology clinic	N/A	31	Plasma oxypurinol
Stamp et al. (2011a), [New Zealand]	Cohort study	N/A	Recruited from a medical centre. Data collected via clinical assessment	12 months	45	Plasma oxypurinol concentration

Stamp et al. (2011b), [New Zealand]	Cohort study	N/A	Recruited from a hospital rheumatology clinic. Data collected via clinical assessment	12 months	35	Percentage reduction in SU level; SU level; SU level <360µmol/L
Stamp et al. (2012), [New Zealand]	Case-control study	Age; sex; diuretics; renal function	EHR review of local databases, physician recall, Centre for adverse Reactions Monitoring	Median (Range) 30 days (1, 1,080)	211	Allopurinol hypersensitivity syndrome; SU level
Stamp et al. (2013), [New Zealand]	Cross-sectional study	N/A	Patients recruited from the general population. Data collected via questionnaires, GP records, and clinical assessment	N/A	57	SU level
Thanassoulis et al. (2010), [Canada]	Nested case control cohort study	Calendar day of admission to cohort	EHR review of the Quebec universal health insurance program	Median (range) 2.1 years (30 days to 7 years)	7,684	Heart failure re-admission or all-cause mortality
Vazquez-Mellado et al. (2001), [Mexico]	Cohort study	N/A	Medical record review from a hospital rheumatology clinic.	Not stated	120	Adverse event including rash, allopurinol hypersensitivity syndrome, fixed pigmented drug eruption, leucocytoclastic vasculitis; SU level
Zhang et al. (2012), [Australia]	Case-crossover study	Within person matching of hazard and control periods	Recruited from community advertisement. Data collected via online self-report questionnaires and medical record review	1 year	633	Gout flare

ACE: Angiotensin-converting enzyme; CaOMH: Calcium oxalate monohydrate; CaODH: Calcium oxalate dihydrate; CaPh: Calcium phosphate; EHR: Electronic health record; GFR: glomerular filtration rate; IQR: Interquartile range; MSK: Musculoskeletal; NSAIDS: Non-steroidal anti-inflammatory drugs; PS: Propensity score; SD: Standard deviation; SU: Serum urate; THIN: The Health Improvement Network; USA: United States of America

Table C2: Summary of how allopurinol use was defined

Article	Treatment comparison group vs. allopurinol group	Was dosage of allopurinol taken into account	How was dosage taken into account	Was duration of allopurinol taken into account	How was duration taken into account
Alvarez-Nemegyei et al. (2005)	N/A	Yes	<300 vs. >300mg/day	No	N/A
Azevedo et al. (2014)	Benzbromarone vs. allopurinol	No	N/A	No	N/A
Cheyoe et al. (2012)	Colchicine vs. allopurinol & colchicine	No	N/A	No	N/A
Choe et al. (2010)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Crittenden et al. (2012)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Dalbeth et al. (2006)	Non-allopurinol vs. allopurinol	Yes	Non-allopurinol use vs. lower than recommended dose vs. recommended dose vs. higher than recommended dose	No	N/A
Dalbeth et al. (2012)	Non-allopurinol or probenecid vs. allopurinol	Yes	Mean 235 vs. 194mg/day	No	N/A
Dalbeth et al. (2013)	Non-allopurinol use vs. allopurinol use	Yes	Mean 215 vs. 213mg/day	No	N/A
Dubreuil et al. (2015)	Non-allopurinol use vs. allopurinol use	No	N/A	Yes	Sensitivity analyses truncating follow-up at 1, 2 and 3 years to address treatment discontinuation
Emmerson et al. (1987)	N/A	Yes	100 vs. 200 vs. 300 vs. 400mg/day	No	N/A
Fessel (1979)	Probenecid vs. allopurinol	No	N/A	No	N/A
Graham et al. (2013)	N/A	Yes	Adjusted for dose that ranged from 50mg to 600mg/day	No	N/A
Halpern et al. (2009)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Hatoum et al. (2014)	Febuxostat vs. allopurinol	No	N/A	No	N/A
Hutton et al. (2009)	Non-allopurinol vs. allopurinol	Yes	Distribution of allopurinol dose between cases and controls	No	N/A

Kim et al. (2013b)	Colchicine vs. allopurinol	No	No	Yes	Analysis stratified by no. of days taking allopurinol (<30, 31-90, 91-120 days)
Kok et al. (2014)	Non-allopurinol vs. allopurinol Uricosurics vs. allopurinol	Yes	<100 vs 100 vs. 200 vs. ≥300mg/day	No	N/A
Lim et al. (2012)	N/A	Yes	Adjusted for incremental dose increase (per 50mg); compared distribution of doses between those attaining and not attaining target SU level	No	N/A
Mak et al. (2009)	N/A	No	N/A	Yes	Adjusted for duration of allopurinol
Marchini et al. (2013)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Meek et al. (2014)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Neogi et al. (2014)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Pandya et al. (2011)	N/A	Yes	Adjusted for average dose of last allopurinol prescription, incremental increases of 50mg/day	No	N/A
Perez-Ruiz et al. (1998)	Benzbromarone vs. allopurinol	Yes	Analysis stratified by 300mg users and 300mg and 450mg users	No	N/A
Perez-Ruiz et al. (2002)	Benzbromarone vs. allopurinol	No	N/A	No	N/A
Perez-Ruiz et al. (2010)	Benzbromarone vs. allopurinol	No	N/A	No	N/A
Roddy et al. (2007b)	Non-allopurinol vs. allopurinol	Yes	100 vs. 200 vs. 300 vs. >300mg/day	No	N/A
Rothenbacher et al. (2011)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Stamp et al. (2000)	N/A	Yes	Recommended dose vs. lower than recommended dose vs. higher than recommended dose	No	N/A

Stamp et al. (2011a)	N/A	Yes	Adjusted for dose of incremental increase of 100mg/day	No	N/A
Stamp et al. (2011b)	Furosemide vs. allopurinol	Yes	Incremental increase of 50mg/day from recommended dose	No	N/A
Stamp et al. (2012)	N/A	Yes	Mean 183.5 vs. 112.2mg/day; Starting dose higher than recommended vs. same or lower dose than recommended based on creatinine clearance or estimated GFR	No	N/A
Stamp et al. (2013)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A
Thanassoulis et al. (2010)	Non-allopurinol vs. allopurinol	Yes	Adjusted for non-allopurinol vs. ≤100 vs. >100mg/day	Yes	Adjusted for non-allopurinol vs. ≤30 days vs. >30 days
Vazquez-Mellado et al. (2001)	N/A	Yes	Recommended dose vs. higher than recommended dose	No	N/A
Zhang et al. (2012)	Non-allopurinol vs. allopurinol	No	N/A	No	N/A

Table C3: Summary of how confounding was controlled for

Article	Unadjusted or adjusted analysis	List of covariates	Statistical methods	Statistical limitations related to confounding
Alvarez-Nemegyei et al. (2005)	Unadjusted	N/A	Chi square test with yates correction	None stated
Azevedo et al. (2014)	Unadjusted	N/A	T-test	No randomisation; confounding by indication; follow-up was not blind
Cheyoe et al. (2012)	Adjusted	BMI; hypertension; diabetes; dyslipidaemia; diuretics; SU level	Logistic regression	Unmeasured covariates
Choe et al. (2010)	Adjusted	Stage of renal function; age; HDL-cholesterol; SU level; benzbromarone; erythrocyte sediment reaction; C-reactive protein	Linear regression	None stated
Crittenden et al. (2012)	Unadjusted	N/A	Chi-square test	Confounding by indication
Dalbeth et al. (2006)	Unadjusted	N/A	Chi-square test; T-test	Patient compliance to allopurinol
Dalbeth et al. (2012)	Unadjusted & adjusted	Sex; ethnicity; confidence to keep SU under control	Chi-square test; T-test; logistic regression	Confounding by indication
Dalbeth et al. (2013)	Unadjusted	N/A	Chi-square test; T-test	None stated
Dubreuil et al. (2015)	Adjusted	BMI; age; sex; hypertension; cardiovascular disease; diabetes; Charlson comorbidity index; statins; fibrates; ACE inhibitors; ARBs; β blockers; calcium channel blockers; aspirin; NSAIDS; loop diuretics; hydrochlorothiazide; losartan; insulin; SU level; cholesterol; albumin; GFR; no. of primary care visits	Cox regression	Residual or unknown confounding
Emmerson et al. (1987)	Unadjusted	N/A	Linear regression	None stated
Fessel (1979)	Unadjusted	N/A	Chi-square test; T-test	None stated
Graham et al. (2013)	Adjusted	Creatinine clearance; SU level	Linear regression	None stated

Halpern et al. (2009)	Adjusted	SU level; age; sex; no. of gout related office visits; disorders of lipid metabolism; non-traumatic joint disorders; hypertension; diseases of the heart; diseases of the urinary system; diabetes without complications	Logistic regression	Missing data
Hatoum et al. (2014)	Adjusted	SU level; age; sex; Charlson comorbidity index; ethnicity; year of gout diagnosis; tophi diagnosis; SU level	Logistic regression	Missing data; compliance with treatment and dose adjustment
Hutton et al. (2009)	Unadjusted	N/A	Odds ratio calculated from the Mantel Haenszel method; Wilcoxon independent groups test	Adherence to allopurinol
Kim et al. (2013b)	Adjusted	Age; sex; comorbidity score; hypertension; chronic kidney disease; renal stones; heart failure; cardiovascular disease; diabetes; hyperlipidaemia; stroke; obesity; COPD; diuretics; β blockers; ACE inhibitors; ARBs; NSAIDS; coxibs; opioids; oral steroids; intravenous steroids; intra-articular steroids; no. of prescription drugs; no. of hospitalisations; no. of ER visits; no. office visits; no. of PCP visits; no. of rheumatology visits	Poisson regression	Confounding by indication; missing data
Kok et al. (2014)	Adjusted	Chronic kidney disease, uremia, gastric ulcer	Cox regression	Unmeasured covariates; Compliance and adherence to treatment
Lim et al. (2012)	Unadjusted	N/A	Cox regression; Mann-Whitney U test	No randomization; time dependency of the number of allopurinol titrations; duration of colchicine prophylaxis

Mak et al. (2009)	Adjusted	Age; sex; ethnicity; duration of gout (years); serum creatinine level; SU level; ischaemic heart disease; LDL cholesterol; HDL cholesterol; triglyceride; creatinine clearance; statin use	Linear regression	Unmeasured confounding
Marchini et al. (2013)	Unadjusted	N/A	T-test; Chi-square test; Fisher's test	None stated
Meek et al. (2014)	Unadjusted	N/A	Chi-square test	Unmeasured confounding
Neogi et al. (2014)	Adjusted	Purine intake, diuretics, colchicine, NSAIDS, no. of alcohol servings	Conditional logistic regression	Residual confounding
Pandya et al. (2011)	Adjusted	Age; sex; stage of chronic kidney disease; specialty of prescribing physician	Logistic regression	None stated
Perez-Ruiz et al. (1998)	Unadjusted	N/A	T-test; Chi-square test	None stated
Perez-Ruiz et al. (2002)	Unadjusted	N/A	T-test	None stated
Perez-Ruiz et al. (2010)	Unadjusted	N/A	T-test	None stated
Roddy et al. (2007b)	Unadjusted	N/A	T-test; Chi-square test; ANOVA	None stated
Rothenbacher et al. (2011)	Adjusted & unadjusted	Sex; age; no. of GP visits; smoking status; alcohol consumption; BMI; ischaemic heart disease; hypertension; hyperlipidaemia; diabetes; renal failure	Cox regression; likelihood ratio test	Unmeasured covariates
Stamp et al. (2000)	Unadjusted	N/A	Fisher's exact test	Compliance with allopurinol; unmeasured covariates
Stamp et al. (2011a)	Unadjusted	N/A	Mixed-effect linear model	None stated
Stamp et al. (2011b)	Unadjusted	N/A	T-test, chi-square test	None stated
Stamp et al. (2012)	Unadjusted and adjusted	Ethnicity; presence of tophi	ANOVA; conditional logistic regression	Unmeasured covariates
Stamp et al. (2013)	Unadjusted	N/A	T-test	None stated

Thanassoulis et al. (2010)	Adjusted	Age; sex; Deyo modified Charlson comorbidity score; hypertension; myocardial infarction; renal failure; cardiac procedures; ACE inhibitors; β blockers; antiplatelet agents; anticoagulants; aldosterone antagonists; diuretics; NSAIDS; corticosteroids	Conditional logistic regression	Unmeasured covariates; confounding by indication; compliance
Vazquez-Mellado et al. (2001)	Unadjusted	N/A	Chi-square test or Fisher's exact test; T-test	None stated
Zhang et al. (2012)	Adjusted	Purine intake; alcohol use; diuretics; colchicine; NSAIDS; cherry intake	Conditional logistic regression	Unmeasured covariates

ANOVA: Analysis of variance; ARB: Angiotensin II receptor blockers; BMI: Body mass index; ER: Emergency room; HDL: High density lipoproteins; LDL: Low density lipoproteins; PCP: Primary care provider; SU: Serum urate

Appendix D ISAC application form

ISAC APPLICATION FORM
PROTOCOLS FOR RESEARCH USING THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

ISAC use only: Protocol Number Date submitted	....	**IMPORTANT** **If you have any queries, please contact ISAC Secretariat:** ISAC@cprd.com

1. Study Title: Modelling effectiveness of treatment in gout using primary care databases

2. Principal Investigator (full name, job title, organisation & e-mail address for correspondence regarding this protocol)
Trishna Rathod; Research assistant in biostatistics & PhD student**;** Research Institute for Primary Care & Health Sciences, Keele University; t.rathod@keele.ac.uk

3. Affiliation (full address)
Research Institute for Primary Care & Health Sciences
Keele University
Keele
Staffordshire
ST5 5BG

4. Protocol's Author (if different from the principal investigator)
N/A

5. List of all investigators/collaborators (*please list the names, affiliations and e-mail addresses* of all collaborators*, other than the principal investigator)
Milisa Blagojevic-Bucknall; Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK; m.bucknall@keele.ac.uk
Edward Roddy; Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK; e.roddy@keele.ac.uk
Kelvin Jordan; Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK; k.p.jordan@keele.ac.uk
Christian Mallen; Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK; c.d.mallen@keele.ac.uk

Please note that your ISAC application form and protocol **must be copied to all e-mail addresses listed above at the time of submission of your application to the ISAC mailbox. Failure to do so will result in delays in the processing of your application.*

6. Type of Institution (please tick one box below)

Academia	☒	Research Service Provider	☐	Pharmaceutical Industry	☐
NHS	☐	Government Departments	☐	Others	☐

7. Financial Sponsor of study

Pharmaceutical Industry *(please specify)*	☐	Academia *(Keele University)*	☒
Government / NHS *(please specify)*	☐	None	☐
Other *(please specify)*	☐		

8. Data source *(please tick one box below)*

Sponsor has on-line access	☒	Purchase of ad hoc dataset	☐
Commissioned study	☐		
Other	☐ *(please specify)*		

9. Has this protocol been peer reviewed by another Committee?

Yes*	☒	No	☐

** Please state in your protocol the name of the reviewing Committee(s) and provide an outline of the review process and outcome.*
The protocol has been reviewed internally by the Keele CPRD Steering Group.

10. Type of Study *(please tick all the relevant boxes which apply)*

Adverse Drug Reaction/Drug Safety	☐	Drug Use	☐	Disease Epidemiology	☒
Drug Effectiveness	☒	Pharmacoeconomic	☐	Other	☐

11. This study is intended for:

Publication in peer reviewed journals	☒	Presentation at scientific conference	☒
Presentation at company/institutional meetings	☒	Other PhD	

12. Does this protocol also seek access to data held under the CPRD Data Linkage Scheme?

Yes	☒	No	☐

13. If you are seeking access to data held under the CPRD Data Linkage Scheme*, please select the source(s) of linked data being requested.

☒ Hospital Episode Statistics	☐ Cancer Registry Data**
☐ MINAP	☒ ONS Mortality Data
☒ Index of Multiple Deprivation/ Townsend Score	
☐ Mother Baby Link	☐ Other: (please specify)

** As part of the ISAC review of linkages, the protocol may be shared - in confidence - with a representative of the requested linked data set(s) and summary details may be shared - in confidence - with the Confidentiality Advisory Group of the Health Research Authority.*

***Please note that applicants seeking access to cancer registry data must provide consent for publication of their study title and study institution on the UK Cancer Registry website. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email kc@cprd.com to discuss this requirement further.*

14. If you are seeking access to data held under the CPRD Data Linkage Scheme, have you already discussed your request with a member of the Research team?

Yes ☒ No* ☐

**Please contact the CPRD Research Team on +44 (20) 3080 6383 or email kc@cprd.com to discuss your requirements before submitting your application.*

Please list below the name of the person/s at the CPRD with whom you have discussed your request: Kendall Chidwick (15th April 2014)

15. If you are seeking access to data held under the CPRD Data Linkage Scheme, please provide the following information:

The number of linked datasets requested: 3

A synopsis of the purpose(s) for which the linkages are required:

Deprivation is a risk factor for increased consultation, morbidity and all-cause mortality which would have an impact on resource allocation to general practices. In England, gout prevalence increases with worsening occupational socio-economic status and gout is associated with inadequate income. The long term effects of allopurinol vs. alternative treatments are unknown and any difference arising between treatment groups could potentially partly be explained by deprivation.

Gout has a strong association with the metabolic syndrome in particular hypertension, hyperlipidaemia, insulin resistance and obesity. This increases the risk of diabetes, vascular events and renal diseases which may result in hospitalization. Admissions to hospital would indicate whether gout patients using allopurinol are hospitalised more often than those on alternative treatments. Integrated hospital episode statistics data is requested only.

Knowing the exact cause of death would identify gout patients whose cause of death is gout comorbidity related and may estimate all-cause mortality or cause-specific mortality.

Is linkage to a local dataset with <1 million patients being requested?

Yes* ☐ No ☒

** If yes, please provide further details:*

16. If you have requested linked data sets, please indicate whether the Principal Investigator or any of the collaborators listed in response to question 5 above, have access to any of the linked datasets in a patient identifiable form, or associated with a patient index.

Yes* ☐ No ☒

** If yes, please provide further details:*

17. Does this protocol involve requesting any additional information from GPs?

Yes* ☐ No ☒

** Please indicate what will be required:*

Completion of questionnaires by the GP[v] Yes ☐ No ☐

Provision of anonymised records (e.g. hospital discharge summaries) Yes ☐ No ☐

Other (please describe)

[v] Any questionnaire for completion by GPs or other health care professional must be approved by ISAC before circulation for completion.

18. Does this protocol describe a purely observational study using CPRD data (this may include the review of anonymised free text)?

Yes* ☒ No** ☐

** Yes: If you will be using data obtained from the CPRD Group, this study does not require separate ethics approval from an NHS Research Ethics Committee.*

*** No: You may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.*

19. Does this study involve linking to patient *identifiable* data from other sources?

Yes ☐ No ☒

20. Does this study require contact with patients in order for them to complete a questionnaire?

Yes ☐ No ☒

N.B. Any questionnaire for completion by patients must be approved by ISAC before circulation for completion.

21. Does this study require contact with patients in order to collect a sample?

Yes* ☐ No ☒

** Please state what will be collected*

22. Experience/expertise available

Please complete the following questions to indicate the experience/expertise available within the team of researchers actively involved in the proposed research, including analysis of data and interpretation of results

	Previous GPRD/CPRD Studies	Publications using GPRD/CPRD data
None	☐	☐
1-3	☐	☐
> 3	☒	☒

	Yes	No
Is statistical expertise available within the research team?	☒	☐

If yes, please outline level of experience

TR is a research assistant in biostatistics and holds an MSc in statistics; MB is a lecturer in statistics and holds a PhD in statistics; KJ is a reader in biostatistics and holds a PhD in statistics.

Is experience of handling large data sets (>1 million records) available within the research team?	☒	☐

If yes, please outline level of experience

KJ and MB have previously worked with large GPRD datasets and in addition have used local general practice database (CiPCA).

Is UK primary care experience available within the research team? ☒ ☐ *If yes, please outline level of experience* CM is a practising GP and Professor of General Practice Research.	
23. References relating to your study Please list up to 3 references (most relevant) relating to your proposed study. 1. Kuo CF, Grainge MJ, Mallen C, Zhang W, Doherty M. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann Rheum Dis* 2014 2. Wei L, Mackenzie IS, Chen Y,Struthers AD, MacDonald TM. Impact of allopurinol use on urate concentration and cardiovascular outcome. *Br J Clin Pharmacol* 2011; 71(4):600-607 3. Roddy E, Zhang W, Doherty M. Concordance of the management of chronic gout in a primary-care population with the EULAR gout recommendations. *Ann Rheum Dis* 2007; 66:1311-1315	

PROTOCOL CONTENT CHECKLIST

In order to help ensure that protocols submitted for review contain adequate information for protocol evaluation, ISAC have produced instructions on the content of protocols for research using CPRD data. These instructions are available on the CPRD website (www.cprd.com/ISAC). All protocols using CPRD data which are submitted for review by ISAC must contain information on the areas detailed in the instructions. IF you do not feel that a specific area required by ISAC is relevant for your protocol, you will need to justify this decision to ISAC.

Applicants must complete the checklist below to confirm that the protocol being submitted includes all the areas required by ISAC, or to provide justification where a required area is not considered to be relevant for a specific protocol. Protocols will not be circulated to ISAC for review until the checklist has been completed by the applicant.

Please note, your protocol will be returned to you if you do not complete this checklist, or if you answer 'no' and fail to include justification for the omission of any required area.

	Included in protocol?		
Required area	Yes	No	If no, reason for omission
Lay Summary (max.200 words)	☒	☐	
Background	☒	☐	
Objective, specific aims and rationale	☒	☐	
Study Type *Descriptive* *Hypothesis Generating* *Hypothesis Testing*	 ☒ ☐ ☒	 ☐ ☒ ☐	This will be a hypothesis testing study with the null hypothesis that there is no different in outcomes between gout patients prescribed and not prescribed allopurinol. However, the study also includes a strong element of methodological research to answer the clinical objectives.
Study Design	☒	☐	
Sample size/power calculation (Please provide justification of sample size in the protocol)	☒	☐	
Study population (including estimate of expected number of relevant patients in the CPRD)	☒	☐	
Selection of comparison group(s) or controls	☒	☐	
Exposures, outcomes and covariates *Exposures are clearly described* *Outcomes are clearly described*	☒ ☒ ☒	☐ ☐ ☐	
Use of linked data (if applicable)	☒	☐	
Data/ Statistical Analysis Plan *There is plan for addressing confounding* *There is a plan for addressing missing data*	☒ ☒ ☒	☐ ☐ ☐	
Patient/ user group involvement †	☐	☒	No PPI was needed in the planning and interpretation of the results for this PhD as it will be mostly based on methodological research.
Limitations of the study design, data sources and analytic methods	☒	☐	
Plans for disseminating and communicating study results	☒	☐	

† ***It is expected that many studies will benefit from the involvement of patient or user groups in their planning and refinement, and/or in the interpretation of the results and plans for further work. This is particularly, but not exclusively true of studies with interests in the impact on quality of life. Please indicate whether or not you intend to engage patients in any of the ways mentioned above.***

Voluntary registration of ISAC approved studies:

Epidemiological studies are increasingly being included in registries of research around the world, including those primarily set up for clinical trials. To increase awareness amongst researchers of ongoing research, ISAC encourages voluntary registration of epidemiological research conducted using MHRA databases. This will not replace information on ISAC approved protocols that may be published in its summary minutes or annual report. It is for the applicant to determine the most appropriate registry for their study. Please inform the ISAC secretariat that you have registered a protocol and provide the location.

Protocol

Lay Summary (Max 200 words)

Gout is the most prevalent inflammatory arthropathy and is largely managed in primary care in the UK. The principal risk factor for gout development is elevated serum uric acid levels. Long term treatment of gout involves using urate-lowering therapies such as allopurinol, however there is lack of high quality evidence on its effectiveness. We aim to use routinely collected data to examine long term effects of allopurinol on different outcomes, including uric acid levels. However using such data to determine treatment effect is problematic as there may be differences between treatment groups, e.g. certain treatments may be preferred for older patients. The treatment effect may then be due to differences between treatment groups, inducing confounding and bias study conclusions. Propensity score (PS) is a statistical approach that we will use to address this. The propensity (likelihood) of each patient to receive allopurinol given their measured characteristics is determined, so that the distribution of characteristics for patients with similar PS should be the same, enabling valid comparison of treatments. Optimal specification and subsequent adjustment for PS play a key role and we aim to explore these aspects of PS approach, and address the potential impact on our findings of patient characteristics not measured in CPRD.

Background

Gout is the most prevalent inflammatory arthropathy, affecting approximately 2.5% of adults in the UK in 2012; this prevalence increases to over 14% among men aged over 75 years[1]. Most gout patients are treated within the primary care setting, with an average general practice having 40 patients consulting for gout each year[2]. The principal risk factor for gout development is increased uric acid levels and the definitive long-term treatment of gout involves using urate-lowering therapies such as allopurinol. However, only a third of gout sufferers are prescribed allopurinol whilst others may be prescribed NSAIDS, analgesics or colchicine. A fifth of allopurinol users fail to lower their uric acid level below the recommended target of 360μmol/L[3].

Allopurinol was developed for the treatment of gout over 50 years ago however, only recently have studies investigating its effect been undertaken. Most of these have been of experimental nature, with allopurinol used as the comparator in three randomized controlled trials (RCTs) of newer urate-lowering drugs such as febuxostat[4-6]. However these trials were limited to 52 weeks follow-up and employed strict inclusion and exclusion criteria which may limit their generalizability to the majority of patients with gout. Therefore an extensive observational study is needed to investigate the effect of allopurinol use (Vs. non-allopurinol use), over a longer period of time on range of outcomes. However, in studies drawing on observational data a major impediment to valid assessment of treatment effect is the lack of randomization that is inherent in RCTs. This may often result in significant differences between treatment groups, with respect to both measured and unmeasured covariates, thus inducing possible confounding which may impact on inferences and conclusions. For instance RCTs using allopurinol as the comparator group excluded gout patients with poor renal function[4-6] in whom gout treatment can be most challenging. A traditional statistical procedure employed to take account of measured covariates is multivariable regression. However the issue of covariate imbalance between treatment groups remains. Propensity score methodology, formalized by Rosenbaum and Rubin in 1983[7], is a possible alternative because it enables balance on measured covariates between those receiving and nor receiving allopurinol to be achieved thus removing some of the bias inherent in treatment allocation. The process involves collapsing information on observed covariates into a single value (the propensity score) which reflects the likelihood of a patient receiving allopurinol given these covariates. Once propensity scores have been adjusted for in the analysis of treatment effect, we can then be confident that any differences in outcome can be attributed to treatment, at least as far as accountability of observed covariates used in construction of propensity scores is concerned. Propensity score methodology has already been applied in studying effect of allopurinol. For example one study found that allopurinol users have a slightly reduced risk of mortality compared to non-allopurinol users[8] whilst in another study allopurinol users were found to have an increased risk of severe cutaneous adverse reactions[9]. Wei *et* al[10] reported that there was no increased risk in cardiovascular events for allopurinol users compared with non-users of urate lowering therapy and that high dose allopurinol users had a significantly lower risk of cardiovascular events and mortality than those on a lower dose. However the drawbacks of majority of such studies utilizing propensity score methodology has been the restriction to older age groups, consideration of only a few outcomes, not making allowances for recurrence of non-fatal outcomes such as gout attacks, renal and cardiovascular disease, assumption of time independence (i.e. they do not change) of certain covariates, and assumption of time independence of allopurinol use. For example a patient may be prescribed allopurinol for a couple of months then stop taking the medication until a couple of years later and so on. Furthermore, routinely collected observational data typically records only a limited set of covariates, and it is possible that some important covariates will remain unmeasured such as family history of gout, purine rich diet, and poor adherence to treatment. Propensity score analysis does not balance for such covariates and hence there is a possibility that some bias from lack of randomization will remain. This will introduce heterogeneity among patients, ultimately resulting in underestimation of data dispersion and over-optimistic results. Common practice is to ignore such heterogeneity.

This project aims to approach estimation of allopurinol effect in a comprehensive and thorough manner. From a clinical aspect this will be achieved by considering a wide range of relevant outcomes among gout patients, relaxing exclusion criteria, specifically in terms of age and possible presence of comorbid conditions such as renal disease, and stratifying analyses on severity of gout. From a statistical aspect, allowance will be made for time-dependent use of allopurinol, time dependence of covariates and recurrence of outcome events, thus giving rise to repeated measures data structure within which both propensity scores and subsequently treatment effect will be estimated. Furthermore, robustness of allopurinol effect estimates to omission of important covariates will be tested. True propensity score is unknown and we will consider making allowance for such uncertainty by modelling propensity score as a latent variable within Bayesian set-up[11].

Objectives, Specific Aims, and Rationale

The main research objective is to investigate the long term effectiveness of allopurinol among a group of gout patients aged 18 and over.

To address this, two specific clinical aims are:

6) Examine effect of allopurinol (Vs. not taking allopurinol) on a range of outcomes: uric acid levels; repeat consultations for gout; hospital admissions; NSAID, analgesic and colchicine usage; allopurinol related side effects of hypersensitivity syndrome, rash, liver function, bone marrow suppression; gout comorbidities (vascular and renal diseases); joint replacement; mortality.

7) Repeat objective 1 stratified on baseline levels of uric acid (</> 480 μmol/L)[4] and on relevant comorbidities (renal and vascular disease) to assess whether effect of allopurinol varies by severity of gout or comorbidity.

Pertaining to statistical methodology, the three specific aims which will be addressed within 1) and 2) above are:

8) On estimation and adjustment for propensity scores: Investigate effect of including different sets of patient characteristics (covariates) in propensity score estimation on treatment effect estimate using three different adjustment approaches (stratification, matching, and weighting). We will consider inclusion of covariates related to treatment alone, those related to outcome alone, those related to both and inclusion of all available information. Extensions will be made to repeated measures setup.

9) In order to address the impact of unobserved covariates, sensitivity of parameter estimates obtained in 1) to omission of covariates (with varying degree of strength of association with outcome and/or treatment) from the propensity score estimation stage will be examined.

10) Investigate modelling propensity score as a latent variable within a Bayesian model, thereby accounting for some of the uncertainty inherent in construction of propensity scores.

Study Type

This will be a hypothesis testing study with the null hypothesis that there is no different in outcomes between gout patients prescribed and not prescribed allopurinol. However, the study also includes a strong element of methodological research to answer the clinical objectives.

Study design

This will be a prospective cohort study. The exposure of interest is allopurinol use vs. non usage of allopurinol in those who consulted for gout between 1997 and 2002 (see below).

Study population

The study sample will consist of all patients aged 18 years and over who had an initial consultation (or Read code) for gout between 1997 and 2002. The list of Read codes used to identify gout consultations is in appendix 1. Gout patients will need to have been registered at their practice for at least two years prior to their initial gout consultation.

Patients under the age of 18 years or who have taken any urate lowering drugs (allopurinol, sulfinpyrazone, probenecid and benzbromarone) during the two years prior to the initial gout consultation between 1997 and 2002 will be excluded from the analysis. Patients who are prescribed other urate lowering drugs (febuxostat, sulfinpyrazone, probenecid and benzbromarone) prior to first prescription of allopurinol will be excluded from analysis. Patients prescribed allopurinol who are subsequently prescribed other urate lowering drugs, their follow up will be censored at date of prescription of other urate lowering drugs

Exposures, Outcomes, and Covariates

Exposure:

- Prescriptions for allopurinol, including dosage, measured from baseline till end of study or death.

Patients with an initial consultation for gout between 1997 and 2002 will be identified and will be classified by exposure of allopurinol use (allopurinol vs. non-allopurinol). However prescription for allopurinol does not coincide with the initial consultation for gout as it is often prescribed a few years later. Defining the follow-up period from the date of the initial consultation for gout would be inappropriate as patients' exposure status would be determined during follow-up. During this time lag between gout consultation and allopurinol prescription, outcomes of interest may have occurred thus introducing bias, therefore a landmark method will be used.

In the landmark method, a fixed time-point (landmark date) would be selected to define the start of the follow-up period after the initial consultation for gout. Up to the landmark date, patients would be classified as either allopurinol or non-allopurinol users. Patients would be excluded from the sample if they have the outcome of interest prior to the landmark date. Three landmark time-points are selected at one, two and three years.

Primary outcome

- Occurrence of and time to uric acid levels <360 µmol/L (from blood tests). All records of uric acid level measurements from baseline till end of study/death will be needed.

Secondary outcomes (All records of these outcomes from baseline till end of study/death will be needed)

- Occurrence of and time to (and between) repeat gout consultations (appendix 1).
- Gout and non-gout related hospital admissions and related time to event.
- Occurrence of and time to use of NSAIDs, analgesics and colchicine usage and allopurinol related side effects of hypersensitivity syndrome, rash, liver function, bone marrow suppression.
- Occurrence of and time to vascular and renal diseases, and joint replacement (appendix 1).
- Occurrence of and time to death and cause of death.

Covariates (measured from baseline till death or end of study)

Socio-demographic and lifestyle

- Gender
- Year of birth
- Date and cause of death
- General practice
- Index of multiple deprivation
- Alcohol consumption
- Smoking status

General health

- Body Mass Index

Comorbidity

- Depression
- Anxiety

- Hypertension
- Hyperlipidaemia
- Diabetes mellitus
- Osteoarthritis
- Diuretic use
- Vascular disease
- Renal disease

Covariates will be collected 2 years before gout consultation and 10 years after or until death or no longer contribute to CPRD. Possible time dependent nature of allopurinol use will be explored. In analysis of outcomes where their recurrence is not of interest or not possible (for example joint replacement), covariate information after the date of such an event will be ignored.

Sample size/power calculation

Restriction is made to population of patients registered at general practices that have consented to HES, deprivation and ONS mortality linkage. It is assumed that the total CPRD annual registered population is 5.5 million patients and that approximately 50% of practices contributing to CPRD have consented to linkage. This yields relevant annual population of approximately 2.75 million. It is estimated that 70,000 of these will have gout, based on the latest figure for prevalence of gout as 2.5%.

A feasibility count was performed using CPRD. Between 1997 and 2002, 33,538 individuals aged 18 years and over had a Read code for gout and were registered with their practice for at least two years prior. The table below shows the proportion of patients who were prescribed allopurinol up to each landmark date.

Landmark date	Number (%) prescribed allopurinol
1 year	10,266 (30.61)
2 years	11,648 (34.74)
3 years	12,671 (37.78)

Adequate power is required to yield reasonable estimates of treatment effect of allopurinol across a wide range of outcomes.

The first primary outcome is uric acid level <360 µmol/L. Based on a single small recent study[3], 77% of those on allopurinol are expected to reach this threshold as opposed to 25% of those not taking allopurinol. Secondary outcomes of recurrent gout attacks and allopurinol hypersensitivity syndrome have been evaluated. A nationwide population based study found 22% of newly diagnosed gout patients prescribed with allopurinol had a recurrent gout attack within a year compared to 14% of non-allopurinol users[12]. The proportion of hypersensitivity syndrome cases amongst allopurinol users is known to be 0.4%[13] compared to 0% amongst non-allopurinol users.

Using a significance level of 0.01, the power to detect a difference in proportion in each outcome outlined above between allopurinol and non-allopurinol users is almost 1.

Use of linked data

Deprivation is a risk factor for increased consultation, morbidity and all-cause mortality which would have an impact on resource allocation to general practices. In England, gout prevalence increases with worsening occupational socio-economic status and gout is associated with inadequate income. The long term effects of allopurinol vs. alternative treatments are unknown and any difference arising between treatment groups could potentially partly be explained by deprivation.

Gout has a strong association with the metabolic syndrome in particular hypertension, hyperlipidaemia, insulin resistance and obesity. This increases the risk of diabetes, vascular events and renal diseases which may result in hospitalization. Admissions to hospital would indicate whether gout patients using allopurinol are hospitalised more often than those on alternative treatments.

Knowing the exact cause of death would identify gout patients whose cause of death is gout comorbidity related.

Data /Statistical Analysis

It is aimed that the project starts with the most basic models aimed at estimating effect of allopurinol on various outcomes. Extensions will then be made to more complex analysis in terms of accounting for data structure and type and using propensity score methodology to account for confounding due to observed covariates. Comparisons of the estimated effects of allopurinol will be made between models. Specific phases of model development, starting with the simplest model are:

Phase 1 (Descriptive statistics)

The baseline characteristics and demographics of allopurinol and non-allopurinol users will be described using proportions and means. The association between allopurinol use and outcomes will be described as incidence rates and relative risks.

Phase 2 (Allopurinol at baseline, unadjusted effect)

(a) All outcomes will initially be considered as single binary occurrences. Cox regression model will be used to estimate association between allopurinol at baseline and time to first occurrence of a particular event of interest. Proportionality of hazards assumption will be appropriately tested throughout and adjustments to the Cox model made or alternative non-PH models used. Right censoring will be taken into account, defined as the end of study or death (where death is not outcome of interest) and the censoring mechanism will be assumed non-informative. Time will be treated as a continuous variable.

(b) Extend (a) to recurrent times to event analyses. We will use Andersen-Gill model, an extension of Cox model, and Cox model with shared frailty component (an unobserved patient specific component aimed at accounting for correlation of times to a particular event within a patient) to analyse effect of allopurinol at baseline on recurrent times to event of interest. This will not be applicable to all outcomes. Note that considerations of other models will be made in case of violation of assumptions of these models.

Phase 3 (Allopurinol at baseline, multivariable and propensity score adjustment)

(c) Multivariable adjustment. Extend (a)-(b) by including observed covariates measured at baseline (i.e. closest to the time point leading up to prescription of allopurinol) directly in the model of allopurinol effect on time to event of interest. This is the most common way of adjusting treatment effects and the results will be compared to those obtained following propensity score adjustment.

(d) Propensity scores. Estimate probability to use allopurinol (i.e. propensity score) at baseline via logistic regression. Different choices of covariates at baseline to be included in this process will be investigated (*baseline* for allopurinol and covariates as defined above):

(i) all observed covariates

(ii) all observed covariates related to allopurinol use only

(iii) all observed covariates related to outcome of interest only

(iv) all observed covariates related to both allopurinol use and outcome of interest

Subsequently three methods of incorporating propensity scores in analyses (a)-(b): will be used during estimation of effect of allopurinol on time to event of interest

- Stratification (patients will be ranked according to their propensity score and then split into strata (i.e. probability of using allopurinol based on quintiles of the propensity score). Within each stratum the effect of allopurinol will be calculated and then pooled across strata to get an overall allopurinol effect.

- Matching (Matched pairs between allopurinol and non-allopurinol users who share similar (definition of "similar" dependent on matching algorithm) propensity score will be created. Allopurinol effect can then be estimated by directly comparing the outcomes between allopurinol users and non-users. There are different algorithms to form a matched sample, most popular are greedy, optimal, nearest neighbour and caliper matching algorithms.

- Inverse probability weighting (allopurinol effect e weighted by using the inverse of the propensity score. This approach ensures the distribution of covariates will be independent of allopurinol use).

Phase 4 (Bayesian modelling)

e) By considering allopurinol use at baseline and a single event occurrence, a fully Bayesian model for the joint distribution of data and parameters will be developed, treating propensity score as a latent variable. Then the treatment effect marginal posterior distribution will incorporate uncertainty in propensity scores as it integrates over the latent variable [11]. A comparison will then be made between the Bayesian credible intervals for treatment effect and the intervals obtained using the standard propensity score approaches. In construction of the Bayesian model, we will initially consider non-informative priors only, followed by prior information based on expert opinion and existing research.

Phase 5 (Incorporating allopurinol use over time (i.e. all periods of allopurinol prescription and non-prescription), multivariable and propensity score adjustment)

f) Use of allopurinol may change on multiple occasions over time. Allopurinol use/non-use for each patient will therefore be defined for the entire study duration. Based on the number of allopurinol tablets and the frequency the patient needs to take them for, an estimate of how long the patient was taking allopurinol will be derived. Care will be taken to account for covariates appropriately by taking the covariate measurement closest in time prior to each occurrence of allopurinol prescription. For periods of non-use of allopurinol the midpoint in time will be taken for covariate measures. Timings of repeated events, in relation to allopurinol use, will be taken into account in the model of effect of allopurinol.

g) Propensity score adjustment. Analysis f) above will be repeated, but using time-varying propensity scores, rather than time-varying covariate adjustment. Essentially two approaches will be considered to estimate such propensity scores, firstly they could be estimated at each time point when a treatment is received and secondly using a random effects model.

Phase 6 (estimating effect of omitting important unobserved covariates)

h) In all analyses pertaining to propensity score methodology above, the covariate found to be most influential (i.e. in terms of being strongly related to treatment and outcome, or to outcome alone) will be omitted from the propensity score estimation stage and resulting allopurinol effects compared to findings using a more correct specification of propensity score.

Phase 7 (subgroup analyses)

Stratification of all analyses above will be made on baseline uric acid levels </> 480 µmol/L and on relevant comorbidities (renal and vascular disease) to assess whether effect of allopurinol varies by severity of gout or comorbidity.

Other considerations:

Missing data

During a consultation, all the patient's symptoms and GP's diagnoses may not be recorded therefore it will be assumed that all problems considered by the GP to be of importance at the time will be recorded and those things not recorded were not of importance. Hence, if a symptom or diagnosis is not recorded, this will not be considered 'missing' data. Missing information on patient socio-demographic, lifestyle and general health characteristics will be treated as true missing and will initially be ignored. Subsequently, we will consider multiple imputation, taking into account the repeated measures structure of the data where appropriate.

Data will be managed and analysed in in Stata v13, R (v3.0.2+) and SAS.

It has been noted to preserve confidentiality at the reporting stage and that cells with less than <5 events will not be reported.

Patient or user group involvement

It was not envisaged PPI is required to aid development of this protocol.

Limitations of the study design, data sources, and analytic methods

There is possibility patients may be misclassified as having gout however, a recent systematic review had validated 183 different diagnoses and had shown the median 89% of cases were confirmed using additional internal or external information[14]. With regards to gout, a study had taken a small subsample and had shown patients with a Read code for gout, a high urate level and prescribed anti-gout medication were all confirmed as having gout whereas those with a read code for gout and prescribed anti-gout medication 86% were confirmed as having gout[15]. We can be 90% sure patients consulting for gout would indeed have gout.

Observational studies evaluating the effectiveness of drug effects will be biased due to residual confounding. The methodology outlined in this study attempts to minimise residual confounding. Firstly propensity score analysis will minimise the effect of confounding by indication. Although residual confounding may persist due to unmeasured covariates, the Bayesian methodology described will model the propensity scores as a latent variable taking into account the uncertainty in the estimation of propensity scores.

Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

The CPRD data will be used for a PhD. Results will be written into a thesis and findings will be presented at internal and external meetings and published in peer-reviewed journals. In reporting the results of this study, the STROBE guidelines on the reporting of observational studies will be followed.

Appendix 1

List of Read codes of consultation for gout

Read Code	Read Term
C34	Gout
N023	Gouty arthritis
EGTON 227	Gout NOS
OX2740G	Gout Acute /ox
1443	H/O: gout
EMISR4QG01	Gouty tophi + Gout NOS
2D52	O/E - auricle of ear – tophi
669	Gout monitoring

List of read codes of joint replacement

The appropriate Read codes for joint replacement will be identified from the Read code chapter 7. An example read code list has been provided as follows:

- hip replacement: 7K2..
- knee replacement: 7K3..
- humerus/shoulder replacement: 7K4..
- other joint replacement: 7K6..
- elbow replacement: 7K7..
- ankle replacement: 7K8..

List of read codes of renal disease

Read Code	Read Term	Read Code	Read Term
K032600	Berger's IgA or IgG nephropathy	K07..00	Renal sclerosis unspecified
K0A0500	Acute neph syn, diffuse mesangiocapillary glomerulonephritis	K00..00	Acute glomerulonephritis
K03y.00	Other nephritis and nephrosis unspecified	K08y500	Acute interstitial nephritis
K0A0200	Acute nephritic syn, diffuse membranous glomerulonephritis	SP01500	Mechanical complication of dialysis catheter
1Z12.00	Chronic kidney disease stage 3	K0A4200	Isolatd proteinur/specfd morphlgcl les df membrn glomneph
K02..00	Chronic glomerulonephritis	K016.00	Nephrotic syndrome, diffuse membranous glomerulonephritis
K032.00	Membranoproliferative nephritis unspecified	K0A0300	Acut neph syn, diffuse mesangial prolifrative glomnephritis
K132.00	Acquired cyst of kidney	B91z111	Renal neoplasm of uncertain behaviour
K040.00	Acute renal tubular necrosis	K0A0600	Acute nephritic syndrome, dense deposit disease
K0C1.00	Nephropathy induced by other drugs meds and biologl substncs	K019.00	Nephrotic syn,diffuse mesangiocapillary glomerulonephritis
K017.00	Nephrotic syn difus mesangial prolifertiv glomerulonephritis	C341z00	Gouty nephropathy NOS
G22..11	Nephrosclerosis	G22z.11	Renal hypertension
C104.11	Diabetic nephropathy	K01x300	Nephrotic syndrome in polyarteritis nodosa
7L1B100	Removal of ambulatory peritoneal dialysis catheter	K072.00	Glomerulosclerosis
K0A2300	Recur+persist haemuria df mesangial prolif glomerulnephritis	K0A4500	Isoltd prteinur+specfd morph les df mesangiocap glomnephr
K0z..00	Nephritis, nephrosis and nephrotic syndrome NOS	K08z.00	Impaired renal function disorder NOS
K013.11	Lipoid nephrosis	PD3D.00	Enlarged kidney
K030.00	Proliferative nephritis unspecified	D310100	Henoch-Schonlein nephritis
K032300	Anaphylactoid glomerulonephritis	K0D..00	End-stage renal disease
K043.00	Acute drug-induced renal failure	K03X.00	Unsp nephrit synd, diff mesang prolif glomerulonephritis
K02y200	Chronic focal glomerulonephritis	K0A3700	Chronic nephritic syn diffuse crescentic glomerulonephritis
K05..00	Chronic renal failure	K0A..00	Glomerular disease

K05..11	Chronic uraemia	K0A3100	Chronic nephritic syndrm focal+segmental glomerular lesions
K0A4300	Isoltd prteinur/spcfd morph lesn df mesngl prolf glomneph	K0A2800	IgA nephropathy
K00y300	Acute diffuse nephritis	K0A2200	Recur+persist haematuria difus membranous glomerulonephritis
K08yz11	Renal acidaemia	C109C00	Non-insulin dependent diabetes mellitus with nephropathy
TB11.00	Kidney dialysis with complication, without blame	K03..12	Nephropathy, unspecified
K0A0400	Ac neph syn difus endocaplry prolifrative glomerulonephritis	K032y00	Nephritis unsp+OS membranoprolif glomerulonephritis lesion
K01x400	Nephrotic syndrome in systemic lupus erythematosus	C109C11	Type II diabetes mellitus with nephropathy
7L1A500	Continuous ambulatory peritoneal dialysis	K03W.00	Unsp nephrit synd, diff endocap prolif glomerulonephritis
C10ED00	Type 1 diabetes mellitus with nephropathy	K00..11	Acute nephritis
K032y13	Mesangioproliferative glomerulonephritis NEC	K041.00	Acute renal cortical necrosis
K0A0700	Acute nephrotic syndrm diffuse crescentic glomerulonephritis	K0A2700	Recur+persist haematuria difus crescentic glomerulonephritis
9Ot3.00	Chronic kidney disease monitoring verbal invite	K034.00	Renal cortical necrosis unspecified
K0A1200	Rapid progres neph syn diffuse membranous glomerulonephritis	K0C4.00	Toxic nephropathy, not elsewhere classified
SP08300	Kidney transplant failure and rejection	K0C2.00	Nephropathy induced by unspec drug medicament or biol subs
K032y11	Hypocomplementaemic persistent glomerulonephritis NEC	K022.00	Chronic membranoproliferative glomerulonephritis
C10FC00	Type 2 diabetes mellitus with nephropathy	7L1A400	Automated peritoneal dialysis
C354711	Renal calcinosis	K04..00	Acute renal failure
4519	Deteriorating renal function	K060.00	Renal impairment
9Ot4.00	Chronic kidney disease monitoring telephone invite	S760111	Renal haematoma without mention of open wound into cavity
K08..00	Impaired renal function disorder	K138z11	Renal infarction
A786.00	Haemorrhagic nephrosonephritis	C108D11	Type I diabetes mellitus with nephropathy
K01A.00	Nephrotic syndrome, dense deposit disease	9Ot1.00	Chronic kidney disease monitoring second letter
K00y100	Acute exudative nephritis	C341.00	Gouty nephropathy
K060.11	Impaired renal function	K03yz00	Other nephritis and nephrosis NOS
K00yz00	Other acute glomerulonephritis NOS	K071.00	Renal fibrosis
C373600	Nephropathic amyloidosis	K0A0100	Acute nephritic syndrome, focal+segmental glomerular lesions
K050.00	End stage renal failure	TB00111	Renal transplant with complication, without blame
K023.00	Chronic rapidly progressive glomerulonephritis	K042.00	Acute renal medullary necrosis
K090.00	Unilateral small kidney	SP15400	Renal failure as a complication of care
K0C0.00	Analgesic nephropathy	7L1B000	Insertion of ambulatory peritoneal dialysis catheter
7L1A.11	Dialysis for renal failure	K01x100	Nephrotic syndrome in diabetes mellitus
7B00z00	Transplantation of kidney NOS	K020.00	Chronic proliferative glomerulonephritis
K021.00	Chronic membranous glomerulonephritis	9Ot0.00	Chronic kidney disease monitoring first letter
K0A3500	Chronic neph syn difus mesangiocapillary glomerulonephritis	K014.00	Nephrotic syndrome, minor glomerular abnormality
K138.11	Renal vascular disorders	K0A3.00	Chronic nephritic syndrome
K0A1300	Rpd prog neph syn df mesangial prolifratv glomerulonephritis	K0A1700	Rapid progres nephritic syn df crescentic glomerulonephritis
66i..00	Chronic kidney disease monitoring	7L1A100	Peritoneal dialysis
1Z13.00	Chronic kidney disease stage 4	K08y000	Hypokalaemic nephropathy
K0A3300	Chron neph syn difus mesangial prolifrtiv glomerulonephritis	TB11.11	Renal dialysis with complication, without blame
1Z11.00	Chronic kidney disease stage 2	K00y000	Acute glomerulonephritis in diseases EC
9Ot..00	Chronic kidney disease monitoring administration	1Z10.00	Chronic kidney disease stage 1
K032z00	Nephritis unsp+membranoprolif glomerulonephritis lesion NOS	K0A0.00	Acute nephritic syndrome
K02z.00	Chronic glomerulonephritis NOS	K01x200	Nephrotic syndrome in malaria
K01x411	Lupus nephritis	K032y14	Mesangiocapillary glomerulonephritis NEC
K01x000	Nephrotic syndrome in amyloidosis	K138.00	Vascular disorders of kidney
K03V.00	Unspecified nephritic syndrome, dense deposit disease	K07z.00	Renal sclerosis NOS
K02y300	Chronic diffuse glomerulonephritis	K01B.00	Nephrotic syndrome, diffuse crescentic glomerulonephritis
K04y.00	Other acute renal failure	K011.00	Nephrotic syndrome with membranous glomerulonephritis

K02y000	Chronic glomerulonephritis + diseases EC	K00y200	Acute focal nephritis
K0...00	Nephritis, nephrosis and nephrotic syndrome	K0A1600	Rapid progressive nephritic syndrome, dense deposit disease
7L1A600	Peritoneal dialysis NEC	7L1A000	Renal dialysis
K0A0000	Acute nephritic syndrome, minor glomerular abnormality	7B00100	Transplantation of kidney from live donor
K070.00	Atrophy of kidney	K001.00	Acute nephritis with lesions of necrotising glomerulitis
K0y..00	Other specified nephritis, nephrosis or nephrotic syndrome	K03y000	Other nephritis and nephrosis in diseases EC
K0A1100	Rapid progres nephritic syn focal+segmental glomerulr lesion	K015.00	Nephrotic syndrome, focal and segmental glomerular lesions
K0A1.00	Rapidly progressive nephritic syndrome		
K02..12	Nephropathy - chronic	7L1C000	Insertion of temporary peritoneal dialysis catheter
K03U.00	Unspecif nephr synd, diff concentric glomerulonephritis	K01..00	Nephrotic syndrome
7B00200	Transplantation of kidney from cadaver	K018.00	Nephrotic syn,difus endocapilary proliftv glomerulonephritis
K081.00	Nephrogenic diabetes insipidus	K08y400	Renal tubular acidosis
K13yz11	Salt-losing nephritis	K000.00	Acute proliferative glomerulonephritis
K032y15	Mixed membranous and proliferative glomerulonephritis NEC	K010.00	Nephrotic syndrome with proliferative glomerulonephritis
K00z.00	Acute glomerulonephritis NOS	K0A3600	Chronic nephritic syndrome, dense deposit disease
K03y200	Other interstitial nephritis	K03..11	Nephritis and nephropathy unspecified
K06..00	Renal failure unspecified	K031.00	Membranous nephritis unspecified
K013.12	Steroid sensitive nephrotic syndrome	K03T.00	Tubulo-interstit nephritis, not specif as acute or chron
TB00100	Kidney transplant with complication, without blame	K02..11	Nephritis - chronic
7L1A200	Haemodialysis NEC	K01y.00	Nephrotic syndrome with other pathological kidney lesions
1Z1..00	Chronic renal impairment	K033.00	Rapidly progressive nephritis unspecified
K13z000	Non-functioning kidney	SP15411	Kidney failure as a complication of care
K03z.00	Unspecified glomerulonephritis NOS	C108D00	Insulin dependent diabetes mellitus with nephropathy
1Z14.00	Chronic kidney disease stage 5	C354700	Nephrocalcinosis
K02y.00	Other chronic glomerulonephritis	K0A2500	Recur+persist hmuria df mesangiocapilary glomerulonephritis
K01z.00	Nephrotic syndrome NOS	K0A3000	Chronic nephritic syndrome, minor glomerular abnormality
K013.00	Nephrotic syndrome with minimal change glomerulonephritis	K04z.00	Acute renal failure NOS
G222.00	Hypertensive renal disease with renal failure	PDz0.00	Unspecified anomaly of kidney
9Ot2.00	Chronic kidney disease monitoring third letter	K032000	Focal membranoproliferative glomerulonephritis
7B00.00	Transplantation of kidney	K0A7.00	Glom disordr in blood diseas+disordr invlvg imun mechansm
K0A3200	Chron nephritic syndrom difuse membranous glomerulonephritis	K012.00	Nephrotic syndrome+membranoproliferative glomerulonephritis
K02yz00	Other chronic glomerulonephritis NOS	K00y.00	Other acute glomerulonephritis
K035.00	Renal medullary necrosis unspecified	K138z00	Renal vascular disorders NOS

List of read codes for vascular disease

Cardiovascular, cerebrovascular and peripheral vascular Read codes will be identified.

List of cardiovascular Read codes

Read Code	Read Term	Read Code	Read Term
14A..00	H/O: cardiovascular disease	32B..00	ECG: Q wave
14A..12	H/O: myocardial problem	32B2.00	ECG: Q wave abnormal
14A4.00	H/O: myocardial infarct >60	44p2.00	Cardiac troponin positive
14A5.00	H/O: angina pectoris	5543	Coronary arteriograph.abnormal
14AH.00	H/O: Myocardial infarction in last year	662..00	Cardiac disease monitoring
14AJ.00	H/O: Angina in last year	662K000	Angina control - good
14AL.00	H/O: Treatment for ischaemic heart disease	662K100	Angina control - poor
14AZ.00	H/O: CVS disease NOS	662K200	Angina control - improving
14N6.00	H/O: cardiac surgery	662K300	Angina control - worsening
182A.00	Chest pain on exertion	662N.00	CHD monitoring

3213100	Exercise ECG abnormal	66f..00	Cardiovascular disease monitoring
3213111	Positive exercise ECG test	6A2..00	Coronary heart disease annual review
322..00	ECG: myocardial ischaemia	6A4..00	Coronary heart disease review
3222	ECG:shows myocardial ischaemia	790H300	Revascularisation of wall of heart
322Z.00	ECG: myocardial ischaemia NOS	792..00	Coronary artery operations
323..00	ECG: myocardial infarction	792..11	Coronary artery bypass graft operations
3232	ECG: old myocardial infarction	7920	Saphenous vein graft replacement of coronary artery
3233	ECG: antero-septal infarct.	7920.11	Saphenous vein graft bypass of coronary artery
3234	ECG:posterior/inferior infarct	7920000	Saphenous vein graft replacement of one coronary artery
3235	ECG: subendocardial infarct	7920100	Saphenous vein graft replacement of two coronary arteries
3236	ECG: lateral infarction	7920200	Saphenous vein graft replacement of three coronary arteries
323Z.00	ECG: myocardial infarct NOS	7920300	Saphenous vein graft replacement of four+ coronary arteries
7920y00	Saphenous vein graft replacement of coronary artery OS	7926	Connection of other thoracic artery to coronary artery
7920z00	Saphenous vein graft replacement coronary artery NOS	7926000	Double anastom thoracic arteries to coronary arteries NEC
7921	Other autograft replacement of coronary artery	7926200	Single anastomosis of thoracic artery to coronary artery NEC
7921.11	Other autograft bypass of coronary artery	7926300	Single implantation thoracic artery into coronary artery NEC
7921000	Autograft replacement of one coronary artery NEC	7926z00	Connection of other thoracic artery to coronary artery NOS
7921100	Autograft replacement of two coronary arteries NEC	7927	Other open operations on coronary artery
7921200	Autograft replacement of three coronary arteries NEC	7927000	Repair of arteriovenous fistula of coronary artery
7921300	Autograft replacement of four of more coronary arteries NEC	7927100	Repair of aneurysm of coronary artery
7921y00	Other autograft replacement of coronary artery OS	7927300	Transposition of coronary artery NEC
7921z00	Other autograft replacement of coronary artery NOS	7927400	Exploration of coronary artery
7922	Allograft replacement of coronary artery	7927500	Open angioplasty of coronary artery
7922.11	Allograft bypass of coronary artery	7927y00	Other specified other open operation on coronary artery
7922000	Allograft replacement of one coronary artery	7927z00	Other open operation on coronary artery NOS
7922100	Allograft replacement of two coronary arteries	7928	Transluminal balloon angioplasty of coronary artery
7922200	Allograft replacement of three coronary arteries	7928.11	Percutaneous balloon coronary angioplasty
7922300	Allograft replacement of four or more coronary arteries	7928000	Percut transluminal balloon angioplasty one coronary artery
7922y00	Other specified allograft replacement of coronary artery	7928100	Percut translum balloon angioplasty mult coronary arteries
7922z00	Allograft replacement of coronary artery NOS	7928200	Percut translum balloon angioplasty bypass graft coronary a
7923	Prosthetic replacement of coronary artery	7928300	Percut translum cutting balloon angioplasty coronary artery
7923.11	Prosthetic bypass of coronary artery	7928y00	Transluminal balloon angioplasty of coronary artery OS
7923000	Prosthetic replacement of one coronary artery	7928z00	Transluminal balloon angioplasty of coronary artery NOS
7923100	Prosthetic replacement of two coronary arteries	7929	Other therapeutic transluminal operations on coronary artery
7923200	Prosthetic replacement of three coronary arteries	7929000	Percutaneous transluminal laser coronary angioplasty
7923300	Prosthetic replacement of four or more coronary arteries	7929100	Percut transluminal coronary thrombolysis with streptokinase
7923z00	Prosthetic replacement of coronary artery NOS	7929111	Percut translum coronary thrombolytic therapy-streptokinase
7924	Revision of bypass for coronary artery	7929200	Percut translum inject therap subst to coronary artery NEC
7924000	Revision of bypass for one coronary artery	7929300	Rotary blade coronary angioplasty
7924100	Revision of bypass for two coronary arteries	7929400	Insertion of coronary artery stent
7924200	Revision of bypass for three coronary arteries	7929500	Insertion of drug-eluting coronary artery stent
7924y00	Other specified revision of bypass for coronary artery	7929600	Percutaneous transluminal atherectomy of coronary artery
7925	Connection of mammary artery to coronary artery	7929y00	Other therapeutic transluminal op on coronary artery OS
7925.11	Creation of bypass from mammary artery to coronary artery	7929z00	Other therapeutic transluminal op on coronary artery NOS
7925000	Double anastomosis of mammary arteries to coronary arteries	792A.00	Diagnostic transluminal operations on coronary artery
7925011	LIMA sequential anastomosis	792A000	Percutaneous transluminal angioscopy
7925012	RIMA sequential anastomosis	792A100	Intravascular ultrasound of coronary artery

7925100	Double implant of mammary arteries into coronary arteries	792Ay00	Diagnostic transluminal operation on coronary artery OS
7925300	Single anastomosis of mammary artery to coronary artery NEC	792Az00	Diagnostic transluminal operation on coronary artery NOS
7925311	LIMA single anastomosis	792B.00	Repair of coronary artery NEC
7925312	RIMA single anastomosis	792B000	Endarterectomy of coronary artery NEC
7925400	Single implantation of mammary artery into coronary artery	792B100	Repair of rupture of coronary artery
7925y00	Connection of mammary artery to coronary artery OS	792B200	Repair of arteriovenous malformation of coronary artery
7925z00	Connection of mammary artery to coronary artery NOS	792By00	Other specified repair of coronary artery
792Bz00	Repair of coronary artery NOS	9Ob5.00	Coronary heart disease monitoring 3rd letter
792C.00	Other replacement of coronary artery	9Ob6.00	Coronary heart disease monitoring verbal invitation
792C000	Replacement of coronary arteries using multiple methods	G3...00	Ischaemic heart disease
792Cy00	Other specified replacement of coronary artery	G3...11	Arteriosclerotic heart disease
792Cz00	Replacement of coronary artery NOS	G3...12	Atherosclerotic heart disease
792D.00	Other bypass of coronary artery	G3...13	IHD - Ischaemic heart disease
792Dy00	Other specified other bypass of coronary artery	G30..00	Acute myocardial infarction
792Dz00	Other bypass of coronary artery NOS	G30..11	Attack - heart
792y.00	Other specified operations on coronary artery	G30..12	Coronary thrombosis
792z.00	Coronary artery operations NOS	G30..13	Cardiac rupture following myocardial infarction (MI)
7932z00	Open operation on heart NOS	G30..14	Heart attack
7938000	Angiocardiography of both right and left sides of heart	G30..15	MI - acute myocardial infarction
7938100	Angiocardiography of right side of heart NEC	G30..16	Thrombosis - coronary
7938200	Angiocardiography of left side of heart NEC	G30..17	Silent myocardial infarction
7938300	Coronary arteriography using two catheters	G300.00	Acute anterolateral infarction
7938400	Coronary arteriography using single catheter	G301.00	Other specified anterior myocardial infarction
7938500	Coronary arteriography NEC	G301000	Acute anteroapical infarction
7938600	Coronary arteriography using three catheters	G301100	Acute anteroseptal infarction
7939.11	Cardiac catheterisation	G301z00	Anterior myocardial infarction NOS
7939000	Catheterisation of both right and left sides of heart NEC	G302.00	Acute inferolateral infarction
7939100	Catheterisation of right side of heart NEC	G303.00	Acute inferoposterior infarction
7939200	Catheterisation of left side of heart NEC	G304.00	Posterior myocardial infarction NOS
7939y00	Other specified catheterisation of heart	G305.00	Lateral myocardial infarction NOS
7939z00	Catheterisation of heart NOS	G306.00	True posterior myocardial infarction
793G.00	Perc translumin balloon angioplasty stenting coronary artery	G307.00	Acute subendocardial infarction
793G000	Perc translum ball angio insert 1-2 drug elut stents cor art	G307000	Acute non-Q wave infarction
793G100	Perc tran ball angio ins 3 or more drug elut stents cor art	G307100	Acute non-ST segment elevation myocardial infarction
793G200	Perc translum balloon angioplasty insert 1-2 stents cor art	G308.00	Inferior myocardial infarction NOS
793G300	Percutaneous cor balloon angiop 3 more stents cor art NEC	G309.00	Acute Q-wave infarct
793Gy00	OS perc translumina balloon angioplast stenting coronary art	G30A.00	Mural thrombosis
793Gz00	Perc translum balloon angioplasty stenting coronary art NOS	G30B.00	Acute posterolateral myocardial infarction
793K.00	Transluminal operations internal mammary artery side branch	G30X.00	Acute transmural myocardial infarction of unspecif site
793K000	Translum occlusion left internal mammary artery side branch	G30X000	Acute ST segment elevation myocardial infarction
88A8.00	Thrombolytic therapy	G30y.00	Other acute myocardial infarction
8B3k.00	Coronary heart disease medication review	G30y000	Acute atrial infarction
8B63.11	Aspirin prophylaxis - IHD	G30y100	Acute papillary muscle infarction
8H2V.00	Admit ischaemic heart disease emergency	G30y200	Acute septal infarction
9Ob..00	Coronary heart disease monitoring administration	G30yz00	Other acute myocardial infarction NOS
9Ob0.00	Attends coronary heart disease monitoring	G30z.00	Acute myocardial infarction NOS
9Ob2.00	Coronary heart disease monitoring default	G31..00	Other acute and subacute ischaemic heart disease
9Ob3.00	Coronary heart disease monitoring 1st letter	G34yz00	Other specified chronic ischaemic heart disease NOS
9Ob4.00	Coronary heart disease monitoring 2nd letter	G34z.00	Other chronic ischaemic heart disease NOS

G310.00	Postmyocardial infarction syndrome	G34z000	Asymptomatic coronary heart disease
G310.11	Dressler's syndrome	G35..00	Subsequent myocardial infarction
G311.00	Preinfarction syndrome	G350.00	Subsequent myocardial infarction of anterior wall
G311.11	Crescendo angina	G351.00	Subsequent myocardial infarction of inferior wall
G311.13	Unstable angina	G353.00	Subsequent myocardial infarction of other sites
G311.14	Angina at rest	G35X.00	Subsequent myocardial infarction of unspecified site
G311100	Unstable angina	G36..00	Certain current complication follow acute myocardial infarct
G311200	Angina at rest	G360.00	Haemopericardium/current comp folow acut myocard infarct
G311300	Refractory angina	G361.00	Atrial septal defect/curr comp folow acut myocardal infarct
G311400	Worsening angina	G362.00	Ventric septal defect/curr comp fol acut myocardal infarctn
G311500	Acute coronary syndrome	G363.00	Ruptur cardiac wall w'out haemopericard/cur comp fol ac MI
G311z00	Preinfarction syndrome NOS	G364.00	Ruptur chordae tendinae/curr comp fol acute myocard infarct
G312.00	Coronary thrombosis not resulting in myocardial infarction	G365.00	Rupture papillary muscle/curr comp fol acute myocard infarct
G31y.00	Other acute and subacute ischaemic heart disease	G366.00	Thrombosis atrium,auric append&vent/curr comp foll acute MI
G31y000	Acute coronary insufficiency	G37..00	Cardiac syndrome X
G31y100	Microinfarction of heart	G38..00	Postoperative myocardial infarction
G31y200	Subendocardial ischaemia	G380.00	Postoperative transmural myocardial infarction anterior wall
G31y300	Transient myocardial ischaemia	G381.00	Postoperative transmural myocardial infarction inferior wall
G31yz00	Other acute and subacute ischaemic heart disease NOS	G384.00	Postoperative subendocardial myocardial infarction
G32..00	Old myocardial infarction	G38z.00	Postoperative myocardial infarction, unspecified
G32..11	Healed myocardial infarction	G3y..00	Other specified ischaemic heart disease
G32..12	Personal history of myocardial infarction	G3z..00	Ischaemic heart disease NOS
G33..00	Angina pectoris	G5...00	Other forms of heart disease
G330.00	Angina decubitus	G5y..00	Other specified heart disease
G330000	Nocturnal angina	G5y1.00	Myocardial degeneration
G330z00	Angina decubitus NOS	G5y2.00	Cardiovascular arteriosclerosis unspecified
G331.00	Prinzmetal's angina	G5yy.00	Other ill-defined heart disease
G331.11	Variant angina pectoris	G5yyz00	Other ill-defined heart disease NOS
G332.00	Coronary artery spasm	G5yz.00	Other heart disease NOS
G33z.00	Angina pectoris NOS	G5z..00	Heart disease NOS
G33z000	Status anginosus	Gyu3.00	[X]Ischaemic heart diseases
G33z100	Stenocardia	Gyu3000	[X]Other forms of angina pectoris
G33z200	Syncope anginosa	Gyu3200	[X]Other forms of acute ischaemic heart disease
G33z300	Angina on effort	Gyu3300	[X]Other forms of chronic ischaemic heart disease
G33z400	Ischaemic chest pain	Gyu3400	[X]Acute transmural myocardial infarction of unspecif site
G33z500	Post infarct angina	Gyu3600	[X]Subsequent myocardial infarction of unspecified site
G33z600	New onset angina	ZV45800	[V]Presence of coronary angioplasty implant and graft
G33z700	Stable angina	ZV45K00	[V]Presence of coronary artery bypass graft
G33zz00	Angina pectoris NOS	ZV45K11	[V]Presence of coronary artery bypass graft - CABG
G34..00	Other chronic ischaemic heart disease	ZV45L00	[V]Status following coronary angioplasty NOS
G340.00	Coronary atherosclerosis	G340000	Single coronary vessel disease
G340.11	Triple vessel disease of the heart	G340100	Double coronary vessel disease
G340.12	Coronary artery disease	G342.00	Atherosclerotic cardiovascular disease
G343.00	Ischaemic cardiomyopathy	G34y000	Chronic coronary insufficiency
G344.00	Silent myocardial ischaemia	G34y100	Chronic myocardial ischaemia
G34y.00	Other specified chronic ischaemic heart disease		

List of cerebrovascular Read codes

Read code	Read term	Read code	Read term
1477	H/O: cerebrovascular disease	7A20311	Carotid endarterectomy and patch
14A7.00	H/O: CVA/stroke	7A20400	Endarterectomy of carotid artery NEC
14A7.11	H/O: CVA	7A20700	Intracranial bypass from carotid artery NEC
14A7.12	H/O: stroke	7A21200	Open embolectomy of carotid artery
14AF.00	H/O sub-arachnoid haemorrhage	7A22300	Percutaneous transluminal insertion stent carotid artery
14AK.00	H/O: Stroke in last year	7A24600	Open embolisation of cerebral artery
5513	Carotid A angiogram abnormal	7A25100	Percutaneous transluminal embolisation of circle of Willis
5C10.00	Carotid artery doppler abnormal	7A25600	Percutaneous transluminal insertion of stent cerebral artery
662M.00	Stroke monitoring	8HBJ.00	Stroke / transient ischaemic attack referral
662e.00	Stroke/CVA annual review	8HTQ.00	Referral to stroke clinic
662o.00	Haemorrhagic stroke monitoring	9N0p.00	Seen in stroke clinic
7A20.00	Reconstruction of carotid artery	9Om..00	Stroke/transient ischaemic attack monitoring administration
7A20000	Replacement of carotid artery using graft	9Om0.00	Stroke/transient ischaemic attack monitoring first letter
7A20100	Intracranial bypass to carotid artery	9Om1.00	Stroke/transient ischaemic attack monitoring second letter
7A20300	Endarterectomy and patch repair of carotid artery	9Om2.00	Stroke/transient ischaemic attack monitoring third letter
F11x200	Cerebral degeneration due to cerebrovascular disease	9Om3.00	Stroke/transient ischaemic attack monitoring verbal invitati
G6...00	Cerebrovascular disease	9Om4.00	Stroke/transient ischaemic attack monitoring telephone invte
G60..00	Subarachnoid haemorrhage	G640000	Cerebral infarction due to thrombosis of cerebral arteries
G600.00	Ruptured berry aneurysm	G641.00	Cerebral embolism
G601.00	Subarachnoid haemorrhage from carotid siphon and bifurcation	G641.11	Cerebral embolus
G602.00	Subarachnoid haemorrhage from middle cerebral artery	G641000	Cerebral infarction due to embolism of cerebral arteries
G603.00	Subarachnoid haemorrhage from anterior communicating artery	G64z.00	Cerebral infarction NOS
G604.00	Subarachnoid haemorrhage from posterior communicating artery	G64z.11	Brainstem infarction NOS
G605.00	Subarachnoid haemorrhage from basilar artery	G64z.12	Cerebellar infarction
G606.00	Subarachnoid haemorrhage from vertebral artery	G64z000	Brainstem infarction
G60X.00	Subarachnoid haemorrh from intracranial artery, unspecif	G64z100	Wallenberg syndrome
G60z.00	Subarachnoid haemorrhage NOS	G64z111	Lateral medullary syndrome
G61..00	Intracerebral haemorrhage	G64z200	Left sided cerebral infarction
G61..11	CVA - cerebrovascular accid due to intracerebral haemorrhage	G64z300	Right sided cerebral infarction
G61..12	Stroke due to intracerebral haemorrhage	G64z400	Infarction of basal ganglia
G610.00	Cortical haemorrhage	G65..00	Transient cerebral ischaemia
G611.00	Internal capsule haemorrhage	G65..12	Transient ischaemic attack
G612.00	Basal nucleus haemorrhage	G65..13	Vertebro-basilar insufficiency
G613.00	Cerebellar haemorrhage	G650.00	Basilar artery syndrome
G614.00	Pontine haemorrhage	G650.11	Insufficiency - basilar artery
G615.00	Bulbar haemorrhage	G651.00	Vertebral artery syndrome
G616.00	External capsule haemorrhage	G651000	Vertebro-basilar artery syndrome
G617.00	Intracerebral haemorrhage, intraventricular	G652.00	Subclavian steal syndrome
G618.00	Intracerebral haemorrhage, multiple localized	G653.00	Carotid artery syndrome hemispheric
G61X.00	Intracerebral haemorrhage in hemisphere, unspecified	G654.00	Multiple and bilateral precerebral artery syndromes
G61X000	Left sided intracerebral haemorrhage, unspecified	G655.00	Transient global amnesia
G61X100	Right sided intracerebral haemorrhage, unspecified	G656.00	Vertebrobasilar insufficiency
G61z.00	Intracerebral haemorrhage NOS	G65y.00	Other transient cerebral ischaemia
G62..00	Other and unspecified intracranial haemorrhage	G65z.00	Transient cerebral ischaemia NOS
G62z.00	Intracranial haemorrhage NOS	G65z100	Intermittent cerebral ischaemia
G63..00	Precerebral arterial occlusion	G65zz00	Transient cerebral ischaemia NOS
G63..11	Infarction - precerebral	G66..00	Stroke and cerebrovascular accident unspecified
G63..12	Stenosis of precerebral arteries	G66..11	CVA unspecified
G630.00	Basilar artery occlusion	G66..12	Stroke unspecified

G631.00	Carotid artery occlusion	G66..13	CVA - Cerebrovascular accident unspecified
G631.11	Stenosis, carotid artery	G660.00	Middle cerebral artery syndrome
G631.12	Thrombosis, carotid artery	G661.00	Anterior cerebral artery syndrome
G632.00	Vertebral artery occlusion	G662.00	Posterior cerebral artery syndrome
G633.00	Multiple and bilateral precerebral arterial occlusion	G663.00	Brain stem stroke syndrome
G634.00	Carotid artery stenosis	G664.00	Cerebellar stroke syndrome
G63y.00	Other precerebral artery occlusion	G665.00	Pure motor lacunar syndrome
G63y000	Cerebral infarct due to thrombosis of precerebral arteries	G666.00	Pure sensory lacunar syndrome
G63y100	Cerebral infarction due to embolism of precerebral arteries	G667.00	Left sided CVA
G63z.00	Precerebral artery occlusion NOS	G668.00	Right sided CVA
G64..00	Cerebral arterial occlusion	G67..00	Other cerebrovascular disease
G64..11	CVA - cerebral artery occlusion	G670.00	Cerebral atherosclerosis
G64..12	Infarction - cerebral	G670.11	Precerebral atherosclerosis
G64..13	Stroke due to cerebral arterial occlusion	G671.00	Generalised ischaemic cerebrovascular disease NOS
G640.00	Cerebral thrombosis	G671000	Acute cerebrovascular insufficiency NOS
G673300	Vertebral artery dissection	G671100	Chronic cerebral ischaemia
G674.00	Cerebral arteritis	G671z00	Generalised ischaemic cerebrovascular disease NOS
G674000	Cerebral amyloid angiopathy	G672.00	Hypertensive encephalopathy
G676.00	Nonpyogenic venous sinus thrombosis	G673.00	Cerebral aneurysm, nonruptured
G676000	Cereb infarct due cerebral venous thrombosis, nonpyogenic	G673000	Dissection of cerebral arteries, nonruptured
G677.00	Occlusion/stenosis cerebral arts not result cerebral infarct	G673100	Carotico-cavernous sinus fistula
G677000	Occlusion and stenosis of middle cerebral artery	G673200	Carotid artery dissection
G677100	Occlusion and stenosis of anterior cerebral artery	G70y000	Carotid artery atherosclerosis
G677200	Occlusion and stenosis of posterior cerebral artery	G70y011	Carotid artery disease
G677300	Occlusion and stenosis of cerebellar arteries	G72y000	Aneurysm of common carotid art
G677400	Occlusion+stenosis of multiple and bilat cerebral arteries	G72y100	Aneurysm of external carotid artery
G679.00	Small vessel cerebrovascular disease	G72y200	Aneurysm of internal carotid artery
G67y.00	Other cerebrovascular disease OS	G755000	Cranial arteritis
G67z.00	Other cerebrovascular disease NOS	Gyu6.00	[X]Cerebrovascular diseases
G68..00	Late effects of cerebrovascular disease	Gyu6300	[X]Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
G680.00	Sequelae of subarachnoid haemorrhage	Gyu6400	[X]Other cerebral infarction
G681.00	Sequelae of intracerebral haemorrhage	Gyu6500	[X]Occlusion and stenosis of other precerebral arteries
G682.00	Sequelae of other nontraumatic intracranial haemorrhage	Gyu6600	[X]Occlusion and stenosis of other cerebral arteries
G683.00	Sequelae of cerebral infarction	Gyu6A00	[X]Other cerebrovascular disorders in diseases CE
G68W.00	Sequelae/other + unspecified cerebrovascular diseases	Gyu6F00	[X]Intracerebral haemorrhage in hemisphere, unspecified
G68X.00	Sequelae of stroke,not specfd as h'morrhage or infarction	Gyu6G00	[X]Cereb infarct due unsp occlus/stenos precerebr arteries
G6W..00	Cereb infarct due unsp occlus/stenos precerebr arteries	ZLEP.00	Discharge from stroke serv
G6X..00	Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs	ZV12511	[V]Personal history of stroke
G6y..00	Other specified cerebrovascular disease	ZV12512	[V]Personal history of cerebrovascular accident (CVA)
G6z..00	Cerebrovascular disease NOS		

List of peripheral vascular Read codes

Read code	Read term	Read code	Read term
662U.00	Peripheral vascular disease monitoring	7A41C00	Bypass leg artery by aorta/deep femoral art anastomosis NEC
7A10300	Axillo-unifemoral PTFE bypass graft	7A41D00	Bypass iliac artery by iliac/iliac artery anastomosis NEC
7A26A00	Endarterectomy and patch repair of vertebral artery	7A41F00	Ilio-femoral prosthetic cross over graft
7A28000	Percutaneous transluminal angioplasty of subclavian artery	7A41y00	Other specified other bypass of iliac artery
7A28200	Percutaneous transluminal angioplasty of vertebral artery	7A41z00	Other bypass of iliac artery NOS
7A30500	Patch angioplasty of renal artery	7A42000	Endarterectomy and patch repair of iliac artery

7A32000	Percutaneous transluminal angioplasty of renal artery	7A42011	Endarterectomy and patch repair of common iliac artery
7A32400	Percutan transluminal balloon angioplasty stenting renal art	7A42012	Iliac endarterectomy and patch
7A32500	Percutaneous transluminal insertion stent into renal artery	7A42100	Endarterectomy of iliac artery NEC
7A33100	Bypass of superior mesenteric artery NEC	7A42111	Endarterectomy of common iliac artery NEC
7A35300	Percutaneous transluminal angioplasty suprarenal artery NEC	7A44400	Percutaneous transluminal insertion of iliac artery stent
7A41000	Emerg bypass iliac art by iliac/femoral art anastomosis NEC	7A47.00	Other emergency bypass of femoral artery or popliteal artery
7A41100	Bypass iliac artery by iliac/femoral artery anastomosis NEC	7A47.11	Other emerg bypass femoral or popliteal art by anastomosis
7A41200	Emerg bypass iliac artery by femoral/femoral art anast NEC	7A47.12	Other emergency bypass of common femoral artery
7A41211	Emergency femoro-femoral prosthetic cross over graft	7A47.13	Other emergency bypass of deep femoral artery
7A41300	Bypass iliac artery by femoral/femoral art anastomosis NEC	7A47.14	Other emergency bypass of popliteal artery
7A41311	Femoro-femoral prosthetic cross over graft	7A47.15	Other emergency bypass of superficial femoral artery
7A41400	Emerg bypass comm iliac art by aorta/com iliac art anast NEC	7A47.16	Other emergency bypass of femoral artery
7A41600	Emerg bypass leg artery by aorta/com fem art anastomosis NEC	7A47000	Emerg bypass femoral art by fem/pop art anast c prosth NEC
7A41900	Bypass common iliac artery by aorta/com iliac art anast NEC	7A47100	Emerg bypass popliteal art by pop/pop art anast c prosth NEC
7A41B00	Bypass leg artery by aorta/com femoral art anastomosis NEC	7A47200	Emerg bypass femoral art by fem/pop a anast c vein graft NEC
7A48C00	Bypass femoral artery by femoral/femoral art anastomosis NEC	7A47300	Emerg bypass pop art by pop/pop art anast c vein graft NEC
7A48D00	Bypass popliteal artery by pop/fem artery anastomosis NEC	7A47400	Emerg bypass femoral art by fem/tib art anast c prosth NEC
7A48E00	Femoro-femoral prosthetic cross over graft	7A47600	Emerg bypass femoral art by fem/tib a anast c vein graft NEC
7A48y00	Other bypass of femoral artery or popliteal artery OS	7A47700	Emerg bypass pop art by pop/tib art anast c vein graft NEC
7A48z00	Other bypass of femoral artery or popliteal artery NOS	7A47800	Emerg bypass femoral art by fem/peron art anast c prosth NEC
7A49100	Endarterectomy and patch repair of popliteal artery	7A47B00	Emerg bypass pop art by pop/peron art anast c vein graft NEC
7A49300	Endarterectomy of popliteal artery NEC	7A47C00	Emerg bypass femoral artery by fem/fem art anastomosis NEC
7A49400	Profundoplasty femoral artery & patch repair deep fem artery	7A47D00	Emerg bypass popliteal artery by pop/fem art anastomosis NEC
7A49500	Profundoplasty and patch repair of popliteal artery	7A47y00	Other emergency bypass of femoral or popliteal artery OS
7A49600	Profundoplasty of femoral artery NEC	7A47z00	Other emergency bypass of femoral or popliteal artery NOS
7A49700	Profundoplasty of popliteal artery NEC	7A48.00	Other bypass of femoral artery or popliteal artery
7A4B000	Percutaneous transluminal angioplasty of femoral artery	7A48.11	Other bypass of femoral or popliteal artery by anastomosis
7A4B100	Percutaneous transluminal angioplasty of popliteal artery	7A48.12	Other bypass of common femoral artery
7A4B900	Percutaneous transluminal insertion of stent femoral artery	7A48.15	Other bypass of popliteal artery
7A56600	Percutaneous transluminal placement peripheral stent artery	7A48.16	Other bypass of superficial femoral artery
9N4h.00	DNA - Did not attend peripheral vascular disease clinic	7A48000	Bypass femoral artery by fem/pop art anast c prosthesis NEC
G701.00	Renal artery atherosclerosis	7A48100	Bypass popliteal artery by pop/pop a anast c prosthesis NEC
G702.00	Extremity artery atheroma	7A48200	Bypass femoral artery by fem/pop art anast c vein graft NEC
G702z00	Extremity artery atheroma NOS	7A48300	Bypass popliteal artery by pop/pop a anast c vein graft NEC
G703.00	Acquired renal artery stenosis	7A48400	Bypass femoral artery by fem/tib art anast c prosthesis NEC
G73..00	Other peripheral vascular disease	7A48600	Bypass femoral artery by fem/tib art anast c vein graft NEC
G73..11	Peripheral ischaemic vascular disease	7A48700	Bypass popliteal artery by pop/tib a anast c vein graft NEC
G73..12	Ischaemia of legs	7A48800	Bypass femoral artery by fem/peron a anast c prosthesis NEC
G73..13	Peripheral ischaemia	7A48A00	Bypass femoral artery by fem/peron a anast c vein graft NEC
G731000	Buerger's disease	7A48B00	Bypass popliteal art by pop/peron art anast c vein graft NEC
G731100	Presenile gangrene	G73yz00	Other specified peripheral vascular disease NOS
G732.00	Peripheral gangrene	G73z.00	Peripheral vascular disease NOS
G732000	Gangrene of toe	G73z000	Intermittent claudication

G732100	Gangrene of foot	G73zz00	Peripheral vascular disease NOS
G733.00	Ischaemic foot	G763.00	Hyperplasia of renal artery
G73y.00	Other specified peripheral vascular disease	G768000	Fibromuscular hyperplasia of arteries NOS
G73y000	Diabetic peripheral angiopathy	G768100	Arterial fibromuscular dysplasia
G73y100	Peripheral angiopathic disease EC NOS	Gyu7400	[X]Other specified peripheral vascular diseases
SP12.00	Peripheral vascular complications of care		

Glossary of Acronyms

CONSORT	Consolidated Standards of Reporting Trials
CPRD	Clinical Practice Research Datalink
EQUATOR	Enhancing the Quality and Transparency of health research network
GP	General Practitioner
ISAC	Independent Scientific Advisory Committee for MHRA database research
MREC	Multi-centre NHS Research Ethic Committee
MHRA	Medicines and Healthcare products Regulatory Agency
NHS	National Health Service
NRR	National Research Register
OXMIS	Oxford Medical Information Systems (codes)
REC	NHS Research Ethics Committee
STROBE	Strengthening the Reporting of Observational studies in Epidemiology
UK	United Kingdom
VRMM	Vigilance and Risk Management division of MHRA

References

1. Kuo CF, Grainge MJ, Mallen C, et al. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann Rheum Dis* 2014; doi:10.1136/annrheumdis-2013-204463
2. Arthritis Research UK National Primary Care Centre, Keele University. *Musculoskeletal Matters Bulletin no 2: Consultations for selected diagnoses and regional problems*. Available from http://www.keele.ac.uk/media/keeleuniversity/ri/primarycare/bulletins/MusculoskeletalMatters2.pdf (accessed 31/07/14).
3. Li-Yu J, Clayburne G, Sieck M, et al. Treatment of chronic gout. Can we determine when urate stores are depleted enough to prevent attacks of gout? *J Rheumatol* 2001; 28(3):577-80
4. Becker MA, Schumacher HR Jr, Wortmann RL, et al. Febuxostat compared with allopurinol in patients with hyperuricaemia and gout. *N Engl J Med* 2005; 353:2450-61
5. Schumacher HR Jr, Becker MA, Wortmann RL, et al. Effects of febuxostat versus allopurinol and placebo in reducing serum urate in subjects with hyperuricaemia and gout: a 28 week, phase III, randomized double-blind, parallel-group trial. *Arthritis Rheum* 2008; 59:1540-8
6. Becker MA, Schumacher HR, Espinoza LR, et al. The urate-lowering efficacy and safety of febuxostat in the treatment of the hyperuricaemia of gout: the CONFIRMS trial. *Arthritis Res Ther* 2010; 12:R63
7. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41-55.
8. Dubreuil M, Zhu Y, Zhang Y, et al. Allopurinol titration and all-cause mortality in the general population. *Ann Rheum Dis* 2014; doi: 10.1136/annrheumdis-2014-205269
9. Kim SC, Newcomb C, Margolis D, et al. Severe cutaneous reactions requiring hospitalization in allopurinol inhibitors: A population-based cohort study. *Arthritis Care Res* 2013; 65: 578-584
10. Wei L, Mackenzie IS, Chen Y, et al. Impact of allopurinol use on urate concentration and cardiovascular outcome. *Br J Clin Pharmacol* 2011; 71:600-7
11. McCandless LC, Gustafson P, Austin PC. Bayesian propensity score analysis for observational data. *Stat Med* 2009; 28:94-112
12. Trifirò G, Morabito P, Cavagna L, et al. Epidemiology of gout and hyperuricaemia in Italy during the years 2005–2009: a nationwide population-based study. *Ann Rheum Dis* 2013; 72(5):694-700
13. Gutiérrez-Macías A, Lizarralde-Palacios E, Martínez-Odriozola P, et al. Fatal allopurinol hypersensitivity syndrome after treatment of asymptomatic hyperuricaemia. *BMJ* 2005*;* 331: 623-4
14. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2009; 69(1):4-14
15. Meier CR, Jick H. Omeprazole, other antiulcer drugs and newly diagnosed gout. *Br J Clin Pharmacol* 1997; 44:175-8

Appendix E Cleaning of SU levels

The UK frequently measures serum urate (SU) in µmol/L however, SU level was found to be recorded inconsistently with various other units of measurement used, for example mg/dL, which is commonly used in the US, and mmol/L which is used in mainland Europe. To convert measurements to µmol/L, measurements recorded as mg/dL was multiplied by 59.48; mmol/L measurements were divided by 1000; measurements per dL were multiplied by 10 to obtain L; international units were assumed to be mmol/L (Taylor, 2013).

For each observed unit, the distribution of SU level was checked and then split into groups based on the gaps in the distribution; the gaps in the distribution indicate the incorrect units were recorded. For each group, if it was believed the wrong unit was recorded a more appropriate unit was allocated based on the SU distribution and if necessary, converted to µmol/L.

Table E1 shows how measurements were standardised. For example, 2,107 measurements did not have units recorded; 2 measurements equalled 0.04 thus assumed the units were mmol/dL; 1,195 measurements ranged between 0.1 and 0.95 thus assumed units were mmol/L, and so on.

Table E1: Standardising SU level to µmol/L

Unit of measurement recorded (in grey) Range of serum urate level	Number of measurements	Assumption of the unit of measurement and the action taken to convert to µmol/L Any other notes
Unknown	2,107	
0.04	2	Assume units were mmol/dL and multiplied by 10000.
0.1 – 0.95	1,195	Assume units were mmol/L and multiplied by 1000.
1.06 – 9.3	7	Assume units were mg/dL and multiplied by 59.48. Two measurements were removed as the patient had a more plausible measurement on the same date.
41 – 56	5	Assume units were µmol/dL and multiplied by 10.
106 – 928	897	Assume units were µmol/L.
5141	1	Cannot assume what the unit was thus measurement was removed.
IU/L (international units per litre)	3,962	
0.04 – 0.05	3	Assume units were mmol/dL and multiple by 10000.
0.15 – 0.96	2,371	Assume units were mmol/L and multiplied by 1000.
3.9 – 11.4	20	Assume units were mg/dL and multiplied by 59.48. One measurement was removed as there was a more plausible measurement 4 days later. One measurement was removed as there was a more plausible measurement one day earlier.
28 – 60	8	Assume units were µmol/dL and multiplied by 10.
107 – 946	1,560	Assume units were µmol/L.
mg/L (milligrams per litre)	1	
66.6	1	Assume units were mg/L and divide by 10 and multiplied by 59.48.
mL/min (millilitre per minute)	2	
36 – 49	2	Assume units were µmol/dL and multiplied by 10.
mmol/d (millimole per day)	1	
0.43	1	Assume units were mmol/L and multiplied by 1000
mmol/L (millimole per litre)	13,851	
0.08	1	Assume units were mmol/dL and multiplied by 10000.
0.1 – 0.97	13,473	Assume units were mmol/L and multiplied by 1000.
1.01 - 7.1	7	Assume units were mg/dL and multiplied by 59.48. One measurement was removed as there was a more plausible measurement on the same date.

43 – 53	2	Assume units were µmol/dL and multipled by 10.
153 – 774	368	Assume units were µmol/L.
mol/L (mol per litre)	57	
0.33 – 0.76	12	Assume units were mmol/L and multiplied by 1000. One measurement was removed as it was entered twice on the same date.
206 – 627	45	Assume units were µmol/L.
mosmol/L (osmoles per litre)	1	
685	1	Assume units were µmol/L.
nmol/L (nano moles per litre)	3	
0.46 – 0.47	2	Assume units were mmol/L and multiplied by 1000. One measurement was removed as it was entered twice on the same date.
369	1	Assume units were µmol/L.
pmol (picomole)	1	
358	1	Assume units were µmol/L.
pmol/L (picomole per litre)	9	
227 – 577	9	Assume units were µmol/L.
U (unit)	1	
639	1	Assume units were µmol/L.
U/L (units per litre)	1	
365	1	Assume units were µmol/L.
µg/L (microgram per litre)	1	
1	1	Cannot assume what the unit was thus remove reading.
µmol (micro mole)	20	
168 -773	20	Assume units were µmol/L.
µmol/L (micromole per litre)	15,890	
0.19 – 0.78	202	Assume units were mmol/L and multiplied by 1000.
43 – 93	5	The reading 43 was assumed, be µmol/dL and then multiplied by 10. The readings 89, 93, 80, and 87 were assumed, be µmol/L.
102 – 995	15,680	Assume units were µmol/L.
1046 – 3070	3	One measurement should be 307. Otherwise assume units were µmol/L.
µmol/min (micromole per minute)	2	

363 – 532	2	Assume units were µmol/L.
microU/L (micro unit per litre)	19	
275 – 592	19	Assume units were µmol/L.
mmol (millimole)	8	
0.37 – 0.5	6	Assume units were mmol/L and multiplied by 1000.
308 – 401	2	Assume units were µmol/L.
mmol/mmol (micromole per micromole)	1	
0.51	1	Assume units were mmol/L and multiplied by 1000.
M	1	
526	1	Assume units were µmol/L.

SU: Serum urate

Appendix F Cleaning of allopurinol prescription data

Table F1 and Table F2 describes the distribution of numeric daily dose and quantity of drugs before and after imputing missing data.

Table F1: Distribution of numeric daily dose by quantity of drugs prior to imputing missing data

	Numeric daily dose										
Quantity	0.5	1	1.5	2	2.5	3	4	5	6	Missing	Total
7	7	12,943	0	0	0	5	0	0	0	1,522	14,477
14	79	180	0	1,246	0	0	0	0	0	1,128	2,633
28	504	146,493	2	411	0	152	5	0	0	9,438	157,005
30	3	6,854	2	10	0	2	0	0	0	231	7,102
56	36	74,214	187	27,154	0	217	30	0	0	9,641	111,479
60	0	8,316	25	1,486	0	5	0	0	0	562	10,394
84	26	15,033	49	1,341	29	2,827	0	0	3	1,627	20,935
90	0	779	0	45	0	53	9	0	0	77	963
100	3	4,485	39	1,270	7	152	4	0	3	971	6,934
112	0	518	6	12,671	0	84	312	0	0	2,201	15,792
120	0	117	2	878	0	6	3	0	0	159	1,165
140	0	2	0	4	0	0	0	5	0	2	13
150	0	2	0	0	0	0	0	0	0	5	7
168	0	237	2	1,252	2	495	4	0	25	290	2,307
180	0	28	0	89	0	15	0	0	0	4	136
Missing	12	1,760	30	759	6	88	85	0	1	507	3,248
Total	670	271,961	344	48,616	44	4,101	452	5	32	28,365	354,590

Table F2: Distribution of numeric daily dose by quantity of drugs after imputing missing data

	Numeric daily dose										
Quantity	0.5	1	1.5	2	2.5	3	4	5	6	Missing	Total
7	7	13,971	0	0	0	5	0	0	0	498	14,481
14	80	196	0	2,379	0	0	0	0	0	1	2,656
28	517	153,425	2	498	0	152	5	0	0	3,861	158,460
30	3	7,046	2	12	0	2	0	0	0	101	7,166
56	36	80,137	217	27,522	0	217	30	0	0	3,948	112,107
60	0	8,862	25	1,505	0	5	0	0	0	73	10,470
84	26	16,396	48	1,397	35	2,921	0	0	3	388	21,213
90	0	859	0	46	0	53	9	0	0	2	969
100	3	5,422	39	1,452	7	152	4	0	3	225	7,307
112	0	532	6	14,401	0	84	397	0	0	536	15,956
120	0	119	2	988	0	6	3	0	0	55	1,173
140	0	2	0	6	0	0	0	5	0	0	13
150	0	2	0	5	0	0	0	0	0	0	7
168	0	238	2	1,483	2	495	4	0	26	116	2,368
180	0	29	0	93	0	15	0	0	0	0	137
Missing	0	0	0	0	0	0	0	0	0	107	107
Total	672	287,235	343	51,789	44	4,107	452	5	33	9,911	354,590

Appendix G Cleaning of BMI measurements

The process of removing implausible height, weight and body mass index (BMI) values was described below (Bhaskaran et al., 2013).

Weight measurements below 20kg and height measurements below 121cm were removed as they were deemed implausible values. Bhaskaran et al. (2013) removed height measurements above 214 cm however this cut off was increased in this thesis to 219cm as one patient had a height of 218cm which was recorded multiple times thus appeared to be plausible.

BMI was then calculated for patients who had height and weight recorded on the same day. If patients had different values of BMI on one day, it was checked if the BMI values were plausible by comparing it to previous recordings; if a decision could not be made on which BMI was plausible, a value was chosen at random.

For the remaining weight measurements that had no height recorded on the same day, if weight was recorded multiple times on the same day, then it was checked if the weight measurements were plausible and if it could not be decided which measurement was the most plausible then one was picked at random. The process was repeated for the height measurements.

BMI was then calculated using the most recent height record prior to recording of weight. For the remaining weight records, BMI was calculated using height that was recorded afterwards. Next, BMI that was automatically generated in Vision software was used.

Finally, BMI under the value of 5 and above 200 was removed.

Appendix H Distribution of baseline covariates

Table H1: Distribution of baseline covariates of eligible patients for each outcome analysis

Baseline covariates	Whole study sample*	Target SU level	Gout hospitalisation	Joint replacement
Number of patients	**16,876**	**1,742**	**14,087**	**16,644**
Demographics				
Age (Mean (SD)) 18-101	62.1 (14.7)	59.4 (14.3)	63.6 (14.3)	61.9 (14.8)
Sex: Female	3,881 (23)	361 (21)	3,475 (25)	3,796 (23)
Deprivation (Mean (SD))	9.1 (5.5)	9.0 (5.6)	9.3 (5.5)	9.1 (5.5)
Comorbidities				
Anxiety	672 (4)	58 (3)	586 (4)	663 (4)
Depression	842 (5)	84 (5)	749 (5)	826 (5)
Cerebrovascular disease	407 (2)	40 (2)	1,936 (14)	2,136 (13)
Coronary heart disease	2,167 (13)	236 (14)	369 (3)	399 (2)
Diabetes	1,047 (6)	107 (6)	947 (7)	1,024 (6)
Hyperlipidaemia	783 (5)	83 (5)	687 (5)	768 (5)
Hypertension	3,137 (19)	371 (21)	2,758 (20)	3,064 (18)
Osteoarthritis	1,106 (7)	129 (7)	1,011 (7)	1,011 (6)
Peripheral vascular disease	257 (2)	22 (1)	240 (2)	252 (2)
Renal disease	217 (1)	28 (1)	190 (1)	210 (1)
Lifestyle factors				
Alcohol consumption				
Ever drinker	9,488 (56)	1,048 (60)	7,993 (57)	9,343 (56)
Never drinker	856 (5)	61 (4)	777 (6)	841 (5)
Missing	6,532 (39)	633 (36)	5,317 (38)	6,460 (39)
Body mass index				
Normal weight	2,517 (15)	183 (11)	2,171 (15)	2,487 (15)
Overweight	4,933 (29)	535 (31)	4,185 (30)	4,861 (29)
Obese	3,219 (19)	416 (24)	2,744 (19)	3,160 (19)
Missing	6,207 (37)	608 (35)	4,987 (35)	6,136 (37)
Smoking status				
Ever smoker	6,436 (38)	773 (44)	5,532 (39)	6,349 (38)
Never smoker	4,847 (29)	423 (24)	4,038 (29)	4,767 (29)
Missing	5,593 (33)	546 (31)	4,517 (32)	5,528 (33)
SU level				
≤360µmol/L	951 (6)	N/A	814 (6)	937 (6)
>360µmol/L	6,062 (36)	497.6 (74.3)[a]	5,023 (36)	5,977 (36)
Missing	9,863 (58)	0 (0.00)	8,250 (59)	9,730 (58)
Medication use				
Analgesics	5,578 (33)	603 (65)	5,017 (36)	5,402 (32)
Colchicine	389 (2)	70 (4)	343 (2)	387 (2)
Diuretics	6,142 (36)	638 (37)	5,479 (39)	5,998 (36)
NSAIDS	8,024 (48)	1,078 (62)	6,824 (48)	7,843 (47)

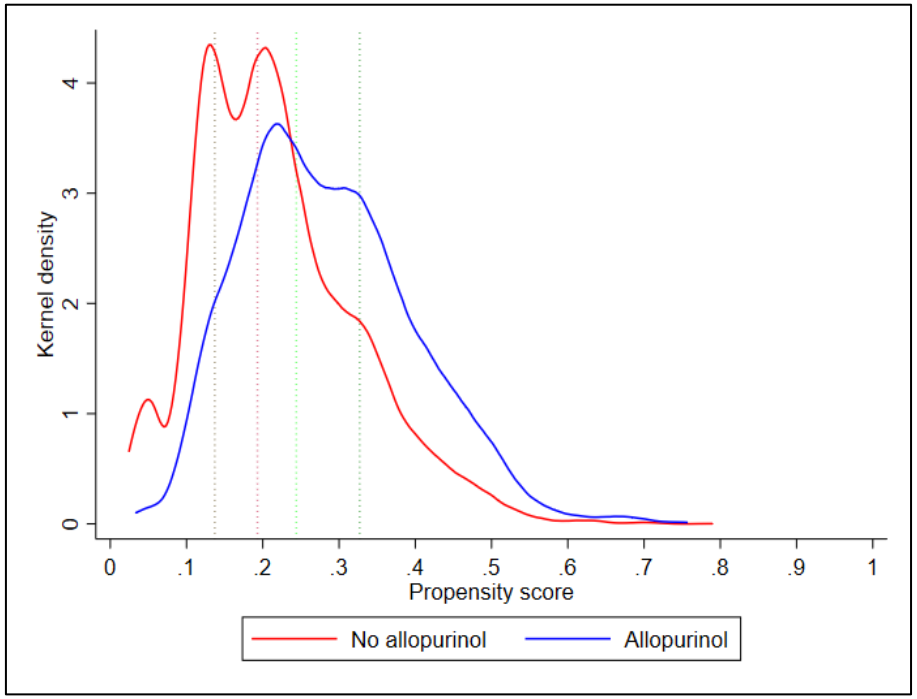
*The study sample was eligible for analysis of outcomes mortality and repeated gout consultations; [a]Mean (SD) presented and ranged from 361 to 905; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate.

Baseline covariates	Cerebrovascular disease	Coronary heart disease	Peripheral vascular disease	Renal disease
Number of patients	**16,253**	**14,063**	**16,519**	**16,508**
Demographics				
Age (Mean (SD))	61.6 (14.7)	60.5 (14.9)	61.9 (14.8)	61.9 (14.7)
Sex: Female	3,687 (23)	3,137 (22)	3,784 (23)	3,752 (23)
Deprivation (Mean (SD))	9.1 (5.5)	9.0 (5.5)	9.1 (5.5)	9.1 (5.5)
Comorbidities				
Anxiety	634 (4)	540 (4)	661 (4)	662 (4)
Depression	788 (5)	678 (5)	823 (5)	825 (5)
Cerebrovascular disease	[redacted]	271 (2)	2,064 (12)	2,064 (13)
Coronary heart disease	2,015 (12)	[redacted]	385 (2)	396 (2)
Diabetes	972 (6)	742 (5)	986 (6)	967 (6)
Hyperlipidaemia	732 (5)	483 (3)	749 (5)	744 (5)
Hypertension	2,939 (18)	2,281 (16)	3,039 (18)	3,027 (18)
Osteoarthritis	1,057 (7)	867 (6)	1,059 (6)	1,068 (6)
Peripheral vascular disease	231 (1)	160 (1)	[redacted]	233 (1)
Renal disease	204 (1)	143 (1)	197 (1)	[redacted]
Lifestyle factors				
Alcohol consumption				
Ever drinker	9,140 (56)	7,662 (54)	9,278 (56)	9,284 (56)
Never drinker	813 (5)	642 (5)	828 (5)	829 (5)
Missing	6,300 (39)	5,759 (41)	6,413 (39)	6,395 (39)
Body mass index				
Normal weight	2,402 (15)	2,058 (15)	2,444 (15)	2,455 (15)
Overweight	4,751 (29)	3,928 (28)	4,813 (29)	4,825 (29)
Obese	3,113 (19)	2,541 (18)	3,169 (19)	3,129 (19)
Missing	5,987 (37)	5,536 (39)	6,093 (37)	6,099 (37)
Smoking status				
Ever smoker	6,170 (38)	5,079 (36)	6,219 (38)	6,281 (38)
Never smoker	4,669 (29)	4,029 (29)	4,786 (29)	4,729 (29)
Missing	5,414 (33)	4,955 (35)	5,514 (33)	5,498 (33)
SU level				
≤360µmol/L	915 (6)	830 (6)	931 (6)	946 (6)
>360µmol/L	5,843 (36)	4,916 (35)	5,917 (36)	5,894 (36)
Missing	9,495 (58)	8,317 (59)	9,671 (59)	9,668 (59)
Medication use				
Analgesics	5,259 (32)	4,260 (30)	5,366 (32)	5,371 (33)
Colchicine	368 (2)	295 (2)	376 (2)	367 (2)
Diuretics	5,725 (35)	4,382 (31)	5,938 (36)	5,855 (35)
NSAIDS	7,749 (48)	6,661 (47)	7,840 (47)	7,854 (48)

The whole study sample was eligible for analysis of outcomes mortality and repeated gout consultations; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate.

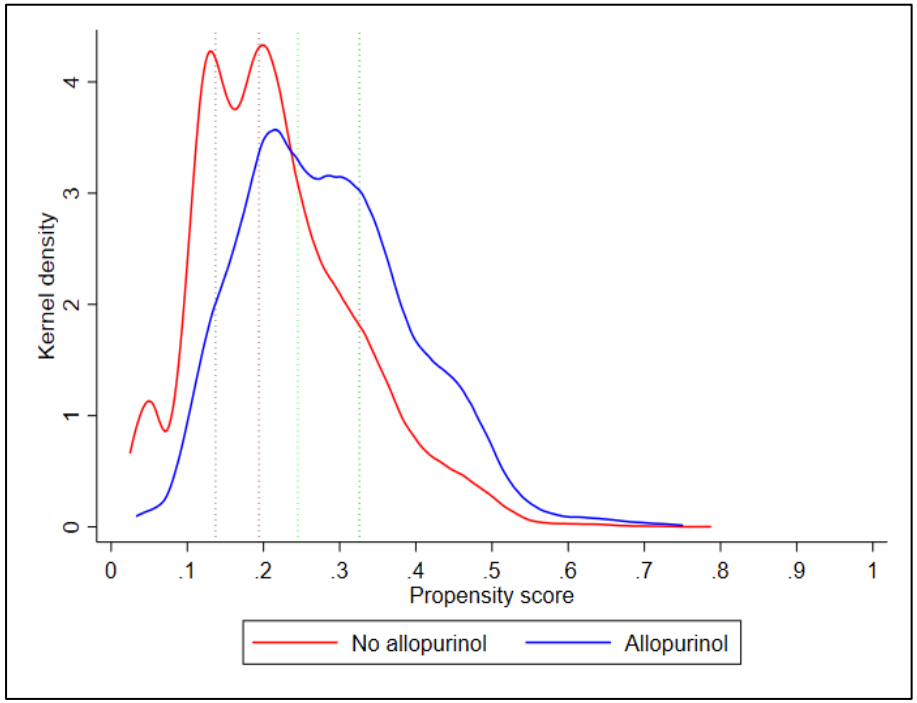
Appendix I Distribution of baseline PS and imbalanced covariates within subclasses

Figure I1: Mortality



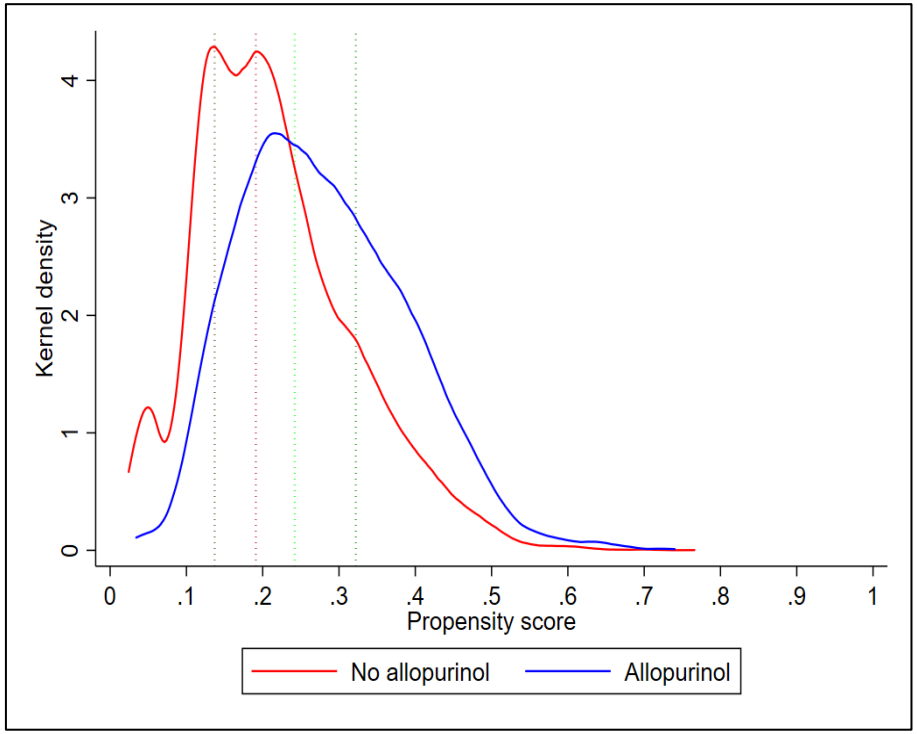
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, BMI, and NSAIDS
S2: Sex, deprivation, BMI, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I2: Gout consultation



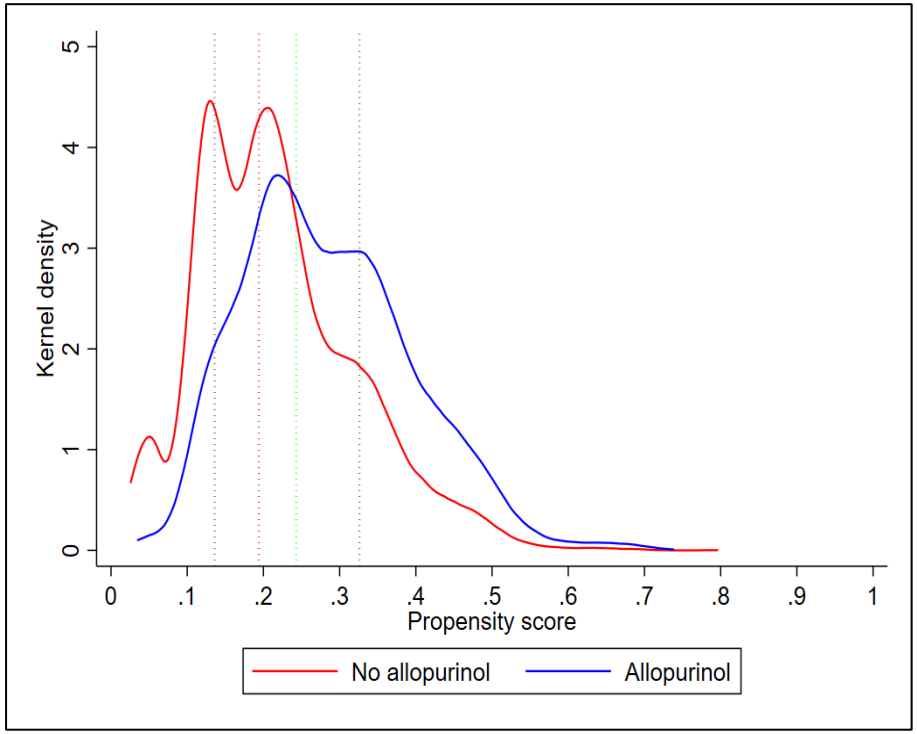
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, BMI, and SU level
S2: Deprivation
S3: N/A
S4: N/A
S5: N/A

Figure I3: Gout hospitalisation



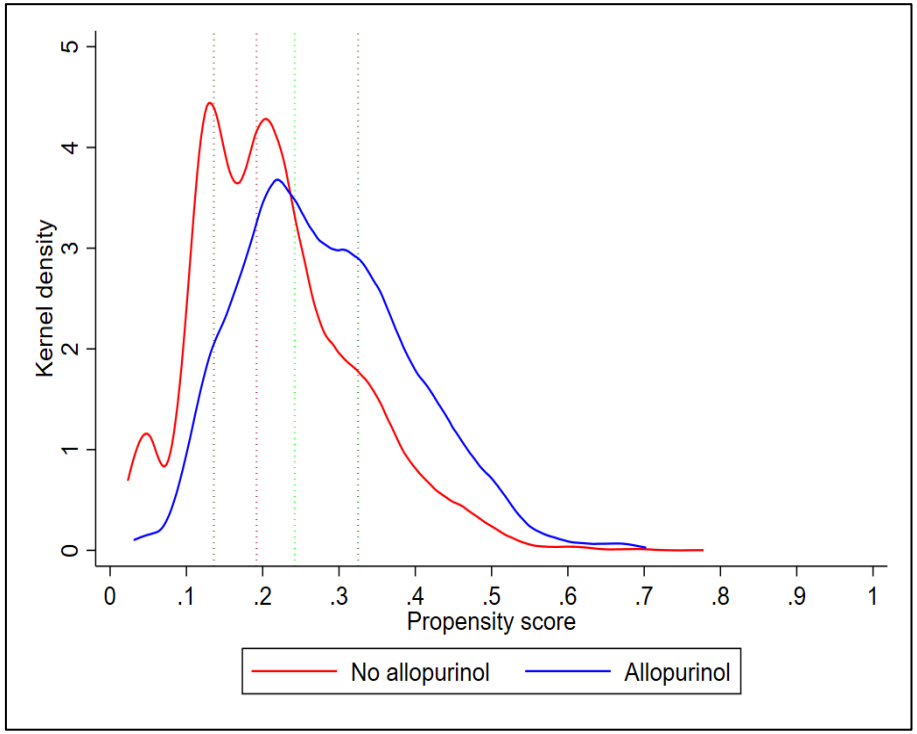
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, BMI, and SU level
S2: Sex, BMI, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I4: Joint replacement



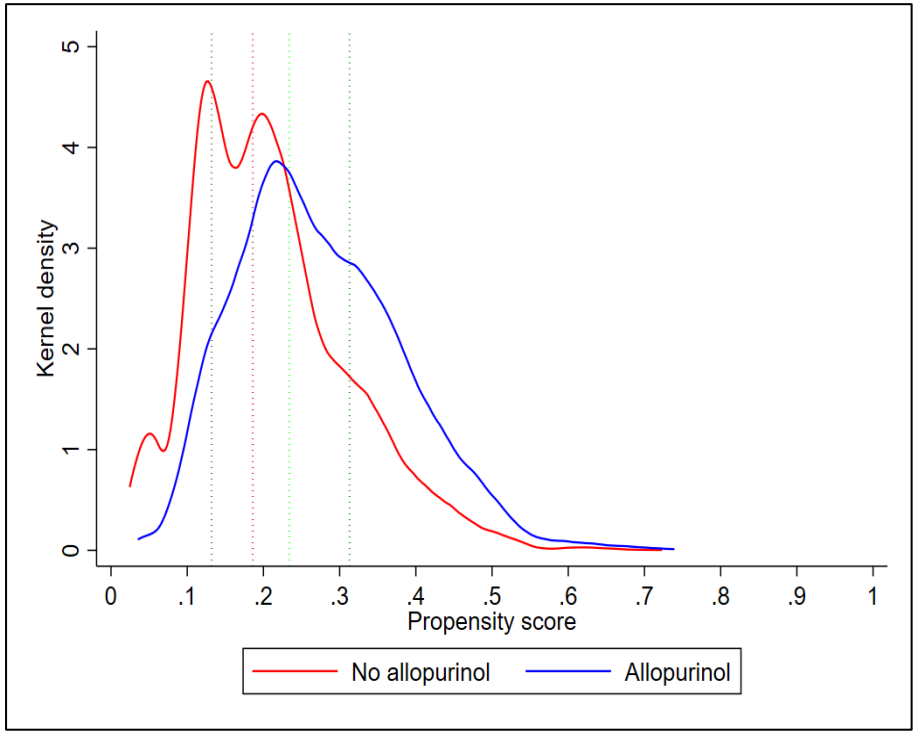
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, and SU level
S2: Deprivation, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I5: Cerebrovascular disease



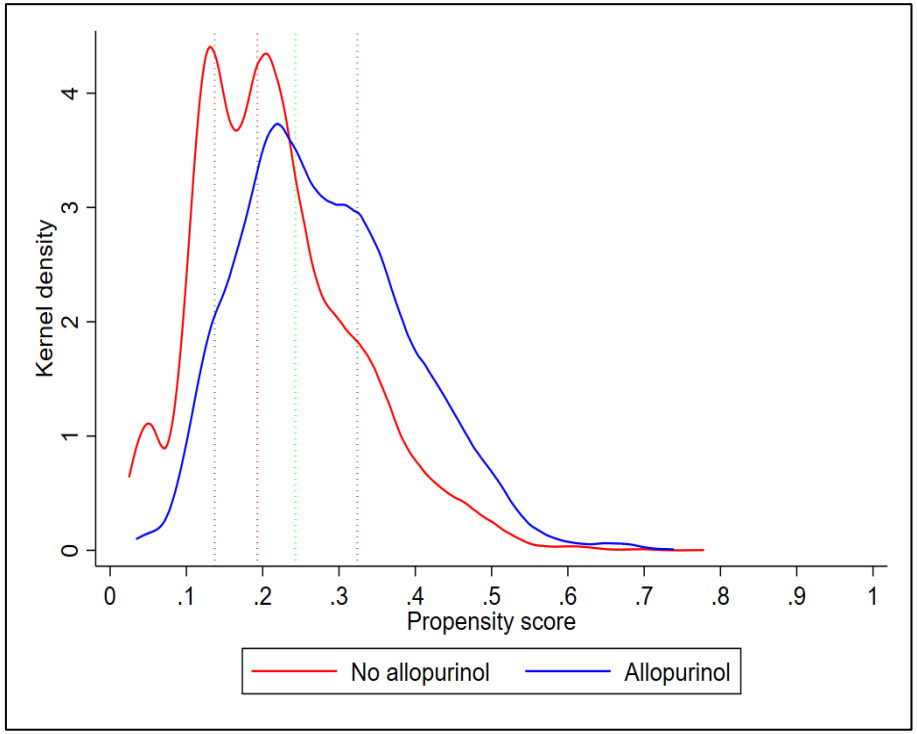
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, smoking status, and SU level
S2: Deprivation, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I6: Coronary heart disease



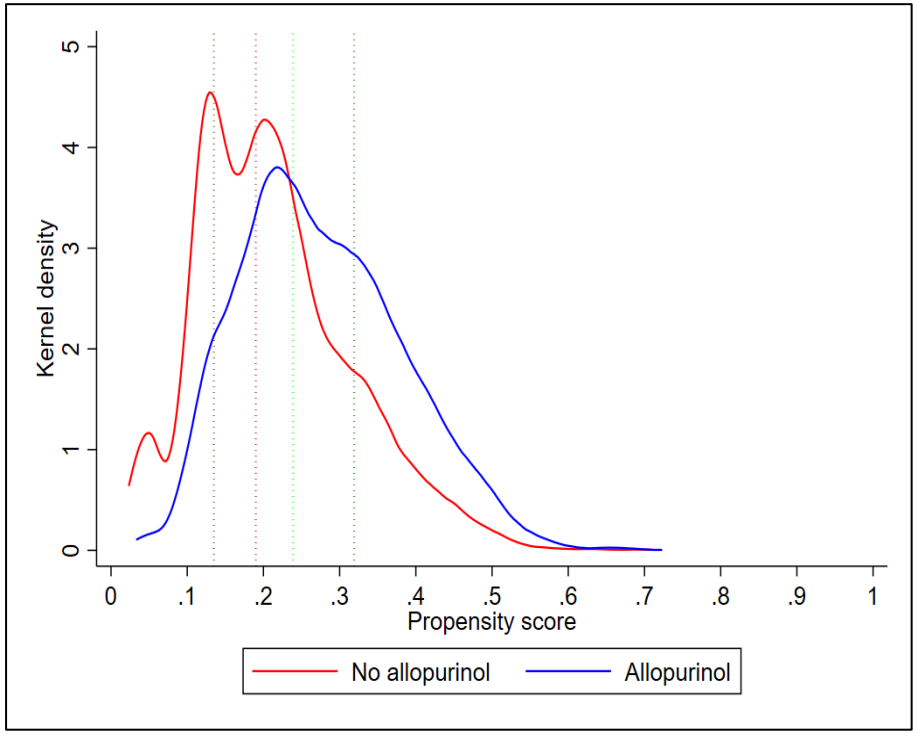
Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, cerebrovascular disease, BMI, SU level, and colchicine
S2: BMI, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I7: Peripheral vascular disease



Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, BMI, and SU level
S2: Sex, deprivation, and NSAIDS
S3: N/A
S4: N/A
S5: N/A

Figure I8: Renal disease



Dotted lines indicate at which propensity scores subclasses (S) were created.
Standardised mean difference >0.10 was observed in covariates;
S1: Sex, alcohol consumption, BMI, SU level, and colchicine
S2: Sex
S3: SU level
S4: N/A
S5: N/A

Appendix J Assessment for proportional hazards

Figure J1: Target serum urate level (unadjusted Cox model)

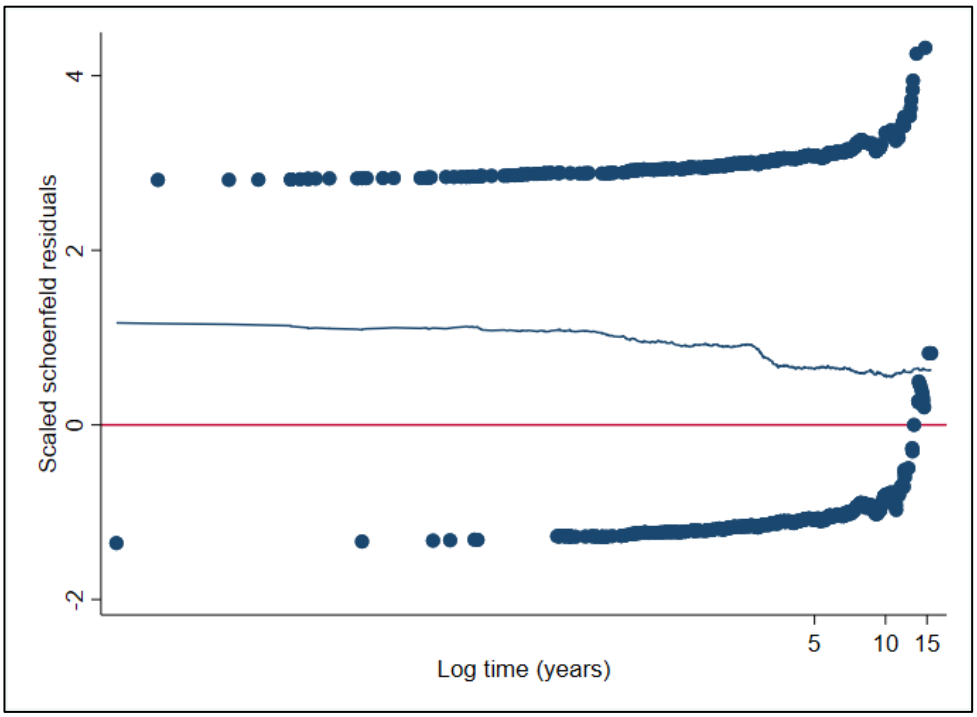


Figure J2: Repeated gout consultations (unadjusted Cox model)

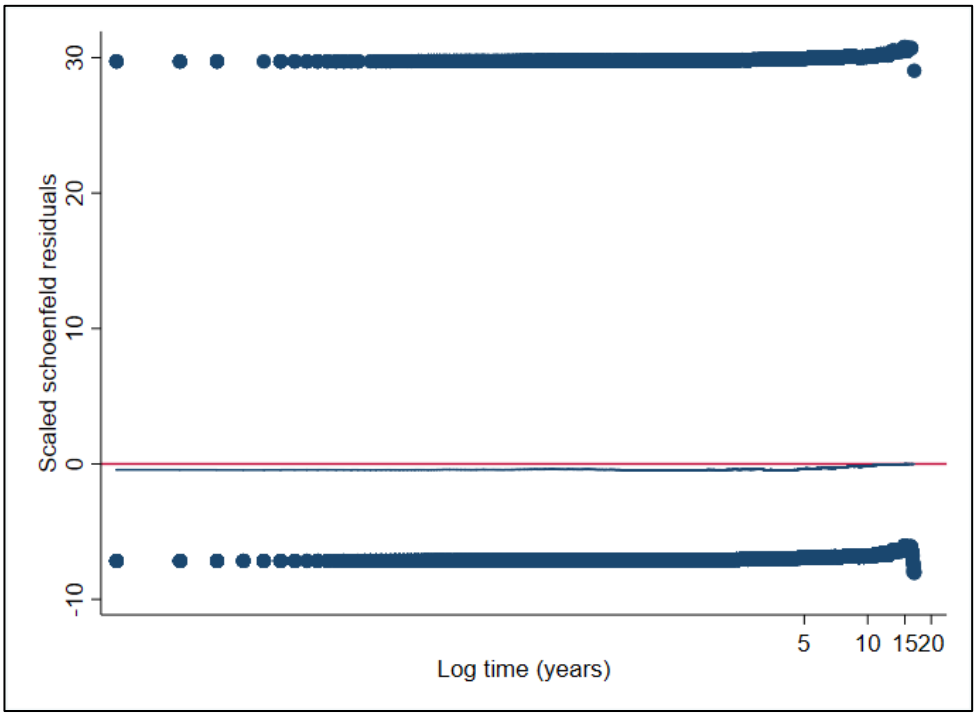


Figure J3: Renal disease (unadjusted Cox model)

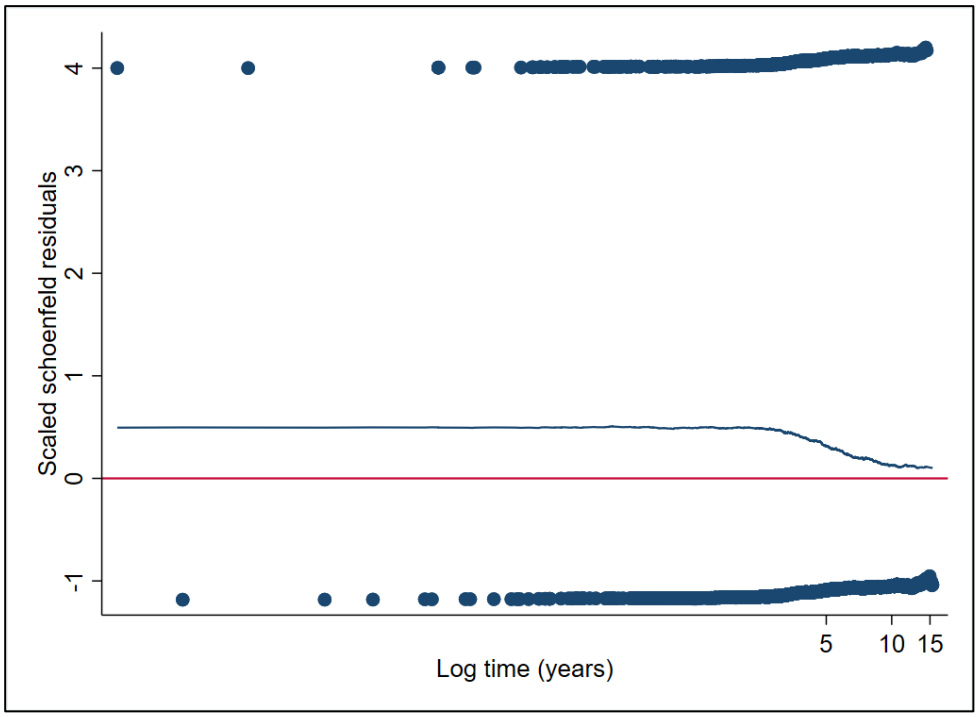


Figure J4: Mortality (PS subclass 1)

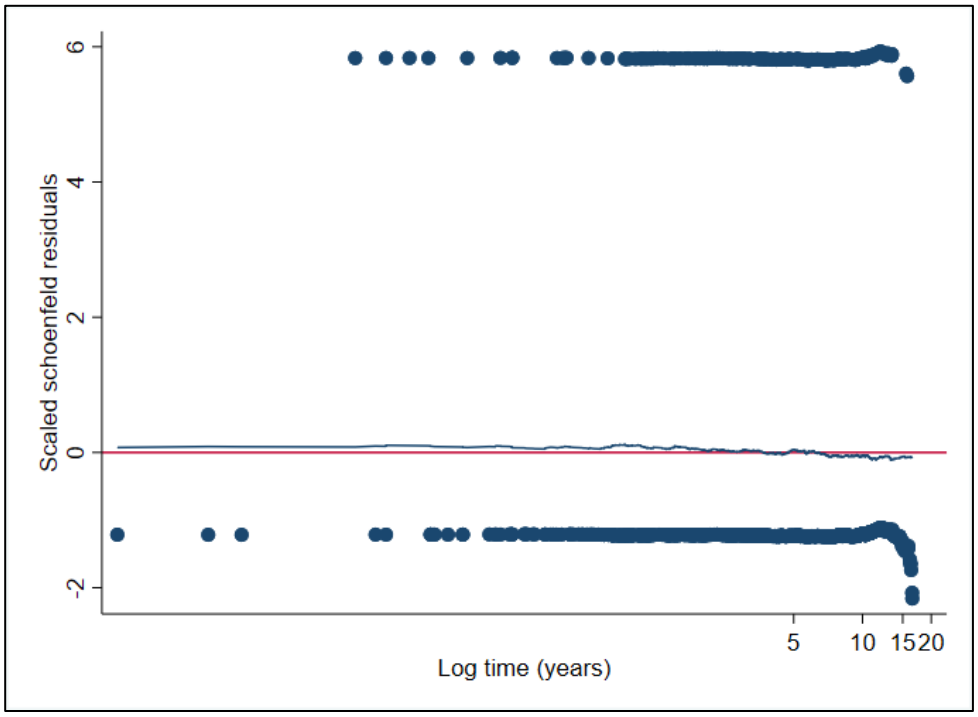
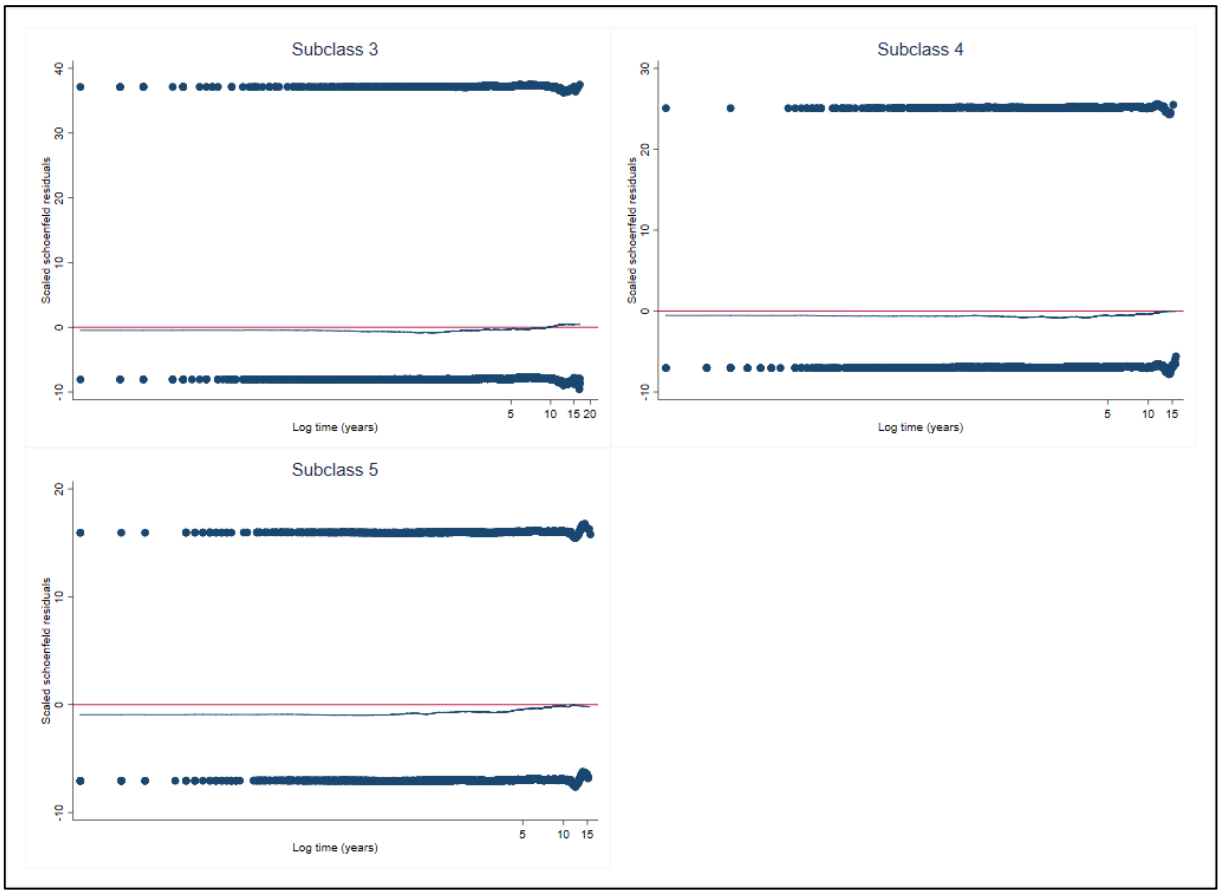


Figure J5: Gout consultations (PS subclasses 3-5)



Appendix K Distribution of covariates over time

Table K1: Distribution of covariates in each year of follow-up

Follow-up year	Age (Mean (SD))	Sex: Female	Deprivation (Mean (SD))	Anxiety	Depression	Cerebrovascular disease	Coronary heart disease	Diabetes	Gout consultation
0 (N=16,876)	62.1 (14.7)	3,881 (23)	9.1 (5.5)	672 (4)	842 (5)	407 (2)	2,167 (13)	1,047 (6)	0 (0)
1 (N=15,873)	62.6 (14.5)	3,608 (23)	9.1 (5.5)	862 (5)	1,106 (7)	550 (3)	2,569 (16)	1,239 (8)	5,541 (35)
2 (N=14,888)	63.2 (14.3)	3,314 (22)	9.1 (5.5)	1,005 (7)	1,259 (8)	622 (4)	2,805 (19)	1,369 (9)	2,571 (17)
3 (N=14,033)	63.7 (14.1)	3,068 (22)	9.1 (5.5)	1,137 (8)	1,384 (10)	677 (5)	2,950 (21)	1,531 (11)	2,342 (17)
4 (N=13,207)	64.2 (13.9)	2,837 (21)	9.0 (5.5)	1,216 (9)	1,473 (11)	730 (6)	3,025 (23)	1,659 (13)	2,159 (16)
5 (N=12,391)	64.7 (13.7)	2,602 (21)	9.0 (5.5)	1,274 (10)	1,547 (12)	734 (6)	3,019 (24)	1,728 (14)	1,853 (15)
6 (N=11,657)	65.3 (13.6)	2,404 (21)	9.0 (5.5)	1,303 (11)	1,598 (14)	744 (6)	3,003 (26)	1,788 (15)	1,688 (14)
7 (N=10,777)	65.8 (13.4)	2,173 (20)	9.0 (5.5)	1,294 (12)	1,577 (15)	715 (7)	2,852 (26)	1,815 (17)	1,499 (14)
8 (N=9,933)	66.3 (13.2)	1,965 (20)	8.9 (5.5)	1,281 (13)	1,578 (16)	692 (7)	2,720 (27)	1,849 (19)	1,307 (13)
9 (N=9,128)	66.8 (13.0)	1,764 (19)	8.9 (5.5)	1,247 (14)	1,557 (17)	688 (8)	2,560 (28)	1,836 (20)	1,177 (13)
10 (N=8,126)	67.3 (12.7)	1,521 (19)	8.9 (5.6)	1,177 (14)	1,462 (18)	645 (8)	2,316 (29)	1,743 (21)	960 (12)
11 (N=6,918)	67.8 (12.4)	1,255 (18)	8.8 (5.6)	1,055 (15)	1,279 (18)	561 (8)	1,985 (29)	1,552 (22)	824 (12)
12 (N=4,764)	68.0 (12.2)	812 (17)	9.0 (5.6)	767 (16)	889 (19)	386 (8)	1,412 (30)	1,116 (23)	550 (12)
13 (N=3,159)	68.3 (11.9)	517 (16)	9.1 (5.6)	539 (17)	598 (19)	260 (8)	918 (29)	758 (24)	346 (11)
14 (N=2,019)	68.9 (11.6)	302 (15)	9.1 (5.6)	342 (17)	405 (20)	179 (9)	591 (29)	489 (24)	218 (11)
15 (N=1,107)	69.4 (11.5)	165 (15)	8.9 (5.6)	193 (17)	229 (21)	101 (9)	321 (29)	258 (23)	93 (8)
16 (N=475)	69.8 (11.4)	66 (14)	9.0 (5.7)	96 (20)	105 (22)	38 (8)	135 (28)	117 (25)	42 (9)

						Alcohol consumption		
Follow-up year	**Hyperlipidaemia**	**Hypertension**	**Osteoarthritis**	**Peripheral vascular disease**	**Renal disease**	**Ever-drinker**	**Never drinker**	**Missing**
0 (N=16,876)	783 (5)	3,137 (19)	1,106 (7)	257 (2)	217 (1)	9,488 (56)	856 (5)	6,532 (39)
1 (N=15,873)	1,171 (7)	3,922 (25)	1,584 (10)	315 (2)	310 (2)	9,882 (62)	784 (5)	5,207 (33)
2 (N=14,888)	1,493 (10)	4,420 (30)	1,876 (13)	348 (2)	359 (2)	9,959 (67)	728 (5)	4,201 (28)
3 (N=14,033)	1,765 (13)	4,802 (34)	2,090 (15)	379 (3)	404 (3)	9,984 (71)	677 (5)	3,372 (24)
4 (N=13,207)	2,020 (15)	5,077 (38)	2,253 (17)	377 (3)	550 (4)	9,892 (75)	608 (5)	2,707 (20)
5 (N=12,391)	2,248 (18)	5,275 (43)	2,360 (19)	377 (3)	970 (8)	9,634 (78)	560 (5)	2,197 (18)
6 (N=11,657)	2,403 (21)	5,372 (46)	2,459 (21)	390 (3)	1,334 (11)	9,347 (80)	506 (4)	1,804 (15)
7 (N=10,777)	2,475 (23)	5,263 (49)	2,470 (23)	366 (3)	1,559 (14)	8,850 (82)	462 (4)	1,465 (14)
8 (N=9,933)	2,528 (25)	5,107 (51)	2,417 (24)	364 (4)	1,756 (18)	8,329 (84)	409 (4)	1,195 (12)
9 (N=9,128)	2,499 (27)	4,933 (54)	2,333 (26)	338 (4)	1,886 (21)	7,762 (85)	373 (4)	993 (11)
10 (N=8,126)	2,364 (29)	4,561 (56)	2,180 (27)	319 (4)	1,858 (23)	7,027 (86)	319 (4)	780 (10)
11 (N=6,918)	2,109 (30)	3,997 (58)	1,930 (28)	268 (4)	1,633 (24)	6,081 (88)	257 (4)	580 (8)
12 (N=4,764)	1,507 (32)	2,813 (59)	1,373 (29)	170 (4)	1,130 (24)	4,258 (89)	180 (4)	326 (7)
13 (N=3,159)	1,023 (32)	1,893 (60)	940 (30)	117 (4)	749 (24)	2,858 (90)	116 (4)	185 (6)
14 (N=2,019)	662 (33)	1,227 (61)	629 (31)	72 (4)	480 (24)	1,847 (91)	72 (4)	100 (5)
15 (N=1,107)	352 (32)	665 (60)	360 (33)	40 (4)	264 (24)	1,032 (93)	34 (3)	41 (4)
16 (N=475)	150 (32)	285 (60)	163 (34)	16 (3)	122 (26)	449 (95)	14 (3)	12 (3)

	Body mass index				Smoking status		
Follow-up years	**Normal weight**	**Overweight**	**Obese**	**Missing**	**Ever smoker**	**Never smoker**	**Missing**
0 (N=16,876)	2,517 (15)	4,933 (29)	3,219 (19)	6,207 (37)	6,436 (38)	4,847 (29)	5,593 (33)
1 (N=15,873)	2,482 (16)	4,967 (31)	3,364 (21)	5,060 (32)	7,047 (44)	4,479 (28)	4,347 (27)
2 (N=14,888)	2,427 (16)	4,914 (33)	3,416 (23)	4,131 (28)	7,424 (50)	4,236 (28)	3,228 (22)
3 (N=14,033)	2,383 (17)	4,813 (34)	3,495 (25)	3,342 (24)	7,747 (55)	4,050 (29)	2,236 (16)
4 (N=13,207)	2,290 (17)	4,670 (35)	3,541 (27)	2,706 (20)	7,894 (60)	3,766 (29)	1,547 (12)
5 (N=12,391)	2,166 (17)	4,528 (37)	3,560 (29)	2,137 (17)	7,827 (63)	3,538 (29)	1,026 (8)
6 (N=11,657)	2,089 (18)	4,338 (37)	3,565 (31)	1,665 (14)	7,667 (66)	3,371 (29)	619 (5)
7 (N=10,777)	1,942 (18)	4,104 (38)	3,426 (32)	1,305 (12)	7,304 (68)	3,113 (29)	360 (3)
8 (N=9,933)	1,797 (18)	3,818 (38)	3,291 (33)	1,027 (10)	6,851 (69)	2,856 (29)	226 (2)
9 (N=9,128)	1,633 (18)	3,529 (39)	3,125 (34)	841 (9)	6,378 (70)	2,599 (28)	151 (2)
10 (N=8,126)	1,478 (18)	3,130 (39)	2,874 (35)	644 (8)	5,719 (70)	2,313 (28)	94 (1)
11 (N=6,918)	1,252 (18)	2,676 (39)	2,527 (37)	463 (7)	4,917 (71)	1,946 (28)	55 (1)
12 (N=4,764)	891 (19)	1,858 (39)	1,751 (37)	264 (6)	3,425 (72)	1,310 (27)	29 (1)
13 (N=3,159)	576 (18)	1,236 (39)	1,183 (37)	164 (5)	2,270 (72)	874 (28)	15 (0)
14 (N=2,019)	375 (19)	795 (39)	764 (38)	85 (4)	1,461 (72)	550 (27)	8 (0)
15 (N=1,107)	213 (19)	423 (38)	428 (39)	43 (4)	805 (73)	298 (27)	4 (0)
16 (N=475)	91 (19)	179 (38)	192 (40)	13 (3)	350 (74)	124 (26)	1 (0)

	SU level						
Follow-up years (N)	**≤360µmol/L**	**>360µmol/L**	**Missing**	**Analgesics**	**Colchicine**	**Diuretics**	**NSAIDS**
0 (N=16,876)	951 (6)	6,062 (36)	9,863 (58)	5,578 (33)	3,881 (23)	6,142 (36)	8,024 (48)
1 (N=15,873)	910 (6)	5,734 (36)	9,229 (58)	4,908 (31)	3,608 (23)	4,874 (31)	12,340 (78)
2 (N=14,888)	879 (6)	5,372 (36)	8,637 (58)	4,207 (28)	3,314 (22)	4,286 (29)	6,178 (41)
3 (N=14,033)	844 (6)	5,066 (36)	8,123 (58)	4,026 (29)	3,068 (22)	4,103 (29)	5,849 (42)
4 (N=13,207)	808 (6)	4,786 (36)	7,613 (58)	3,931 (30)	2,837 (21)	3,897 (30)	5,471 (41)
5 (N=12,391)	763 (6)	4,514 (36)	7,114 (57)	3,798 (31)	2,602 (21)	3,625 (29)	4,975 (40)
6 (N=11,657)	710 (6)	4,250 (36)	6,697 (57)	3,684 (32)	2,404 (21)	3,387 (29)	4,557 (39)
7 (N=10,777)	659 (6)	3,913 (36)	6,205 (58)	3,461 (32)	2,173 (20)	3,050 (28)	4,059 (38)
8 (N=9,933)	605 (6)	3,577 (36)	5,751 (58)	3,252 (33)	1,965 (20)	2,757 (28)	3,544 (36)
9 (N=9,128)	542 (6)	3,295 (36)	5,291 (58)	3,109 (34)	1,764 (19)	2,428 (27)	3,167 (35)
10 (N=8,126)	484 (6)	2,885 (36)	4,757 (59)	2,729 (34)	1,521 (19)	2,098 (26)	2,710 (33)
11 (N=6,918)	412 (6)	2,401 (35)	4,105 (59)	2,367 (34)	1,255 (18)	1,743 (25)	2,261 (33)
12 (N=4,764)	252 (5)	1,469 (31)	3,043 (64)	1,641 (34)	812 (17)	1,161 (24)	1,540 (32)
13 (N=3,159)	158 (5)	862 (27)	2,139 (68)	1,103 (35)	517 (16)	779 (25)	1,009 (32)
14 (N=2,019)	87 (4)	497 (25)	1,435 (71)	692 (34)	302 (15)	484 (24)	630 (31)
15 (N=1,107)	31 (3)	226 (20)	850 (77)	356 (32)	165 (15)	249 (22)	343 (31)
16 (N=475)	10 (2)	87 (18)	378 (80)	151 (32)	66 (14)	98 (21)	136 (29)

N (%) presented unless otherwise stated; NSAIDS: Non-steroidal anti-inflammatory drugs; SD: Standard deviation; SU: Serum urate

Appendix L Time-varying PS model specification and distribution

Table L1: PS model specification for mortality analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=109,684	Allopurinol N=45,647	Overall N=155,331
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.003 (0.002, 0.06) ($8*10^{-5}$, 0.99)	0.93 (0.66, 0.99) (0.01, 0.99)	76,177 (69%)	702 (2%)	76,879 (49%)
2	Main effects model + FP1 terms for deprivation$^{(0)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(3)}$ + linear term for age	0.01 (0.01, 0.06) ($3*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.01, 0.99)	48,406 (44%)	1,132 (2%)	49,538 (32%)
3	Main effects model + FP2 terms for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$, cumulative allopurinol use$^{(1, 2)}$, follow-up time$^{(3, 3)}$	0.01 (0.004, 0.06) ($2*10^{-4}$, 0.99)	0.93 (0.70, 0.98) (0.01, 0.99)	61,767 (56%)	6,797 (15%)	68,564 (44%)
4	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time	0.01 (0.01, 0.06) ($2*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.01, 0.99)	28,739 (26%)	398 (1%)	29,137 (19%)
5	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time + SU level*NSAIDS	0.01 (0.01, 0.06) ($2*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.01, 0.99)	22,520 (21%)	395 (1%)	22,915 (15%)
6	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time + SU level*NSAIDS + SU level*hypertension	0.01 (0.01, 0.06) ($2*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.01, 0.99)	16,696 (15%)	417 (1%)	17,113 (11%)

7	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time + SU level*NSAIDS + SU level*hypertension + diuretics*sex	0.01 (0.01, 0.06) ($1*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.004, 0.99)	13,152 (12%)	457 (1%)	13,609 (9%)
8	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time + SU level*NSAIDS + SU level*hypertension + diuretics*sex + alcohol consumption*sex	0.01 (0.01, 0.06) ($2*10^{-4}$, 0.99)	0.92 (0.68, 0.98) (0.003, 0.99)	10,584 (10%)	683 (1%)	11,267 (7%)
9	Main effects model + FP2 term for age$^{(3, 3)}$, deprivation$^{(-2, -1)}$ + FP1 term for cumulative allopurinol use$^{(0.5)}$ + linear terms for follow-up time + SU level*NSAIDS + SU level*hypertension + diuretics*sex + alcohol consumption*sex + analgesic*sex	0.01 (0.01, 0.06) ($1*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.003, 0.99)	10,217 (9%)	436 (1%)	10,653 (7%)

Specification highlighted in green is the chosen PS model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score; SU: Serum urate

Figure L1: Comparison of PS distribution for mortality analysis

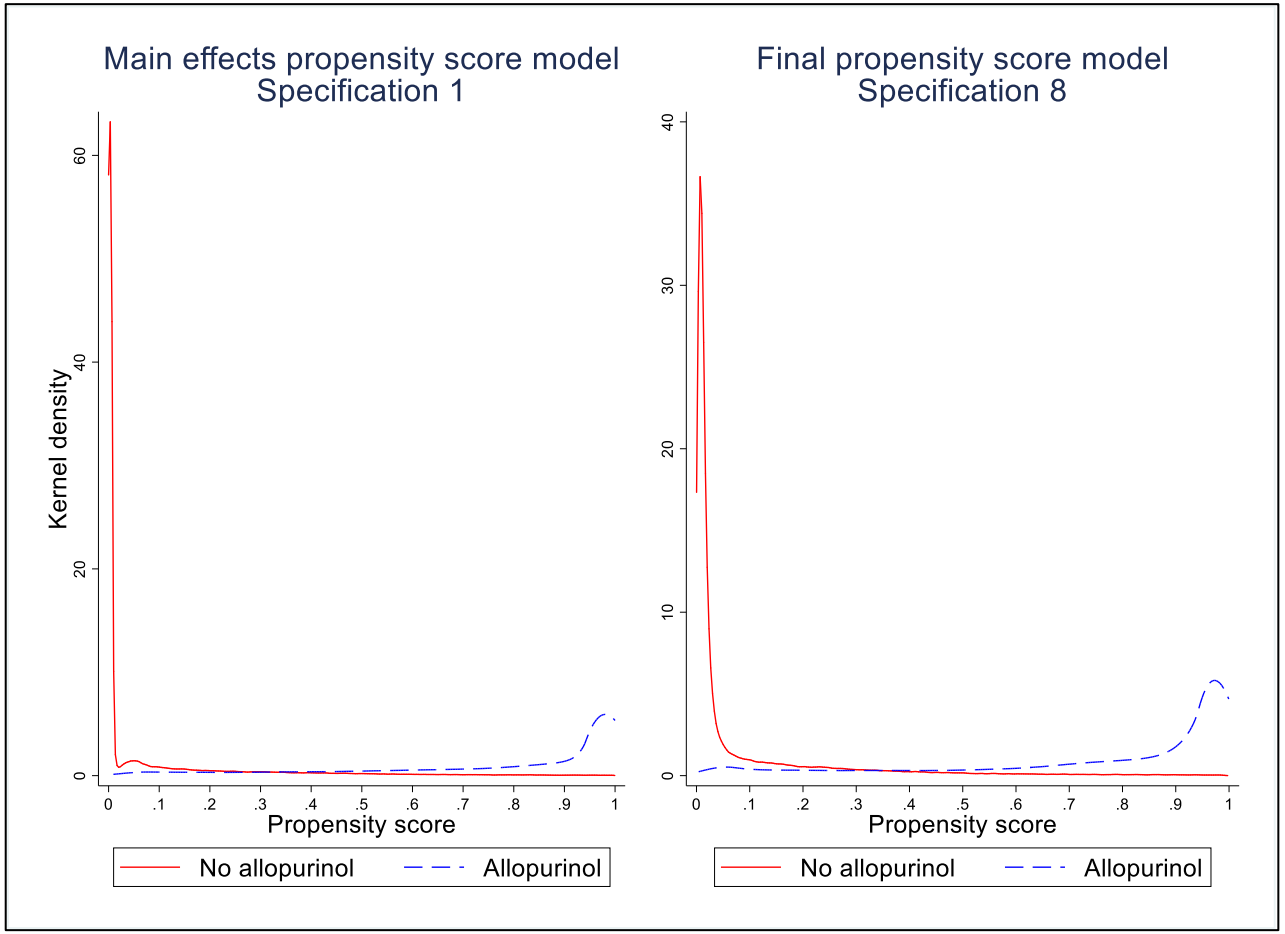


Table L2: PS model specification for gout hospitalisation analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=88,617	Allopurinol N=33,451	Overall N=122,068
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.002 (0.001, 0.04) ($9*10^{-5}$, 0.99)	0.94 (0.67, 0.99) (0.01, 0.99)	63,658 (72%)	2,026 (6%)	65,684 (54%)
2	Main effects model + FP1 terms for age$^{(3)}$, deprivation$^{(3)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(0.5)}$	0.01 (0.01, 0.05) ($3*10^{-4}$, 0.99)	0.92 (0.67, 0.98) (0.01, 0.99)	26,608 (30%)	2,671 (8%)	29,279 (24%)
3	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation$^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$, follow-up time$^{(3, 3)}$	0.01 (0.01, 0.04) ($3*10^{-4}$, 0.99)	0.92 (0.71, 0.98) (0.01, 0.99)	23,549 (27%)	3,301 (10%)	26,850 (22%)
4	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation $^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$ + FP1 term for follow-up time$^{(0.5)}$	0.01 (0.01, 0.04) ($20*10^{-4}$, 0.98)	0.92 (0.69, 0.98) (0.003, 0.99)	17,771 (20%)	5,304 (16%)	23,075 (19%)
5	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation $^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$ + FP1 term for follow-up time$^{(0.5)}$ + SU level*NSAIDS	0.01 (0.01, 0.04) ($2*10^{-4}$, 0.98)	0.92 (0.69, 0.98) (0.003, 0.99)	9,541 (11%)	5,203 (16%)	14,744 (12%)
6	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation$^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$ + FP1 term for follow-up time$^{(0.5)}$ + SU level*NSAIDS + renal disease*gout consultation	0.01 (0.01, 0.04) ($20*10^{-4}$, 0.99)	0.92 (0.69, 0.98) (0.003, 0.99)	9,076 (10%)	2,875 (9%)	11,951 (10%)
7	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation$^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$ + FP1 term for follow-up time$^{(0.5)}$ + SU level*NSAIDS + renal disease*gout consultation + diuretics*sex	0.01 (0.01, 0.04) ($2*10^{-4}$, 0.99)	0.92 (0.69, 0.98) (0.003, 0.99)	7,048 (8%)	2,409 (7%)	9,457 (8%)

8	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation$^{(3, 3)}$, cumulative allopurinol use$^{(-0.5, 0)}$ + FP1 term for follow-up time$^{(0.5)}$ + SU level*NSAIDS + renal disease*gout consultation + diuretics*sex + diuretics*hyperlipidaemia	0.01 (0.01, 0.04) ($2*10^{-4}$, 0.99)	0.92 (0.69, 0.98) (0.002, 0.99)	6,639 (7%)	2,002 (6%)	8,641 (7%)

Specification highlighted in green is the chosen PS model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; Propensity score; SU: Serum urate

Figure L2: Comparison of PS distribution for gout hospitalisation analysis

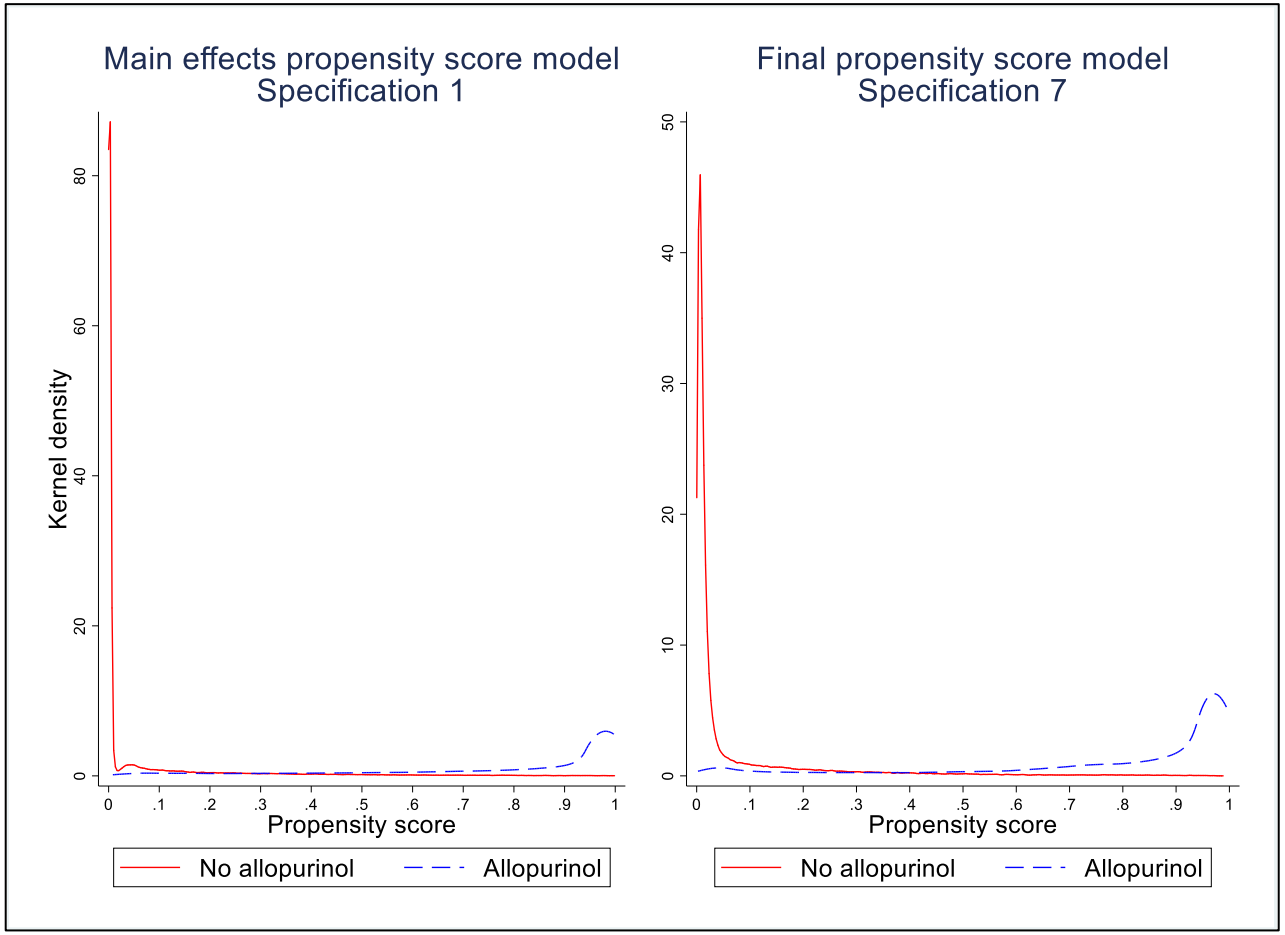


Table L3: PS model specification for joint replacement analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=104,827	Allopurinol N=43054	Overall N=147881
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.003 (0.002, 0.05) ($1*10^{-4}$, 0.99)	0.93 (0.67, 0.99) (0.01, 0.99)	72,043 (69%)	2,542 (6%)	74,585 (50%)
2	Main effects model + FP1 terms for age$^{(-2)}$, deprivation$^{(3)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$	0.01 (0.01, 0.06) ($7*10^{-4}$, 0.99)	0.91 (0.69, 0.98) (0.01, 0.99)	22,985 (22%)	954 (2%)	23,939 (16%)
3	Main effects model + FP2 terms for age$^{(2, 3)}$, deprivation$^{(-2, 3)}$, cumulative allopurinol use$^{(1, 2)}$, follow-up time$^{(1, 1)}$	0.01 (0.004, 0.05) ($3*10^{-4}$, 0.99)	0.92 (0.70, 0.98) (0.01, 0.99)	56,524 (54%)	4,812 (11%)	61,336 (41%)
4	Main effects model + FP2 terms for age$^{(2, 3)}$ + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation	0.01 (0.01, 0.06) ($50*10^{-4}$, 0.99)	0.91 (0.69, 0.98) (0.01, 0.99)	21,510 (21%)	830 (2%)	22,340 (15%)
5	Main effects model + FP2 terms for age$^{(2, 3)}$ + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation + colchicine*coronary heart disease	0.01 (0.01, 0.06) ($5*10^{-4}$, 0.99)	0.91 (0.69, 0.98) (0.01, 0.99)	19,283 (18%)	851 (2%)	20,134 (14%)
6	Main effects model + FP2 terms for age$^{(2, 3)}$ + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation + colchicine*coronary heart disease + body mass index*osteoarthritis	0.01 (0.01, 0.06) ($5*10^{-4}$, 0.99)	0.91 (0.6, 0.98) (0.01, 0.99)	17,177 (16%)	846 (2%)	18,023 (12%)

7	Main effects model + FP2 terms for age$^{(2, 3)}$ + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation + colchicine*coronary heart disease + body mass index*osteoarthritis + alcohol consumption*gout consultation	0.01 (0.007, 0.061) ($5*10^{-4}$, 0.99)	0.91 (0.69, 0.97) (0.005, 0.99)	15,979 (15%)	771 (2%)	16,750 (11%)
8	Main effects model + FP2 terms for age$^{(2, 3)}$ + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation + colchicine*coronary heart disease + body mass index*osteoarthritis + alcohol consumption*gout consultation + diuretic*analgesic	0.01 (0.01, 0.06) ($4*10^{-4}$, 0.99)	0.91 (0.69, 0.98) (0.01, 0.99)	14,548 (14%)	617 (1%)	15,165 (10%)

Specification highlighted in green is the chosen PS model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score

Figure L3: Comparison of PS distribution for joint replacement analysis

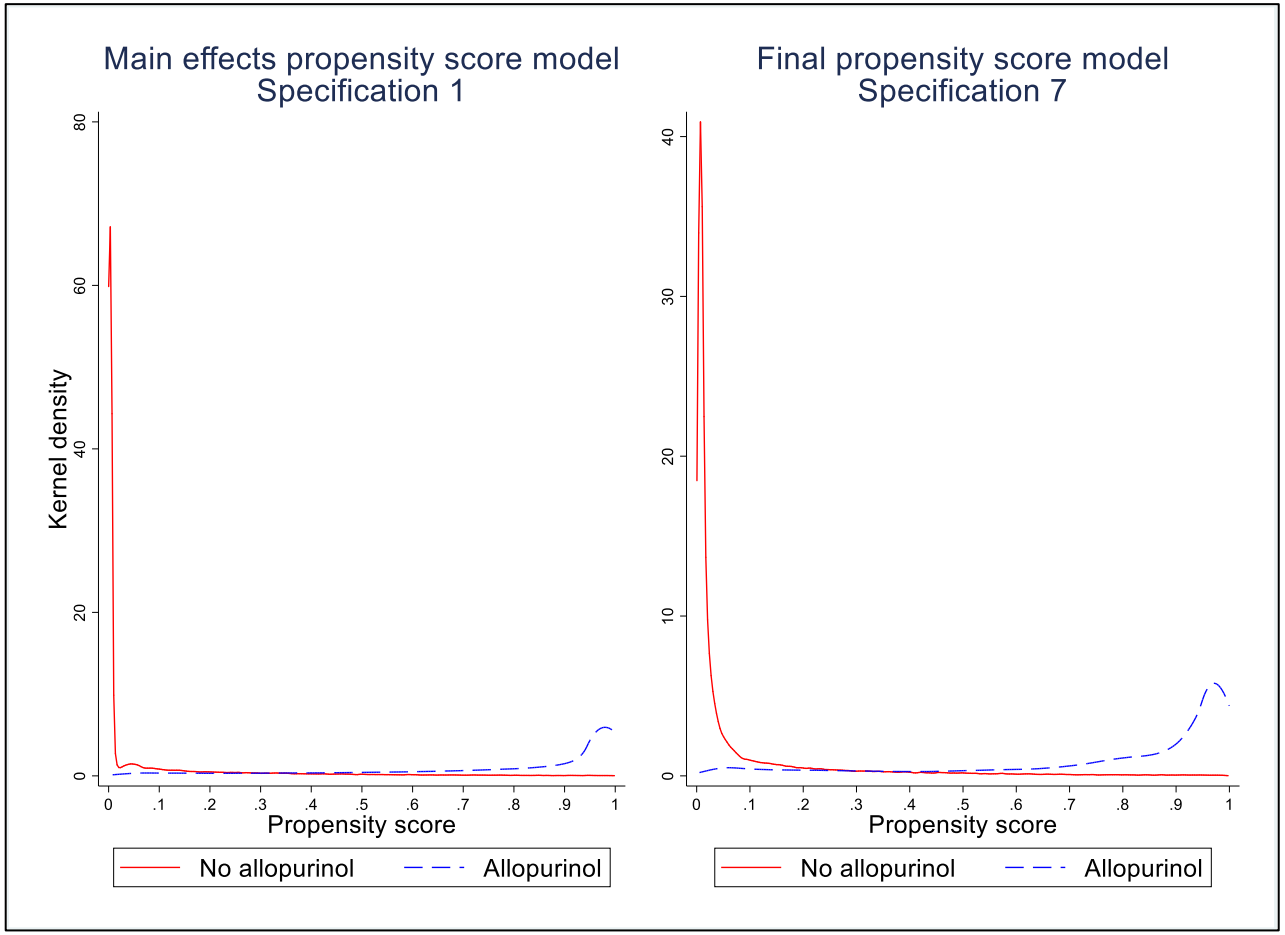


Table L4: PS model specification for cerebrovascular disease analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=102,769	Allopurinol N=42,339	Overall N=145,108
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.003 (0.002, 0.05) ($1*10^{-5}$, 0.99)	0.93 (0.66, 0.99) (0.01, 0.99)	70,628 (69)	756 (2%)	71,384 (49%)
2	Main effects model + FP1 terms for age$^{(-1)}$, deprivation$^{(-1)}$, cumulative allopurinol use$^{(0)}$, follow-up time$^{(0)}$	0.003 (0.002, 0.05) ($8*10^{-5}$, 0.99)	0.93 (0.66, 0.99) (0.01, 0.99)	72,162 (70%)	6,364 (15%)	78,526 (54%)
3	Main effects model + FP2 terms for age$^{(2, 3)}$, deprivation$^{(-1, 3)}$, cumulative allopurinol use$^{(0, 3)}$, follow-up time$^{(-2, -2)}$	0.003 (0.002, 0.05) ($6*10^{-5}$, 0.99)	0.93 (0.68, 0.99) (0.01, 0.99)	72,073 (70%)	10,061 (24%)	82,134 (57%)
4	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time + sex*follow-up time	0.003 (0.002, 0.05) ($9*10^{-5}$, 0.99)	0.93 (0.65, 0.99) (0.01, 0.99)	71,439 (70%)	739 (2%)	72,178 (50%)

Specification highlighted in green is the chosen PS model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score

Figure L4: PS distribution for cerebrovascular disease analysis

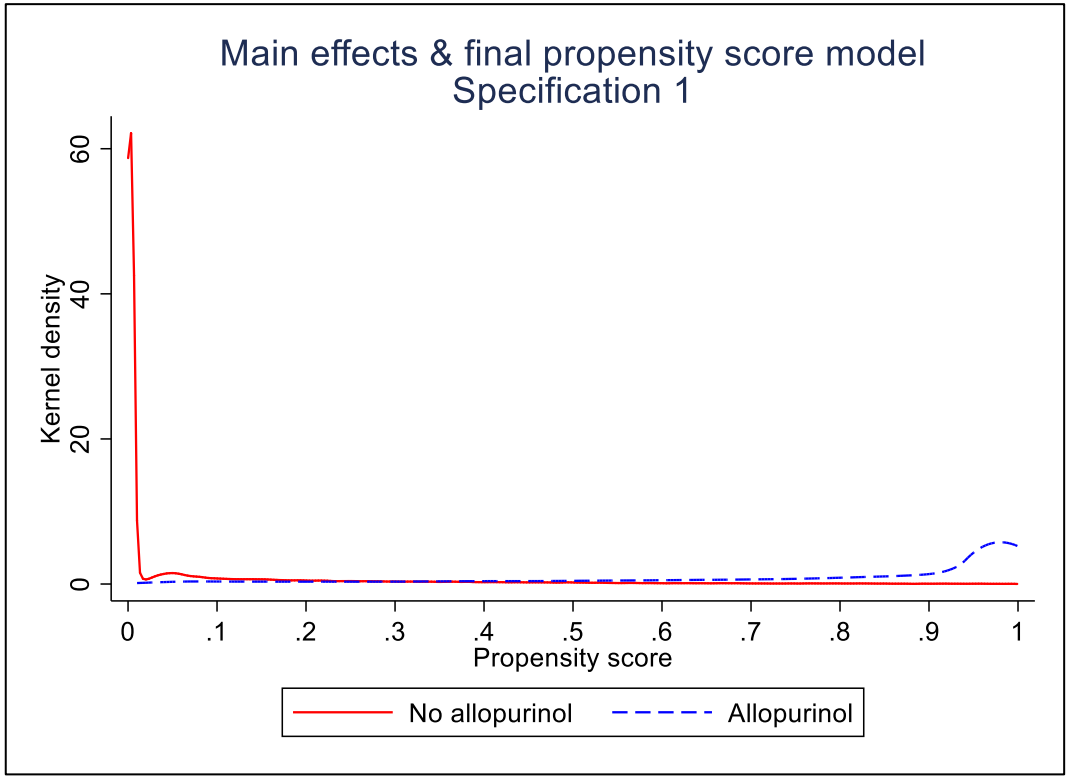


Table L5: PS model specification for coronary heart disease analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=84,348	Allopurinol N=32,308	Overall N=116,656
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.003 (0.002, 0.049) ($8.00*10^{-5}$, 0.999)	0.914 (0.638, 0.984) (0.010, 0.999)	58,716 (67%)	364 (1%)	59,080 (51%)
2	Main effects model + FP1 terms for age$^{(-2)}$, deprivation$^{(-1)}$, follow-up time$^{(0.5)}$ + linear term cumulative allopurinol use	0.003 (0.002, 0.049) ($9.00*10^{-5}$, 0.999)	0.914 (0.638, 0.984) (0.009, 0.999)	58,611 (69%)	375 (1%)	58,986 (51%)
3	Main effects model + FP2 terms for age$^{(2, 2)}$, deprivation$^{(-2 -2)}$, cumulative allopurinol use$^{(0, 1)}$, follow-up time$^{(-2, -0.5)}$	0.003 (0.002, 0.047) ($9.00*10^{-5}$, 0.997)	0.917 (0.647, 0.983) (0.010, 0.999)	58,687 (70%)	2,194 (7%)	60,881 (52%)
4	Main effects model + FP2 terms for age$^{(2, 2)}$, deprivation$^{(-2, -2)}$, follow-up time$^{(-2, -0.5)}$ + linear term for cumulative allopurinol use	0.003 (0.002, 0.05) ($7*10^{-5}$, 0.99)	0.92 (0.64, 0.98) (0.009, 0.99)	58,257 (69%)	534 (2%)	58,791 (50%)
5	Main effects model + FP2 terms for age$^{(2, 2)}$, deprivation$^{(-2, -2)}$, follow-up time$^{(-2, -0.5)}$ + linear term for cumulative allopurinol use + smoking status*diuretics	0.003 (0.002, 0.05) ($8*10^{-5}$, 0.99)	0.92 (0.64, 0.98) (0.008, 0.99)	57,508 (68%)	498 (2%)	58,006 (50%)

Specification highlighted in green is the chosen propensity model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score

Figure L5: Comparison of PS distribution for coronary heart disease analysis

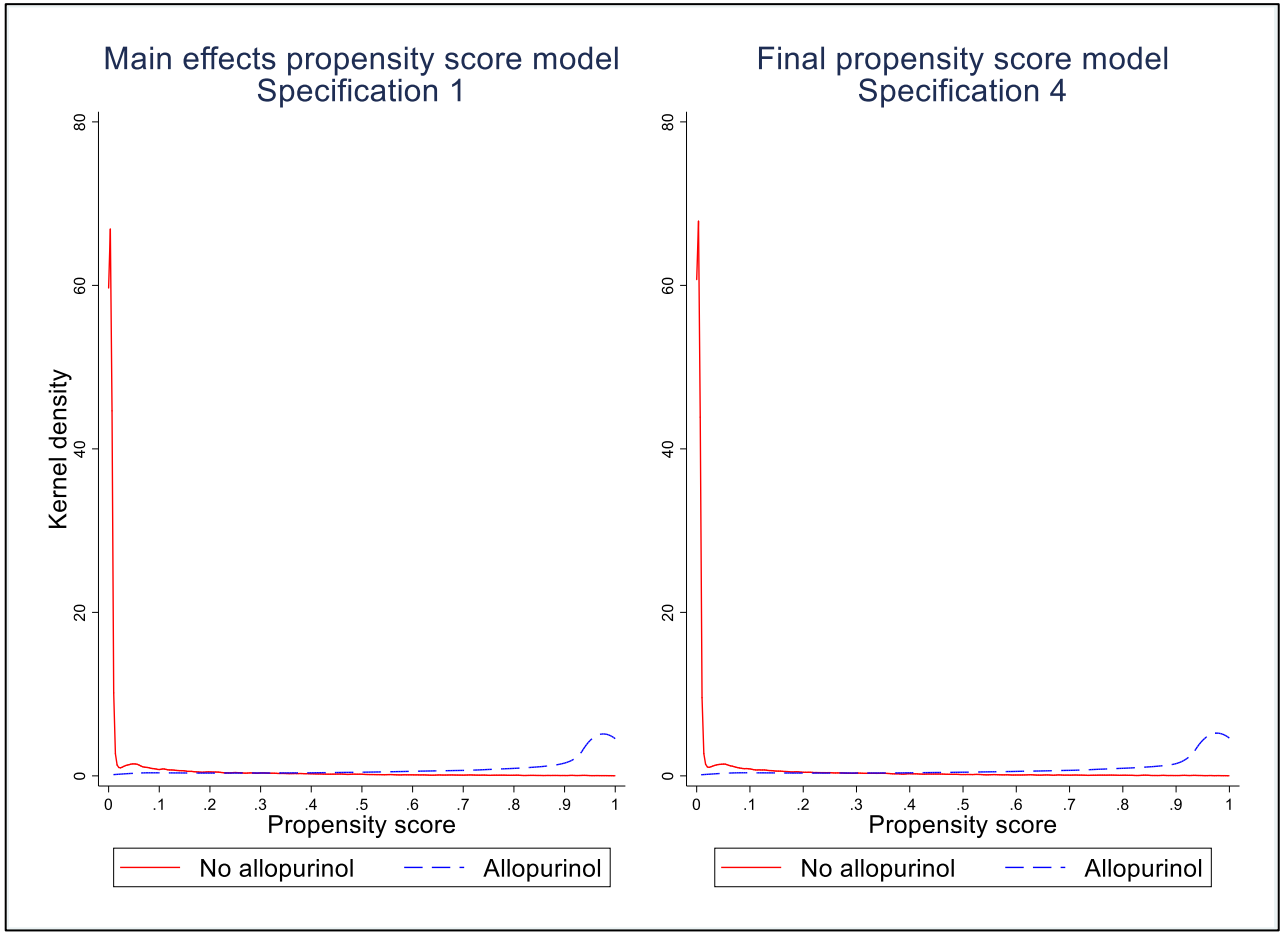


Table L6: PS model specification for peripheral vascular disease analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=106,173	Allopurinol N=43,978	Overall N=150,151
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.003 (0.002, 0.05) ($9*10^{-5}$, 0.99)	0.93 (0.65, 0.99) (0.01, 0.99)	74,245 (70%)	1,079 (2%)	75,324 (50%)
2	Main effects model + FP1 terms for age$^{(-2)}$, deprivation$^{(0.5)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(0)}$	0.02 (0.01, 0.07) ($5*10^{-4}$, 0.99)	0.92 (0.66, 0.98) (0.01, 0.99)	2,616 (25%)	384 (1%)	27,100 (18%)
3*	Main effects model + FP2 terms for age$^{(1, 2)}$, deprivation$^{(0, 2)}$, cumulative allopurinol use$^{(-0.5, 3)}$, follow-up time$^{(-0.5, 3)}$	-	-	-	-	-
4	Main effects model + FP2 terms for follow-up time$^{(-0.5, 3)}$ + FP1 terms for age$^{(-2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms for deprivation	0.02 (0.01, 0.07) ($5*10^{-4}$, 0.99)	0.91 (0.66, 0.98) (0.01, 0.99)	24,750 (23%)	569 (1%)	25,319 (17%)
5	Main effects model + FP2 terms for follow-up time$^{(-0.5, 3)}$ + FP1 terms for age$^{(-2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms for deprivation + smoking status*diuretics	0.02 (0.01, 0.07) ($6*10^{-4}$, 0.99)	0.91 (0.67, 0.97) (0.01, 0.99)	21,456 (20%)	371 (1%)	21,827 (15%)
6	Main effects model + FP2 terms for follow-up time$^{(-0.5, 3)}$ + FP1 terms for age$^{(-2)}$, cumulative allopurinol use$^{(0.5)}$ + linear terms for deprivation + smoking status*diuretics + SU level*hypertension	0.02 (0.01, 0.07) ($5*10^{-4}$, 0.99)	0.91 (0.66, 0.98) (0.01, 0.99)	21,086 (20%)	342 (1%)	21,428 (14%)

*Propensity score model did not converge; Specification highlighted in green is the chosen propensity model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score; SU: Serum urate

Figure L6: Comparison of PS distribution for peripheral vascular disease analysis

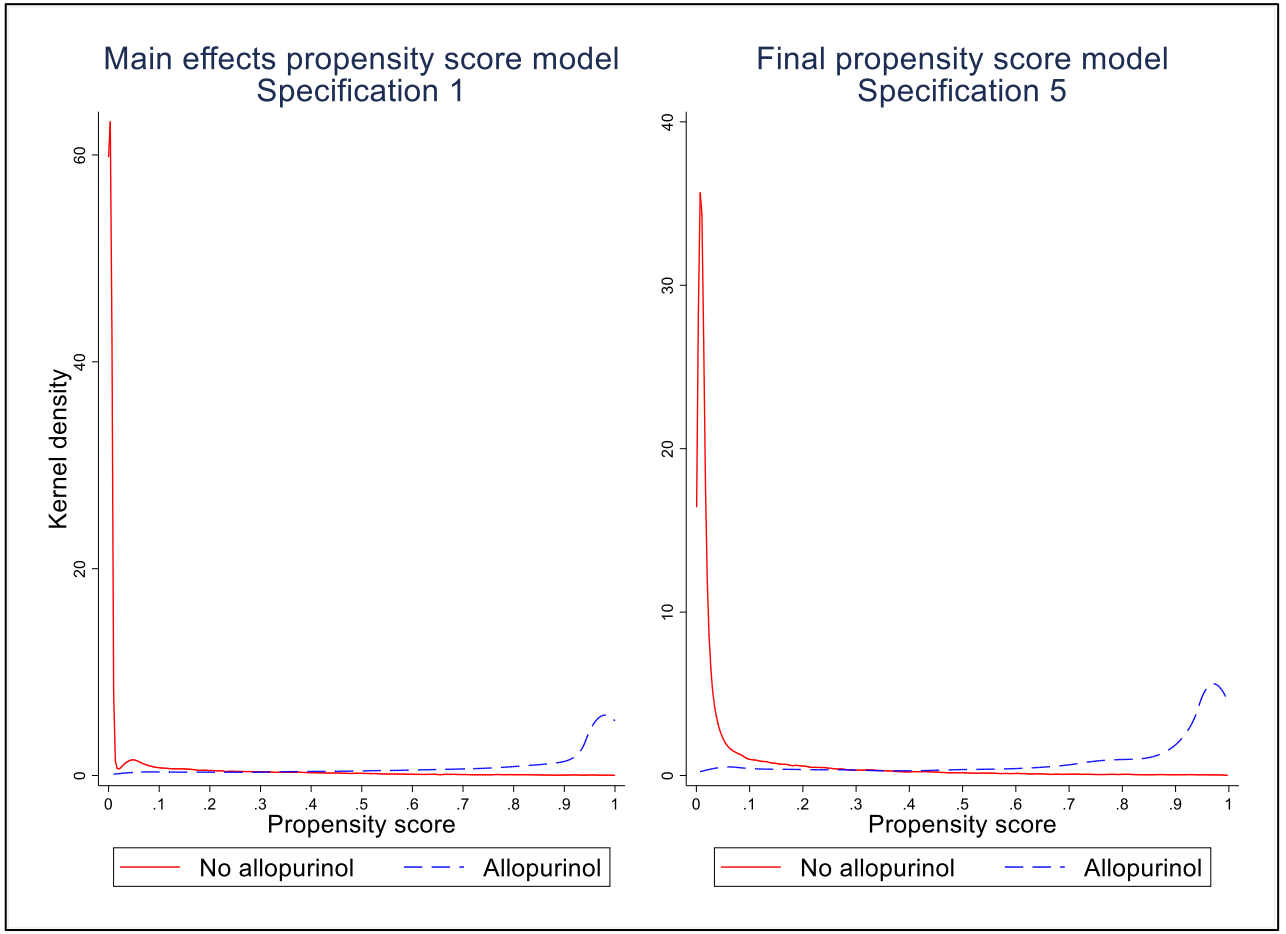
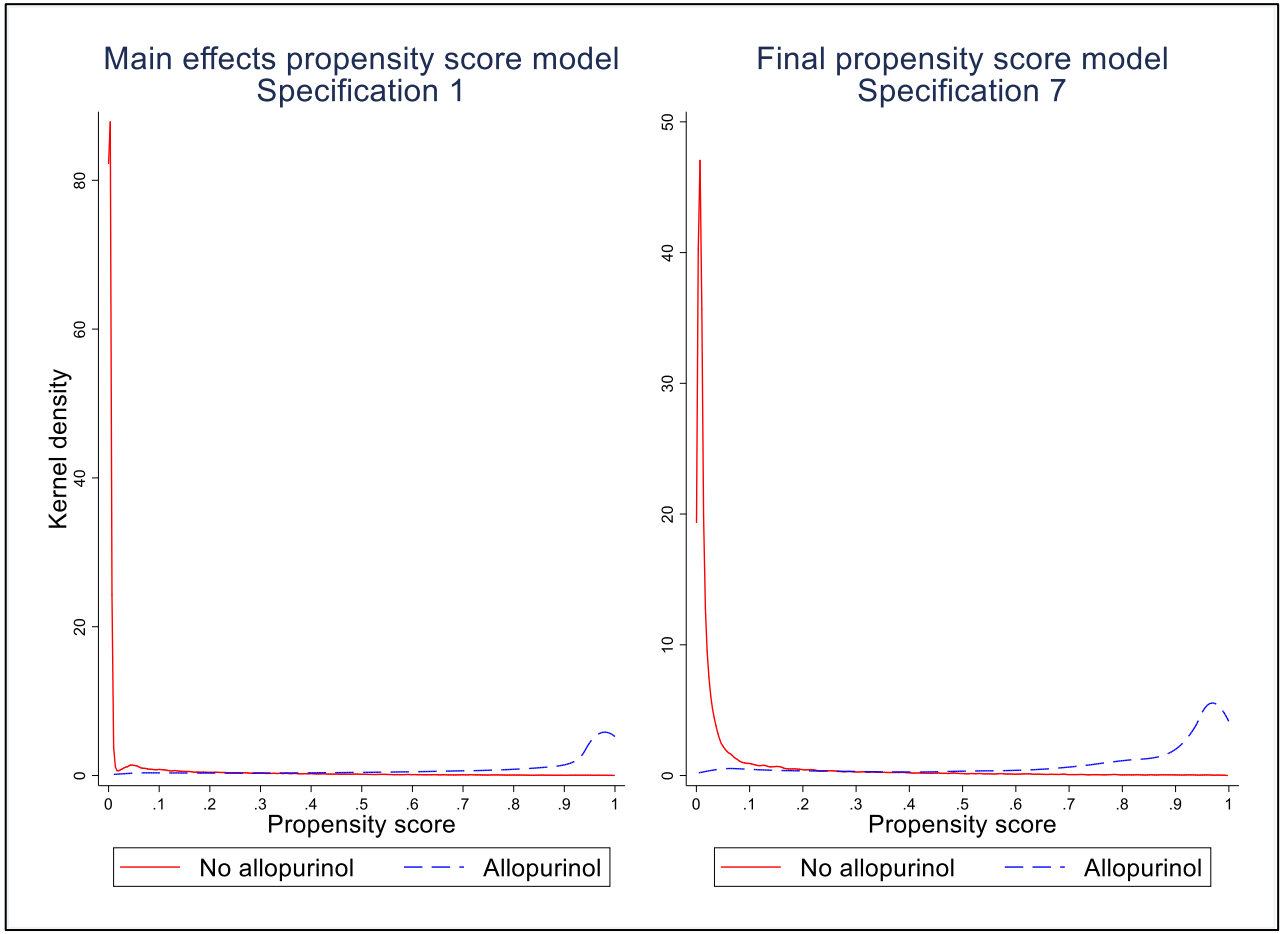


Table L7: PS model specification for renal disease analysis

	PS model specification	Median PS (IQR) (Range)		Number of intervals outside the region of common support		
		No allopurinol	Allopurinol	No allopurinol N=98,618	Allopurinol N=37,370	Overall N=135,988
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time	0.002 (0.001, 0.04) ($9*10^{-5}$, 0.99)	0.93 (0.66, 0.99) (0.01, 0.99)	71,622 (73%)	1,871 (5%)	73,493 (54%)
2	Main effects model + FP1 terms for age$^{(0)}$, deprivation$^{(-1)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$	0.01 (0.01, 0.05) ($40*10^{-4}$, 0.99)	0.91 (0.68, 0.97) (0.01, 0.99)	22,481 (23%)	792 (2%)	23,273 (17%)
3	Main effects model + FP2 terms for age$^{(3, 3)}$, deprivation$^{(0.5, 2)}$, cumulative allopurinol use$^{(1, 2)}$, follow-up time$^{(1, 1)}$	0.01 (0.003, 0.04) ($2*10^{-4}$, 0.99)	0.92 (0.69, 0.98) (0.01, 0.99)	53,895 (55%)	4,216 (11%)	58,111 (43%)
4	Main effects model + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear terms for age, deprivation	0.01 (0.01, 0.05) ($5*10^{-4}$, 0.99)	0.91 (0.68, 0.97) (0.01, 0.99)	25,972 (26%)	763 (2%)	26,735 (20%)
5	Main effects model + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear terms for age, deprivation + SU level*colchicine	0.01 (0.01, 0.05) ($4*10^{-4}$, 0.99)	0.91 (0.68, 0.97) (0.01, 0.99)	21,211 (22%)	766 (2%)	21,977 (16%)
6	Main effects model + FP1 terms for cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$ + linear terms for age, deprivation + SU level*colchicine + SU level*body mass index	0.01 (0.01, 0.05) ($3*10^{-4}$, 0.99)	0.91 (0.68, 0.97) (0.01, 0.99)	20,998 (21%)	730 (2%)	21,728 (16%)

Specification highlighted in green is the chosen propensity model; Values in brackets indicate which fractional polynomial terms were used; FP1: Fractional polynomials of dimension 1; FP2: Fractional polynomials of dimension 2; IQR: Interquartile range; PS: Propensity score; SU: Serum urate

Figure L7: Comparison of PS distribution for renal disease analysis



Appendix M Varying number of time-varying PS subclasses

Table M1: PS and outcome distribution, imbalanced covariates in mortality analysis

PS range	Deaths in no allopurinol intervals N (%)	Deaths in allopurinol intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: <0.001, 0.01	1029 (31)	0 (0)	116,498 (75) 16,123 (96)	Balance achieved
2: 0.01, 0.04	1311 (39)	13 (1)		
3: 0.04, 0.60	857 (26)	345 (21)		
4: 0.60, 0.99	149 (4)	1272 (78)		
5 subclasses				
1: <0.001, 0.01	812 (24)	0 (0)	93,198 (60) 14,542 (86)	Balance achieved
2: 0.01, 0.02	870 (26)	0 (0)		
3: 0.02, 0.14	1285 (38)	78 (5)		
4: 0.14, 0.79	302 (9)	445 (27)		
5: 0.79, 0.99	77 (2)	1107 (68)		
6 subclasses				
1: <0.001, 0.01	682 (20)	0 (0)	103,554 (67) 15,703 (93)	Balance achieved
2: 0.01, 0.01	670 (20)	0 (0)		
3: 0.01, 0.04	988 (30)	13 (1)		
4: 0.04, 0.28	716 (21)	116 (7)		
5: 0.28, 0.89	246 (7)	503 (31)		
6: 0.89, 0.99	44 (1)	998 (61)		
7 subclasses				
1: <0.001, 0.01	571 (17)	0 (0)	88,760 (57) 13,634 (81)	Cumulative allopurinol use
2: 0.01, 0.01	592 (18)	0 (0)		
3: 0.01, 0.02	660 (20)	2 (0)		
4: 0.02, 0.10	1075 (32)	49 (3)		
5: 0.10, 0.43	231 (7)	136 (8)		
6: 0.44, 0.93	193 (6)	577 (35)		
7: 0.93, 1.00	24 (1)	866 (53)		

Propensity score subclassification was performed on 155,331 intervals from 16,876 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M2: PS and outcome distribution, imbalanced covariates in gout hospitalisation analysis

PS range	Hospitalisations in non-allopurinol intervals N (%)	Hospitalisations in allopurinol intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: <0.00, 0.01	194 (21)	0 (0)	91,551 (75) 13,480 (96)	Balance achieved
2: 0.01, 0.02	308 (33)	2 (0)		
3: 0.02, 0.52	343 (37)	145 (14)		
4: 0.52, 0.99	78 (8)	902 (86)		
5 subclasses				
1: <0.001, 0.01	143 (15)	0 (0)	97,654 (80) 13,587 (96)	Balance achieved
2: 0.01, 0.01	231 (25)	1 (0)		
3: 0.01, 0.09	341 (37)	42 (4)		
4: 0.09, 0.75	169 (18)	237 (23)		
5: 0.75, 0.99	39 (4)	769 (73)		
6 subclasses				
1: <0.001, 0.01	121 (13)	0 (0)	81,378 (67) 13,259 (94)	Balance achieved
2: 0.01, 0.01	178 (19)	0 (0)		
3: 0.01, 0.02	203 (22)	2 (0)		
4: 0.02, 0.20	262 (28)	72 (7)		
5: 0.20, 0.86	141 (15)	305 (29)		
6: 0.86, 0.99	18 (2)	670 (64)		
7 subclasses				
1: <0.001, 0.01	96 (10)	0 (0)	87,191 (71) 13,411 (95)	Cumulative allopurinol use
2: 0.01, 0.01	131 (14)	0 (0)		
3: 0.01, 0.02	185 (20)	1 (0)		
4: 0.02, 0.06	255 (28)	25 (2)		
5: 0.06, 0.35	142 (15)	74 (7)		
6: 0.35, 0.91	106 (11)	360 (34)		
7: 0.91, 0.99	8 (1)	589 (56)		

Propensity score subclassification was performed on 122,068 intervals from 14,087 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M3: PS and outcome distribution, imbalanced covariates in joint replacement analysis

PS range	Joint replacement in non-allopurinol intervals N (%)	Joint replacement in allopurinol Intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: 0.001, 0.01	194 (28)	0 (0)	110,910 (75) 16,030 (96)	Balance achieved
2: 0.01, 0.04	311 (45)	1 (0)		
3: 0.04, 0.60	161 (23)	34 (11)		
4: 0.60, 1.00	26 (4)	288 (89)		
5 subclasses				
1: 0.001, 0.01	144 (21)	0 (0)	88,728 (60) 15,455 (93)	Balance achieved
2: 0.01, 0.02	247 (36)	0 (0)		
3: 0.02, 0.13	218 (32)	8 (2)		
4: 0.13, 0.79	64 (9)	73 (23)		
5: 0.80, 1.00	19 (3)	242 (75)		
6 subclasses				
1: 0.001, 0.01	114 (16)	0 (0)	98,587 (67) 15,880 (95)	Balance achieved
2: 0.01, 0.01	184 (27)	0 (0)		
3: 0.01, 0.04	207 (30)	1 (0)		
4: 0.04, 0.26	136 (20)	13 (4)		
5: 0.26, 0.88	40 (6)	93 (29)		
6: 0.88, 1.00	11 (2)	216 (67)		
7 subclasses				
1: 0.001, 0.01	94 (14)	0 (0)	84,503 (57) 14,603 (88)	Balance achieved
2: 0.01, 0.01	132 (19)	0 (0)		
3: 0.01, 0.02	203 (29)	0 (0)		
4: 0.02, 0.09	157 (23)	7 (2)		
5: 0.09, 0.43	66 (10)	10 (3)		
6: 0.43, 0.92	32 (5)	113 (35)		
7: 0.92, 1.00	8 (1)	193 (60)		

Propensity score subclassification was performed on 147,881 intervals from 16,644 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M4: PS and outcome distribution, imbalanced covariates in cerebrovascular disease analysis

PS range	Cerebrovascular events in non-allopurinol intervals N (%)	Cerebrovascular events in allopurinol intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: <0.001, 0.002	149 (13)	0 (0)		
2: 0.002, 0.02	820 (74)	0 (0)	108,831 (75)	Cumulative allopurinol use
3: 0.03, 0.60	118 (11)	67 (15)	15,173 (93)	
4: 0.60, 0.99	27 (2)	371 (85)		
5 subclasses				
1: <0.001, 0.002	104 (9)	0 (0)		
2: 0.002, 0.004	333 (30)	0 (0)	87,064 (60)	Balance achieved
3: 0.004, 0.14	594 (53)	8 (2)	13,259 (82)	
4: 0.14, 0.79	69 (6)	103 (24)		
5: 0.79, 0.99	14 (1)	327 (75)		
6 subclasses				
1: <0.001, 0.002	90 (8)	0 (0)		
2: 0.002, 0.003	199 (18)	0 (0)		
3: 0.003, 0.02	680 (61)	0 (0)	72,554 (50)	Balance achieved
4: 0.02, 0.30	82 (7)	27 (6)	7,565 (47)	
5: 0.30, 0.89	54 (5)	149 (34)		
6: 0.89, 1.00	9 (1)	262 (60)		
7 subclasses				
1: <0.001, 0.002	81 (7)	0 (0)		
2: 0.002, 0.003	127 (11)	0 (0)		
3: 0.003, 0.005	321 (29)	0 (0)	82,918 (57)	Balance achieved
4: 0.005, 0.10	482 (43)	4 (1)	12,341 (76)	
5: 0.010, 0.46	63 (6)	42 (10)		
6: 0.46, 0.94	36 (3)	157 (36)		
7: 0.94, 1.00	4 (0)	235 (54)		

Propensity score subclassification was performed on 145,108 intervals from 16,253 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M5: PS and outcome distribution, imbalanced covariates in coronary heart disease analysis

PS range	Coronary heart events: Non-allopurinol intervals N (%)	Coronary heart events: Allopurinol intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: <0.001, 0.002	259 (12)	0 (0)		
2: 0.002, 0.01	1258 (60)	0 (0)	58,328 (50)	Cumulative allopurinol use
3: 0.01, 0.54	495 (24)	124 (14)	7,473 (53)	
4: 0.54, 0.99	68 (3)	746 (86)		
5 subclasses				
1: <0.001, 0.002	200 (10)	0 (0)		
2: 0.002, 0.004	530 (25)	0 (0)	69,993 (60)	
3: 0.004, 0.11	1168 (56)	12 (1)	11,503 (82)	Balance achieved
4: 0.11, 0.74	143 (7)	186 (21)		
5: 0.74, 0.99	39 (2)	672 (77)		
6 subclasses				
1: <0.001, 0.002	169 (8)	0 (0)		
2: 0.002, 0.003	283 (14)	0 (0)		
3: 0.003, 0.01	1065 (51)	0 (0)	58,328 (50)	Balance achieved
4: 0.01, 0.25	437 (21)	50 (6)	7,473 (53)	
5: 0.25, 0.85	107 (5)	268 (31)		
6: 0.85, 0.99	19 (1)	552 (63)		
7 subclasses				
1: <0.001, 0.001	155 (7)	0 (0)		
2: 0.001, 0.002	167 (8)	0 (0)		
3: 0.002, 0.004	530 (25)	0 (0)	66,660 (57)	
4: 0.004, 0.07	1023 (49)	4 (0)	10,848 (77)	Balance achieved
5: 0.07, 0.40	115 (6)	78 (9)		
6: 0.40, 0.91	81 (4)	348 (40)		
7: 0.91, 0.99	9 (0)	440 (51)		

Propensity score subclassification was performed on 116,656 intervals from 14,063 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M6: PS and outcome distribution, imbalanced covariates in peripheral vascular disease analysis

PS range	Peripheral vascular events: non-allopurinol intervals N (%)	Peripheral vascular events: Allopurinol Intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: 0.001, 0.01	120 (28)	0 (0)	112,613 (75) 15,948 (97)	Balance achieved
2: 0.01, 0.05	166 (38)	4 (2)		
3: 0.05, 0.59	132 (31)	40 (20)		
4: 0.59, 0.99	14 (3)	159 (78)		
5 subclasses				
1: 0.001, 0.01	93 (22)	0 (0)	120,120 (80) 16,103 (97)	Balance achieved
2: 0.01, 0.02	141 (33)	1 (0)		
3: 0.02, 0.13	153 (35)	12 (6)		
4: 0.13, 0.78	40 (9)	50 (25)		
5: 0.78, 0.99	5 (1)	140 (69)		
6 subclasses				
1: 0.001, 0.01	77 (18)	0 (0)	100,100 (67) 15,769 (95)	Balance achieved
2: 0.007, 0.01	102 (24)	0 (0)		
3: 0.01, 0.05	107 (25)	4 (2)		
4: 0.05, 0.27	115 (27)	17 (8)		
5: 0.27, 0.88	26 (6)	61 (30)		
6: 0.88, 0.99	5 (1)	121 (60)		
7 subclasses				
1: 0.001, 0.01	65 (15)	0 (0)	107,250 (71) 15,861 (96)	Balance achieved
2: 0.01, 0.01	76 (18)	0 (0)		
3: 0.01, 0.03	107 (25)	1 (0)		
4: 0.03, 0.09	125 (29)	8 (4)		
5: 0.09, 0.43	37 (9)	18 (9)		
6: 0.43, 0.92	18 (4)	69 (34)		
7: 0.92, 0.99	4 (1)	107 (53)		

Propensity score subclassification was performed on 150,151 intervals from 16,519 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Table M7: PS and outcome distribution, imbalanced covariates in renal disease analysis

Propensity score range	Renal disease events: non-allopurinol intervals N (%)	Renal disease events: allopurinol Intervals N (%)	Number to be analysed in outcome analysis Intervals N (%) Patients N (%)[a]	Imbalanced covariates[b]
4 subclasses				
1: <0.001, 0.01	750 (30)	0 (0)	101,991 (75) 15,886 (96)	Cumulative allopurinol use
2: 0.01, 0.03	1216 (48)	6 (0)		
3: 0.03, 0.52	445 (18)	152 (11)		
4: 0.52, 0.99	130 (5)	1226 (89)		
5 subclasses				
1: <0.001, 0.01	561 (22)	0 (0)	81,592 (60) 15,565 (94)	Balance achieved
2: 0.01, 0.01	1021 (40)	0 (0)		
3: 0.01, 0.09	587 (23)	49 (4)		
4: 0.09, 0.75	313 (12)	291 (21)		
5: 0.75, 0.99	59 (2)	1044 (75)		
6 subclasses				
1: <0.001, 0.01	454 (18)	0 (0)	90,658 (67) 15,762 (95)	Balance achieved
2: 0.01, 0.01	712 (28)	0 (0)		
3: 0.01, 0.03	800 (31)	6 (0)		
4: 0.03, 0.21	311 (12)	88 (6)		
5: 0.21, 0.85	227 (9)	345 (25)		
6: 0.85, 0.99	37 (1)	945 (68)		
7 subclasses				
1: <0.001, 0.01	378 (15)	0 (0)	97,134 (71) 15,821 (96)	Balance achieved
2: 0.01, 0.01	537 (21)	0 (0)		
3: 0.01, 0.02	814 (32)	2 (0)		
4: 0.02, 0.06	371 (15)	27 (2)		
5: 0.06, 0.36	263 (10)	88 (6)		
6: 0.36, 0.90	158 (6)	416 (30)		
7: 0.90, 0.99	20 (1)	851 (61)		

Propensity score subclassification was performed on 135,988 intervals from 16,508 patients; Results highlighted in green indicated the number of subclasses used for outcome analysis; [a]At least one interval from the patient would be included in outcome analysis; [b]Overall SMD >0.10 and assessed in subclasses with outcome occurring in both no allopurinol and allopurinol intervals; PS: Propensity score

Appendix N PS model specification

Table N1: Specification of the PS score model for each outcome analysis

Outcome	PS model specification
Target SU level	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time, SU level
2	Main effects model + FP1 terms for age$^{(-2)}$, deprivation $^{(3)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$, SU level$^{(-0.5)}$
3	Main effects model + FP2 terms for age$^{(-2, -2)}$, deprivation $^{(1, 2)}$, cumulative allopurinol use$^{(-0.5, 0.5)}$, follow-up time$^{(-2, 0.5)}$, SU level$^{(-2 -2)}$
4	Main effects model + FP2 terms for age$^{(-2, -2)}$, cumulative allopurinol use$^{(-0.5, 0.5)}$ + FP1 terms for deprivation $^{(3)}$, SU level$^{(-0.5)}$ + linear terms for follow-up time
Mortality	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, deprivation $^{(2)}$, cumulative allopurinol use$^{(0.5)}$, follow-up time$^{(-2)}$
3	Main effects model + FP2 terms for age$^{(3, 3)}$, deprivation $^{(2, 3)}$, cumulative allopurinol use$^{(-1, 0)}$, follow-up time$^{(-2, -2)}$
4	Main effects model + FP2 terms for age$^{(3, 3)}$, deprivation $^{(2, 3)}$, cumulative allopurinol use$^{(-1, 0)}$ + linear terms for follow-up time
5	Main effects model + FP2 terms for age$^{(3, 3)}$, deprivation $^{(2, 3)}$, cumulative allopurinol use$^{(-1, 0)}$ + linear terms for follow-up time + interaction terms between SU level*hypertension + interaction term diuretic use*follow-up time

Outcome	PS model specification
Gout hospitalisation	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-1)}$ + linear terms for deprivation
3	Main effects model + FP2 terms for age$^{(-2, 2)}$, deprivation $^{(3, 3)}$, cumulative allopurinol use$^{(-1, -0.5)}$, follow-up time$^{(0.5, 3)}$
4	Main effects model + FP2 terms for deprivation $^{(3, 3)}$ + FP1 terms for age$^{(3)}$, follow-up time$^{(-1)}$ + linear terms for cumulative allopurinol use
Joint replacement	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-1)}$ + linear term for deprivation
3	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation $^{(-0.5, 0)}$, cumulative allopurinol use$^{(-1, -0.5)}$, follow-up time$^{(0.5, 3)}$
4	Main effects model + FP2 terms for deprivation $^{(-0.5, 0)}$ + FP1 terms for follow-up time$^{(-1)}$ + linear terms for age, cumulative allopurinol use
5	Main effects model + FP2 terms for deprivation $^{(-0.5, 0)}$ + FP1 terms for follow-up time$^{(-1)}$ + linear terms for age, cumulative allopurinol use + interaction term coronary heart disease*hypertension

Outcome	PS model specification
Cerebrovascular disease	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear term for deprivation
3	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation $^{(3, 3)}$, cumulative allopurinol use$^{(-1, -0.5)}$, follow-up time$^{(-1, 3)}$
4	Main effects model + FP2 terms for deprivation $^{(3, 3)}$ + FP1 terms for cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear terms for age
5	Main effects model + FP2 terms for deprivation $^{(3, 3)}$ + FP1 terms for cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear terms for age + interaction term coronary heart disease*sex
Coronary heart disease	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, deprivation $^{(2)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-1)}$
3	Main effects model + FP2 terms for age$^{(-2, 2)}$, deprivation $^{(0, 0)}$, cumulative allopurinol use$^{(-1, -0.5)}$, follow-up time$^{(-1, 3)}$
4	Main effects model + FP2 terms for age$^{(-2, 2)}$, deprivation $^{(0, 0)}$, cumulative allopurinol use$^{(-1, -0.5)}$ + FP1 terms for follow-up time$^{(-1)}$

Outcome	PS model specification
Peripheral vascular disease	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear terms for deprivation
3	Main effects model + FP2 terms for age$^{(-2, 3)}$, deprivation $^{(3, 3)}$, cumulative allopurinol use$^{(-1, -0.5)}$, follow-up time$^{(-0.5, 3)}$
4	Main effects model + FP2 terms for deprivation $^{(3, 3)}$ + FP1 terms for follow-up time$^{(-2)}$ + linear terms for age, cumulative allopurinol use
5	Main effects model + FP2 terms for deprivation $^{(3, 3)}$ + FP1 terms for follow-up time$^{(-2)}$ + linear terms for age, cumulative allopurinol use + interaction term SU level*cumulative allopurinol use
Renal disease	
1	Main effects model + linear terms for age, deprivation, cumulative allopurinol use, follow-up time
2	Main effects model + FP1 terms for age$^{(3)}$, deprivation $^{(2)}$, cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$
3	Main effects model + FP2 terms for age$^{(-2, 2)}$, deprivation $^{(0, 0.5)}$, cumulative allopurinol use$^{(-2, -0.5)}$, follow-up time$^{(0, 3)}$
4	Main effects model + FP2 terms for deprivation $^{(0, 0.5)}$ + FP1 terms for cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear terms for age
5	Main effects model + FP2 terms for deprivation $^{(0, 0.5)}$ + FP1 terms for cumulative allopurinol use$^{(-0.5)}$, follow-up time$^{(-2)}$ + linear terms for age + interaction term coronary heart disease*sex + interaction term diabetes*sex

Values in brackets indicate the fractional polynomials used; PS: Propensity score; SU: Serum urate