Long-term outcomes and the effects of maintenance treatments in bipolar disorder

Joseph F Hayes
UCL
PhD Thesis
2017

Declaration: I, Joseph Hayes, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Acknowledgements

I would like to thank the following for their specific contributions to the work involved in this thesis:

Supervisors: Professor David Osborn, Dr Kate Walters, Dr Louise Marston and Professor Michael King, who went above and beyond the role, always pulled in the same direction and made it a very enjoyable time.

Funders: The Medical Research Council awarded me a Population Health Scientist Fellowship, which funded this PhD and an MSc in Epidemiology at the London School of Hygiene and Tropical Medicine, which was an amazing experience without which I would not have had the skills required to complete this work.

Collaborators: Professor John Geddes, University of Oxford, who contributed wise words to the published versions of the maintenance medication studies, and Dr Alexandra Pitman, UCL, who made invaluable contributions to the suicide/self-harm study.

Upgrade examiner: Dr Laura Horsfall, for an interesting few hours!

Colleagues: Alexandra Pitman, Hind Khalifeh, James Kirkbride, Vishal Bhavsar, Francesca Solmi, Jen Dykxhoorn, Nomi Werbeloff, Gemma Lewis and others in the Division of Psychiatry, UCL, for tech support, discussions about statistics, writing, time keeping, and coffee.

Family: Especially Stella and Conrad for support, sustenance and sleepless nights, and Roma for finding all the typos!
Abstract

Objectives
To determine 1) mortality and morbidity in people with bipolar disorder, and 2) the impact of maintenance medication on relapse/reoccurrence and adverse events.

Methods

Objective 1: I conducted a meta-analysis of studies examining mortality in bipolar disorder populations. I then carried out a cohort study in United Kingdom primary care electronic health records to understand rates of mortality and morbidity in bipolar disorder relative to the general population.

Objective 2: I completed a network meta-analysis of the efficacy of maintenance mood stabiliser medications (lithium, valproate, olanzapine and quetiapine) in preventing relapse. I then carried out a series of cohort studies in primary care electronic health records. These studies examined 1) the effectiveness and tolerability of these medications, 2) the rates of renal, endocrine, hepatic and metabolic adverse events, and 3) the rates of self-harm, accidental injury and suicide. Propensity score methods were used to address issues of confounding.

Results

Objective 1: All-cause and cause specific mortality was elevated in people with bipolar disorder (summary standardised mortality ratio 2.05; 95% CI 1.89 to 2.23). In a cohort of 17,341 with bipolar disorder, mortality rates increased from the mid-2000s relative to the general population (hazard ratio increased by 0.14 per year; 95% CI 0.10 to 0.19).
Objective 2: Trials comparing lithium, valproate, olanzapine, quetiapine and placebo did not show superiority of one drug. In the electronic health records cohort studies individuals prescribed lithium went for longer before treatment failure (for example valproate had hazard ratio 1.20; 95% CI 1.10 to 1.32 compared with lithium), had increased mild (but not severe) renal failure (hazard ratio for valproate: 0.56; 95% CI 0.45 to 0.69 compared with lithium), hypo- and hyperthyroidism and hypercalcemia rates. However, they had lower rates of clinically significant weight gain (hazard ratio for >15% weight gain with valproate: 1.62; 95% CI 1.31 to 2.01 compared with lithium) and there was no difference in hepatotoxicity, cardiovascular events or diabetes mellitus rates. Additionally, people taking lithium had lower self-harm (hazard ratio for alternatives: 1.51; 95% CI 1.21 to 1.88 compared to lithium) and accidental injury rates.

Conclusions

Bipolar disorder is associated with increased mortality and morbidity, and the mortality gap with the general population has widened in recent years. Despite limited trial evidence, lithium appears to offer the best opportunity for mood stabilisation. Lithium is associated with increased renal and endocrine dysfunction, but these risks are offset by the potential of more frequent weight gain with alternative drugs. Furthermore, lithium may be associated with specific anti-suicidal effects. These risk and benefits should be considered when individual treatment decisions are made.
Acronyms

ACEi – angiotensin converting enzyme inhibitor
ACU – acceptable computer usage
ALT - alanine transaminase
AMR – acceptable mortality reporting
APC – annual percentage change
AST - aspartate aminotransferase
BMI – body mass index
BPD – bipolar disorder
CKD – chronic kidney disease
CVD – cardiovascular disease
CVE – cerebrovascular event
CI – confidence interval
COPD – chronic obstructive pulmonary disease
CPRD – Clinical Practice Research Datalink
DSM – Diagnostic and Statistical Manual of Mental Disorders
ECT – electroconvulsive therapy
EHR – electronic health record
EMIS – Egton Medical Information Systems
FDA – Food and Drug Administration
FGA – first generation antipsychotic
GP – general practitioner
HR – hazard ratio
HES – Hospital episode statistics
ICD – International Classification of Diseases
IHD – ischemic heart disease
IPTW – inverse probability treatment weighting
IQR – interquartile range
MI – myocardial infarction
NHS – National Health Service
NICE – National Institute for Health and Care Excellence
NMA – network meta-analysis
NSAID – non-steroidal anti-inflammatory drug
ONS – Office for National Statistics
PS – propensity score
PYAR – person-years at risk
QOF – Quality and Outcomes Framework
RCT – randomised controlled trial
RR – risk ratio
SES – socioeconomic status
SGA – second generation antipsychotic
SMI – severe mental illness
SMR – standardised mortality ratio
T2DM – type 2 diabetes mellitus
THIN – The Health Improvement Network
TSH – thyroid stimulating hormone
UCL – University College London
UK – United Kingdom
US – United States
Contents

Thesis summary ..................................................................................................................... 16

Chapter 1  Introduction ......................................................................................................... 17

1.1  Summary ....................................................................................................................... 17

1.2  Bipolar disorder .......................................................................................................... 18

1.2.1  ICD-10 and DSM-5 criteria for bipolar disorder diagnosis ...................................... 18

1.3  Research priorities in bipolar disorder ......................................................................... 24

1.3.1  What are the rates of negative outcomes for people with bipolar disorder? ............ 24

1.3.2  Which medications reduce long-term adverse outcome rates in bipolar disorder? ... 25

1.4  Important long-term outcomes in bipolar disorder ....................................................... 25

1.4.1  Mortality in bipolar disorder .................................................................................... 26

1.4.2  Physical comorbidity in bipolar disorder .................................................................. 28

1.5  Maintenance treatment in bipolar disorder .................................................................... 30

1.5.1  Prescription rates for bipolar disorder in the UK ...................................................... 31

1.5.2  Lithium for the treatment of bipolar disorder ......................................................... 33

1.5.3  Valproate .................................................................................................................. 40

1.5.4  Olanzapine and Quetiapine ..................................................................................... 43

1.5.5  Head-to-head comparisons of lithium, valproate, olanzapine and quetiapine as maintenance mood stabilisers in bipolar disorder .............................................................. 47

1.6  Limitations of randomised controlled trials in bipolar disorder ................................. 48

1.6.1  Problems with trial designs in bipolar disorder ....................................................... 48

1.6.2  Recruitment problems in pharmacological trials for bipolar disorder ................... 49

1.6.3  Length of follow-up in pharmacological trials for bipolar disorder maintenance treatment .................................................................................................................. 49

1.6.4  Choosing meaningful outcomes for maintenance treatment trials ....................... 50

1.6.5  Lack of trials with active comparator groups in bipolar disorder maintenance treatment .................................................................................................................. 50

1.6.6  Overcoming the limitations of trials of maintenance treatment .............................. 51

1.7  Addressing the gaps in our understanding about long-term outcomes in bipolar disorder with observational data ................................................................. 51

1.7.1  Data Source: Primary care electronic health records ............................................... 51

1.7.2  Mortality and morbidity in bipolar disorder .............................................................. 52

1.7.3  Effectiveness and tolerability of maintenance medication ....................................... 52

1.7.4  Adverse events associated with maintenance medication in bipolar disorder ........ 53

1.7.5  Self-harm rates with different maintenance medications ......................................... 54
1.8 Objectives of the studies included in this thesis 54
1.8.1 Long-term outcomes in bipolar disorder 54
1.8.2 Effectiveness and adverse effects of maintenance treatments 55

Chapter 2 Methods – justification and overview 56
2.1 Summary 56
2.2 Systematic review and meta-analysis 57
  2.2.1 Summarising studies of mortality in people with bipolar disorder 57
  2.2.2 Network meta-analysis for comparing the efficacy and tolerability of maintenance mood stabilisers 57
2.3 Data Source for cohort studies – primary care electronic health records 58
  2.3.1 Comparison of The Health Improvement Network with other primary care electronic health records 58
  2.3.2 Coverage of The Health Improvement Network database 59
  2.3.3 Data recording in The Health Improvement Network 59
  2.3.4 Ethical approval for the cohort studies 60
2.4 Code lists for defining exposures, outcomes and covariates 60
2.5 Inclusion criteria for the cohort studies 61
  2.5.1 Primary care practice inclusion criteria 61
  2.5.2 Bipolar disorder and schizophrenia inclusion criteria 62
  2.5.3 General population frequency matching for the study of mortality and morbidity 63
2.6 Data extraction from The Health Improvement Network database 64
2.7 Definition of exposures, outcomes and other covariates 64
2.8 Benefits of using The Health Improvement Network for bipolar disorder outcomes and medication response research 66
  2.8.1 Size, variable recording, reliability, validity and generalisability of available data 66
2.9 Potential problems with the use of THIN for bipolar disorder outcomes and medication response research 67
  2.9.1 Causal inference in observational data 67
  2.9.2 Using propensity scores to achieve conditional exchangeability 68
  2.9.3 Confounding by indication in observational studies 69
  2.9.4 Other types of bias in studies using routine data 69
  2.9.5 Lack of recording of important covariates 70
2.10 Study design to minimise bias and confounding 73
  2.10.1 Mortality and morbidity studies 73
  2.10.2 Maintenance mood stabiliser studies 74
2.11 Approach to analysis to minimise bias and confounding

2.11.1 Propensity score estimation

2.11.2 Survival analyses regression techniques

Chapter 3 A systematic review and meta-analysis of mortality in individuals with bipolar disorder

3.1 Summary

3.1.1 Objective

3.1.2 Method

3.1.3 Results

3.1.4 Conclusion

3.2 Introduction

3.3 Methods

3.3.1 Identification of studies

3.3.2 Inclusion and exclusion criteria

3.3.3 Data extraction

3.3.4 Statistical methods

3.4 Results

3.5 Discussion

3.5.1 Main findings

3.5.2 Potential limitations

3.5.3 Conclusions

3.5.4 Implications of the findings from this systematic-review and meta-analysis

Chapter 4 Mortality and morbidity in individuals with bipolar disorder and schizophrenia, compared to the general population

4.1 Summary

4.1.1 Objective

4.1.2 Method

4.1.3 Results

4.1.4 Conclusion

4.2 Introduction

4.3 Methods

4.3.1 Study design and setting

4.3.2 Participants

4.3.3 Outcomes

4.3.4 Statistical analyses
Chapter 4

4.4 Results

4.4.1 Clinical and demographic features

4.4.2 All-cause mortality

4.4.3 Cardiovascular disease mortality and cardiovascular disease diagnoses

4.4.4 Suicide and self-harm

4.5 Discussion

4.5.1 Main findings

4.5.2 Comparison with existing literature

4.5.3 Potential explanations for these findings

4.5.4 Strengths and limitations

4.5.5 Conclusions

4.5.6 Implications of the findings from this study

Chapter 5

Systematic review and network meta-analysis comparing the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder

5.1 Summary

5.1.1 Objective

5.1.2 Method

5.1.3 Results

5.1.4 Conclusion

5.2 Introduction

5.3 Methods

5.3.1 Identification of studies

5.3.2 Inclusion and exclusion criteria

5.3.3 Data extraction

5.3.4 Statistical analysis

5.4 Results

5.5 Discussion

5.5.1 Main findings

5.5.2 Strengths and limitations

5.5.3 Conclusions

5.5.4 Implications of the findings from this network meta-analysis

Chapter 6

Comparison of the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder
Chapter 6  Comparison of the adverse renal, endocrine, hepatic and metabolic events during treatment with different maintenance mood stabiliser medications for bipolar disorder

6.1  Summary  ......................................................... 146
6.1.1  Objective ....................................................... 146
6.1.2  Methods ......................................................... 146
6.1.3  Results ......................................................... 146
6.1.4  Conclusion ..................................................... 147

6.2  Introduction ...................................................... 148

6.3  Methods .......................................................... 149
6.3.1  Study design & setting ...................................... 149
6.3.2  Participants .................................................... 149
6.3.3  Main outcome .................................................. 151
6.3.4  Observed pre-treatment variables for propensity score estimation ........................................... 151
6.3.5  Statistical Analysis ............................................ 152

6.4  Results ............................................................. 153

6.5  Discussion ........................................................ 161
6.5.1  Main findings .................................................. 161
6.5.2  Comparison with previous literature ...................... 161
6.5.3  Strengths and limitations .................................... 162
6.5.4  Conclusions ..................................................... 165
6.5.5  Implications of the findings of this study .................. 165

Chapter 7  Comparison of the adverse renal, endocrine, hepatic and metabolic events during treatment with different maintenance mood stabiliser medications for bipolar disorder .......................................................... 167

7.1  Summary .......................................................... 167
7.1.1  Objective ....................................................... 167
7.1.2  Methods ......................................................... 167
7.1.3  Results ......................................................... 167
7.1.4  Conclusions .................................................... 168

7.2  Introduction ...................................................... 170

7.3  Methods .......................................................... 171
7.3.1  Study design & Setting ...................................... 171
7.3.2  Participants .................................................... 171
7.3.3  Exposure ......................................................... 171
7.3.4  Main outcomes .................................................. 172
7.3.5  Propensity score estimation using observed pre-treatment variables ........................................... 173
7.3.6  Statistical Analysis ............................................ 174
Chapter 8  Comparison of self-harm, accidental injury and suicide in individuals with bipolar disorder during treatment with different maintenance mood stabiliser medications for bipolar disorder

8.1 Summary

8.1.1 Objective

8.1.2 Methods

8.1.3 Results

8.1.4 Conclusions

8.2 Introduction

8.3 Method

8.3.1 Study design, setting and participants

8.3.2 Exposure

8.3.3 Outcomes

8.3.4 Propensity score estimation

8.3.5 Statistical analysis

8.4 Results

8.4.1 Self-harm

8.4.2 Accidental injury

8.4.3 Suicide

8.5 Discussion

8.5.1 Main findings

8.5.2 Comparison with previous literature

8.5.3 Strengths and limitations

8.5.4 Conclusions

8.5.5 Implications of the findings of this study

Chapter 9  Discussion, implications and conclusions

9.1 Summary

9.2 Key findings
9.2.1 Mortality and morbidity in bipolar disorder 210
9.2.2 Maintenance treatment for bipolar disorder 212

9.3 Implications of findings from my mortality and morbidity studies 214
9.3.1 Key points for clinicians and patients 214
9.3.2 Key points for policy makers 215
9.3.3 Key points for researchers 215
9.3.4 All-cause mortality in people with bipolar disorder 216
9.3.5 Cause-specific mortality in people with bipolar disorder 217
9.3.6 Why is there a still “mortality gap”? 219
9.3.7 What should be done to address the “mortality gap”? 221

9.4 Implications of findings from my maintenance treatment studies 224
9.4.1 Key points for clinicians and patients 224
9.4.2 Key points for policy makers 225
9.4.3 Key points for researchers 225
9.4.4 Effectiveness and tolerability of maintenance mood stabiliser medications 226
9.4.5 Adverse events during maintenance mood stabiliser treatment 227
9.4.6 Self-harm, accidental injury and suicide during treatment with maintenance mood stabiliser medications 232

9.5 Strengths and limitations of studies making up this thesis 234
9.5.1 Use of electronic health records 234
9.5.2 Diagnostic validity of bipolar disorder in THIN 235
9.5.3 Comparing characteristics of my cohort of people with bipolar disorder with other cohort studies 237
9.5.4 Defining drug exposure using EHR 238
9.5.5 Using drug monotherapy to investigate effects 239
9.5.6 Follow-up time for patients with bipolar disorder in THIN 240
9.5.7 Validity of outcomes used in this thesis 240
9.5.8 Comparing bipolar disorder mortality with general population rates 241
9.5.9 Approach to handling missing data in THIN 242
9.5.10 Unmeasured and residual confounding 244
9.5.11 Are propensity score approaches superior to traditional techniques? 246
9.5.12 Limitations related to study design 246

9.6 Bipolar disorder now and in the future 247
9.6.1 Is bipolar disorder becoming more common? 247
9.6.2 Underfunded and under researched 248

9.7 Future questions for bipolar disorder research 249
9.7.1 Can early diagnosis improve functioning and reduce adverse long-term outcomes? 249
9.7.2 Can staging models improve outcomes? ................................................................. 249
9.7.3 How do we better use the drugs that we have? .................................................... 250
9.7.4 Could a better understanding of mechanism of lithium help develop new treatments? 251

9.8 Other research directions ....................................................................................... 252
  9.8.1 Study design ......................................................................................................... 252
  9.8.2 Electronic health records .................................................................................... 253

9.9 Dissemination ........................................................................................................... 254

9.10 Conclusions ............................................................................................................ 255

Figures

Figure 1.i Example of a possible trajectory of mood episodes in bipolar disorder .............. 19
Figure 1.ii ICD-10 criteria for the diagnosis of hypomania or mania .................................. 20
Figure 1.iii ICD-10 categories for manic episodes and bipolar disorder ......................... 22
Figure 1.iv Prescribing for bipolar disorder in the UK: Proportion of time spent on treatment with lithium, valproate and any second generation antipsychotic .............................................. 32
Figure 1.v Percentage of patients with clinically significant weight gain (>7%) in short-term (3-8 weeks) placebo-controlled RCTs. .............................................................................................................. 45
Figure 2.i Flow diagram of included patients for each study ............................................. 65
Figure 2.ii Potential distribution of propensity scores for two drugs ............................... 79
Figure 3.i Flow diagram of the published articles evaluated for inclusion in this meta-analysis...... 89
Figure 3.ii SMR of all-cause mortality ............................................................................. 96
Figure 4.i Flow diagram of included patients .................................................................... 109
Figure 4.ii All-cause mortality rate in bipolar disorder and schizophrenia 2000-2014 ............ 114
Figure 4.iii All-cause mortality adjusted hazard ratio for bipolar disorder and schizophrenia compared to the general population 2000-2014 ......................................................... 116
Figure 5.i Flow diagram of the trials evaluated for inclusion in this systematic review and NMA 135
Figure 5.ii Network plot of included trials: Relapse and reoccurrence ............................ 138
Figure 5.iii Network plot of included trials: Discontinuation ......................................... 138
Figure 5.iv Forest plot for relapse or reoccurrence ....................................................... 140
Tables

Table 1.i Studies of lithium’s effects on eGFR published since 2012 .............................................. 35
Table 1.ii Summary of recognised adverse effects from long-term use ............................................. 38
Table 2.i Diagnostic stability of bipolar disorder and schizophrenia .................................................... 63
Table 3.i Studies include in the meta-analysis .................................................................................. 93
Table 3.ii Summary standardised mortality ratios for all-cause and cause specific mortalities ............. 97
Table 3.iii Summary SMRs by subgroup for all-cause mortality, and natural and unnatural death 98
Table 4.i Cohort characteristics ........................................................................................................ 112
Table 4.ii All-cause mortality ........................................................................................................... 115
Table 4.iii Cardiovascular disease mortality and cardiovascular disease diagnoses .......................... 119
Table 4.iv Suicide and self-harm ........................................................................................................... 121
Table 5.i Included trials .................................................................................................................. 136
Table 5.ii Relative risk of relapse or reoccurrence (orange) and all-cause discontinuation (green) according to NMA ......................................................................................................................... 139
Table 5.iii Probability of each treatment being ranked best-worst; for preventing relapse/reoccurrence, and for lowest all-cause discontinuation ........................................................................................................... 142
Table 6.i Characteristics of patients with bipolar disorder prescribed lithium, valproate, olanzapine or quetiapine monotherapy ........................................................................................................... 155
Table 6.ii Rates of treatment failure by drug .................................................................................. 156
Table 6.iii Supplementary analyses using PS adjusted model ......................................................... 160
Table 7.i Patient characteristics .................................................................................................... 176
Table 7.ii Patients included for each outcome, N (%) .................................................................... 177
Table 7.iii Adverse effects during maintenance treatment ............................................................ 177
Table 7.iv Median number (and interquartile range) of tests per year of drug exposure in patients included in analyses .............................................................................................................. 185
Table 7.v Sensitivity analyses to account for missing blood tests by 1) including all individuals and 2) performing inverse probability weighting .................................................................................. 185
Table 8.i Patient characteristics .................................................................................................... 200
Table 8.ii Rates of self-harm, accidental injury, and suicide by mood stabiliser ......................... 201
Thesis summary

This thesis has two overarching aims: To define the mortality and morbidity in people with bipolar disorder (BPD), relative to other individuals with schizophrenia and the general population, and to understand the longer-term effects of maintenance medication on relapse/reoccurrence and adverse events in individuals with BPD.

These aims were refined through completion of a systematic review and meta-analysis of premature mortality in BPD, and a network meta-analysis of efficacy of maintenance medication for BPD. Findings from these reviews informed a number of cohort studies using longitudinal data from United Kingdom (UK) primary care electronic health records (EHRs). Completed studies examine rates of all-cause and cause specific (suicide, cardiovascular) mortality in BPD relative to people with schizophrenia and the general population and compare rates of cardiovascular disease (CVD) diagnosis, and self-harm in these groups. Further studies examine rates of monotherapy treatment failure, rates of adverse physical health events and rates of self-harm, accidental injury and suicide in patients with BPD prescribed one of the four most common maintenance medications for mood stabilisation (lithium, valproate, olanzapine or quetiapine). EHRs provide ways of addressing these questions that have not been previously possible, with long follow-up times, and large, generalisable cohorts of individuals with BPD.

Finally, the implications of findings from these studies in terms of clinical practice, policy and patient impact are explored, set in the context of previous research. Overall strengths and limitations are discussed, and directions for future research are set out.
Chapter 1 Introduction

1.1 Summary

This chapter sets out the context for my thesis. It provides an overview of the characteristics of BPD, including potential international diagnostic differences and comparisons with other mental disorders, such as depression and schizophrenia. It goes on to outline the key research priorities that are addressed by this thesis and the current United Kingdom context in which these sit. Aims and objectives for each study are introduced, and existing literature relating to each objective is described and discussed. Previous literature on mortality and morbidity in BPD is summarised. Commonly used drug treatments in BPD are discussed, including the potential benefits and harms of each. Limitations of randomised controlled trials (RCTs) of maintenance treatments are examined in this context.
1.2 Bipolar disorder

BPD is a life-long, recurrent, episodic illness with high rates of hospitalisation, suicide and comorbidity (Saunders and Goodwin, 2010). It affects at least 1% of the UK population (Das Gupta and Guest, 2002). In 2013 BPD was the sixteenth most common cause of years lived with disability in the world; similar to more prevalent conditions such as asthma or Alzheimer’s disease (Ferrari et al., 2016). It is projected to cost the UK over £6.5 billion per year by 2017 (McCrone et al., 2008). Despite this, BPD research is significantly underfunded, even when compared to other severe mental illness (SMI) such as schizophrenia (Young, 2006).

1.2.1 ICD-10 and DSM-5 criteria for bipolar disorder diagnosis

The World Health Organisation’s International Classification of Disease, 10th revision (ICD-10) and the American Psychiatric Associations Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) are the most commonly used diagnostic classification systems in psychiatry (American Psychiatric Association, 2013, World Health Organisation, 1992). Both define BPD, previously known as manic depressive illness, as a severe chronic mood disorder characterised by episodes of mania, hypomania and alternating (or intertwining) episodes of depression (Figure 1.1).
1.2.1.1 Manic and hypomanic episodes

Mania and hypomania are states of elevated mood and increased motor drive (Grande et al., 2015) (Figure 1.i). ICD-10 criteria for hypomania are that the mood is elevated (or irritable) to a degree that is abnormal for the individual concerned for at least four consecutive days. At least three of the symptoms shown in Figure 1.ii must be present, leading to some interference in personal functioning. Mania is defined by similar, though more severe, symptoms (Figure 1.ii). In mania the mood change must be sustained for at least seven days (unless it is severe enough to result in hospitalisation) and lead to severe interference with personal functioning.
Psychotic symptoms in mania will most commonly be mood congruent, such as grandiose delusions or auditory hallucinations telling the individual that they have superhuman powers. However mood incongruent psychotic symptoms also occur, such as affectively neutral topics, persecutory delusions or delusions of reference.

1.2.1.2 Depressive episodes

Patients with BPD spend a substantial proportion of time suffering from syndromal or sub-syndromal depressive symptoms (Grande et al., 2015) (Figure 1.i). Major depressive episodes in BPD are similar to those experienced in unipolar major depression (National Institute for Health and Care Excellence, 2014). Patients suffer depressed mood and experience profound loss of interest in activities, coupled with other symptoms such as fatigue, weight loss or gain, difficulty sleeping or staying awake, psychomotor slowing, feelings of worthlessness, excessive guilt and suicidal thoughts or actions (World Health Organisation, 1992). For patients presenting with a first episode of depression, it may not
be possible to distinguish between those who will go on to suffer recurrent unipolar depression and those who will develop BPD. However, evidence suggests there may be subtle differences between bipolar and unipolar depression. In particular, depression in the course of BPD may be more likely to show signs of psychomotor retardation, to have melancholic features (such as feelings of worthlessness and marked anhedonia), to show features of atypical depression (such as hypersomnia and weight gain) and to show psychotic features – especially in young people (Bowden, 2005, Forty et al., 2008, Goldberg et al., 2001). Patients experiencing a first episode of depression who display these features and have a family history of BPD may be at increased risk of developing BPD (McGuffin et al., 2003).

Sub-syndromal depressive symptoms are common in patients with BPD and are often associated with significant interpersonal or occupational disability (Bonnin et al., 2012). The management of these chronic, low-grade depressive symptoms is therefore of major importance, but is also a substantial treatment challenge.

1.2.1.3 Differences in diagnostic criteria between ICD and DSM

Some key differences exist in the ICD and DSM diagnostic criteria for BPD (American Psychiatric Association, 2013, World Health Organisation, 1992). The most notable is the DSM’s differentiation of the disorder into bipolar I and bipolar II subtypes. DSM requires at least one manic episode to have occurred for a diagnosis of bipolar I and at least one hypomanic episode for bipolar II. Although major depressive episodes will typically occur, they are not needed for the diagnosis to be made. Conversely ICD does not include the I/II differentiation, although clinicians using ICD will commonly refer to the diagnosis in these terms, and much of the research literature is based on this differentiation (National Institute for Health and Care Excellence, 2014). It is likely that the I/II nomenclature will be
adopted in ICD-11 (de Dios et al., 2014). ICD-10 requires that an individual experiences two mood episodes (with at least one being hypomania or mania) before a diagnosis of BPD be made. As such ICD criteria are stricter (Figure 1.iii). Criteria for defining mania versus hypomania are also stricter in ICD-10. In DSM mania describes all elevated mood states “with functional impairment” and is qualified by severity, i.e., mild, moderate or severe (with or without psychosis). Hypomania is defined as any elevated mood state “without social or occupational dysfunction” lasting at least four days (American Psychiatric Association, 2013), with a number of researchers suggesting that this period should be shortened to two days (Angst et al., 2003). In ICD-10 the boundary between hypomania and mania is defined by the effect on social functioning. The threshold for a diagnosis of mania is made when the individual experiences a “severe or complete” disruption of work and social activity (World Health Organisation, 1992).

Figure 1.iii ICD-10 categories for manic episodes and bipolar disorder

<table>
<thead>
<tr>
<th>F30 Manic episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>F30.0 Hypomania</td>
</tr>
<tr>
<td>F30.1 Mania without psychotic symptoms</td>
</tr>
<tr>
<td>F30.2 Mania with psychotic symptoms</td>
</tr>
<tr>
<td>.20 With mood-congruent psychotic symptoms</td>
</tr>
<tr>
<td>.21 With mood-incongruent psychotic symptoms</td>
</tr>
<tr>
<td>F30.8 Other manic episodes</td>
</tr>
<tr>
<td>F30.9 Manic episode, unspecified</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F31 Bipolar affective disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>F31.0 Bipolar affective disorder, current episode hypomanic</td>
</tr>
<tr>
<td>F31.1 Bipolar affective disorder, current episode manic without psychotic symptoms</td>
</tr>
<tr>
<td>F31.2 Bipolar affective disorder, current episode manic with psychotic symptoms</td>
</tr>
<tr>
<td>.20 With mood-congruent psychotic symptoms</td>
</tr>
<tr>
<td>.21 With mood-incongruent psychotic symptoms</td>
</tr>
<tr>
<td>F31.3 Bipolar affective disorder, current episode mild or moderate depression</td>
</tr>
<tr>
<td>.30 Without somatic syndrome</td>
</tr>
<tr>
<td>.31 With somatic syndrome</td>
</tr>
<tr>
<td>F31.4 Bipolar affective disorder, current episode severe depression without psychotic symptoms</td>
</tr>
<tr>
<td>F31.5 Bipolar affective disorder, current episode severe depression with psychotic symptoms</td>
</tr>
<tr>
<td>.50 With mood-congruent psychotic symptoms</td>
</tr>
<tr>
<td>.51 With mood-incongruent psychotic symptoms</td>
</tr>
<tr>
<td>F31.6 Bipolar affective disorder, current episode mixed</td>
</tr>
<tr>
<td>F31.7 Bipolar affective disorder, currently in remission</td>
</tr>
<tr>
<td>F31.8 Other bipolar affective disorders</td>
</tr>
<tr>
<td>F31.9 Bipolar affective disorder, unspecified</td>
</tr>
</tbody>
</table>

*From (World Health Organisation, 1992).*
These differences will clearly lead to more individuals being diagnosed under DSM criteria than ICD, and this may partially explain worldwide differences in prevalence and incidence estimates of BPD (Ferrari et al., 2011, Hardoon et al., 2013, Merikangas et al., 2007). Data for this thesis are from the UK, and as such, individuals identified as having BPD should be regarded as fulfilling ICD criteria for diagnosis. Additionally, codes available in primary care EHR do not specifically identify bipolar I versus bipolar II.

1.2.1.4 Bipolar disorder, depression and schizophrenia as separate diagnostic entities

BPD appears to share more similarities with schizophrenia than major depressive disorder, despite the commonality of depressive symptoms. There is growing evidence of an aetiological and prognostic overlap between BPD and schizophrenia (Laursen et al., 2007). Genetic evidence supporting non-independence of BPD and schizophrenia has come from family, twin and linkage studies and from studies of individual genes (Craddock et al., 2006, 2009). BPD and schizophrenia also share neurotransmitter dysfunction similarities and there is a treatment overlap in terms of response to second generation antipsychotic (SGA) medication (dopamine receptor antagonists) (Möller, 2003).

Traditionally, individuals with BPD have been considered to have better outcomes than those with schizophrenia. This appears to be due to assumptions about seemingly normal inter-episode function and an absence of significant cognitive impairment (Zarate Jr et al., 2000). However, more recent studies have highlighted problems with psychosocial dysfunction in euthymic individuals with BPD, suggesting there is a discontinuity between clinical and functional outcomes (Martinez - Aran et al., 2007). The complex presentation of BPD may have hindered the identification of this impairment. Reduced health related
quality of life and functioning is now apparent, however, the impact of this impairment on adverse outcomes relative to schizophrenia is still not well understood.

1.3 Research priorities in bipolar disorder

There is remarkably little epidemiological evidence regarding the prognosis and treatment of bipolar disorder. A number of fundamental epidemiological questions about BPD, its treatment, and the resultant outcomes remain poorly answered. This thesis aims to address two key unanswered questions: 1) what are the rates of negative outcomes in BPD? And 2) which medications reduce adverse outcome rates? Getting answers to these questions is important to individuals with BPD and to clinicians providing their care and will facilitate informed decision making about available treatment options. Answers will also be important to the National Health Service (NHS) in informing the best use of resources.

These clinical research questions are identified as priorities by the National Institute for Health and Care Excellence (National Institute for Health and Care Excellence, 2014), and the James Lind Alliance Priority Setting Partnership (Robotham et al., 2016) in the UK. For example, included in the James Lind Alliance top 10 for BPD are questions about effective medication, managing suicide risk, side-effects and adverse effects. They have also been identified as priorities by patients, clinicians and researchers in North America (Crowley et al., 2014, Michalak et al., 2012), Australia (Banfield et al., 2011) and across Europe (van Os and Wahlbeck, 2014).

1.3.1 What are the rates of negative outcomes for people with bipolar disorder?

There have been very few studies of the rates of mortality and comorbidities in BPD using nationally representative community based cohorts. Existing studies have relied mainly on hospital discharge data or case-control designs. The availability of primary care data
reflecting outpatient diagnoses is important as it allows inclusion of people not in contact with secondary care; this might be because of milder or more stable illness, or because of a refusal to engage with psychiatric services. This will result in more generalisable rate estimates.

1.3.2 Which medications reduce long-term adverse outcome rates in bipolar disorder?
The NICE guidelines identify a need to understand the relative benefits in terms of quality of life of individuals with BPD whilst taking different maintenance medications (National Institute for Health and Care Excellence, 2014). Lithium, valproate, olanzapine and quetiapine appear to reduce the risk of relapse when used long-term, but these four medications are also associated with a number of side effects, some of which can adversely affect physical health and may contribute to premature mortality (Correll et al., 2015, Joukamaa et al., 2006). Current research investigating the relative effectiveness and tolerability of these drugs and their impact on quality of life is insufficient due to both its scarcity and low quality.

1.4 Important long-term outcomes in bipolar disorder
There are a limited number of studies that examine long-term outcomes. We know patients with BPD are frequently hospitalised (between 10% and 50% per year (Adler et al., 2012, Tohen et al., 1990)), 15% die by suicide, and approximately half will attempt suicide (Baldessarini and Tondo, 2003, Tondo et al., 2003). Patients with BPD are also more likely to suffer medical and psychiatric comorbidities and have excess all-cause mortality (Crump et al., 2013, Krishnan, 2005). However, limited numbers of participants and follow-up periods in these studies mean that we have little understanding regarding long-term outcomes for people with bipolar disorder in the UK.
1.4.1 Mortality in bipolar disorder

Research has suggested that BPD is associated with premature mortality. Where previously it was believed this was mostly attributable to unnatural causes such as suicide, homicide and accidents, it has also been shown that patients with BPD are at risk of premature death from a range of medical illnesses (Roshanaei-Moghaddam and Katon, 2009). In 1998, Harris and Barraclough reviewed mortality in all mental disorders (Harris and Barraclough, 1998); six studies contributed to their meta-analysis of mortality in BPD. A more recent review published in 2009 included 13 studies of mortality from general medical illnesses (Roshanaei-Moghaddam and Katon, 2009). Since these publications, a number of large database studies have derived elevated mortality estimates for BPD; however, there are no recent estimates from the UK.

I conducted a systematic review and meta-analysis of all-cause and cause specific mortalities (Chapter 3). Cause specific mortalities were grouped as natural and unnatural. Natural deaths were then further divided into circulatory causes (for example myocardial infarction (MI) and cerebrovascular events (CVE)), respiratory causes (such as chronic obstructive pulmonary disease (COPD)), infectious causes (such as pneumonia or sepsis) and deaths from cancer. Unnatural deaths were divided into deaths by suicide and other violent causes.

The most commonly used measure of effect in assessing mortality is the standardised mortality ratio (SMR). In general, SMRs remove confounding by age and sex only, and so they may be limited when comparing a patient group (such as those with BPD) who do not match the general population structure in terms of social deprivation or physical health at baseline. In addition, the SMR does not allow direct comparisons between study groups: for example comparisons of mortality between BPD and schizophrenia.
From this systematic review, particular areas were identified for further investigation and are addressed in Chapter 4 of the thesis: all-cause mortality, CVD mortality, suicide, CVD and self-harm.

1.4.1.1 All-cause mortality in bipolar disorder

Whilst it there is a recognition that mortality is elevated in BPD (Hoang et al., 2011, Roshanaei-Moghaddam and Katon, 2009), we do not have contemporary information regarding mortality trends in people with BPD. In recent years, in the UK, a number of attempts have been made to understand and address the disparity in mortality between individuals with SMI (including BPD) and the general population. These include target setting for general practitioners (Colton and Manderscheid, 2006), establishing early intervention in psychosis services (Edwards and McGorry, 2002), independent review (Schizophrenia Commission, 2012), and Government policy and spending initiatives (Department of Health, 2011, 2014). It remains unclear if these changes have resulted in improved life expectancy for individuals with BPD relative to the general population.

1.4.1.2 Cardiovascular disease mortality in bipolar disorder

CVD has previously been shown to be the leading cause of death in individuals with BPD (Laursen et al., 2013, Roshanaei-Moghaddam and Katon, 2009, Weiner et al., 2011). However there are few estimates of CVD mortality compared to other SMIs, such as schizophrenia, and studies that investigate the potential role of medication, lifestyle factors and other comorbidities are even more limited (Osborn et al., 2007). Studies that do exist suggest that elevated rates of CVD death are not explained only by medication, smoking or physical inactivity (Kilbourne et al., 2009, Osborn et al., 2007). One previous UK EHR study has investigated rates of CVD death in SMI up until 2002, including people with BPD and schizophrenia (Osborn et al., 2007). This study found a 3-fold increase in CVD deaths in
those aged 18 to 49 and a 2-fold increase in those aged 50 to 75 for any SMI diagnosis. Overall rates for specific SMI were not reported. However, in the 18 to 49 year old group the point estimate suggested BPD patients had elevated CVD mortality, but this could not be confirmed because of wide confidence intervals (CIs), due to a low event rate (HR 2.13; 95% CI 0.77 to 5.93). In the 50 to 75 year age group individuals with BPD had an elevated CVD rate, similar to those with schizophrenia (HR 1.52; 95% CI 1.18–1.95 and HR 1.96; 95% CI 1.63 to 2.35 respectively).

1.4.1.3 Suicide in bipolar disorder

Suicide is a leading public health priority in BPD; 10 to 15% of individuals with BPD will die by suicide (Hawton et al., 2005). However, it remains unclear if BPD conveys a bigger risk than other SMI, with previous research inconsistent in this regard (Mortensen et al., 2000, Osborn et al., 2008, Tidemalm et al., 2008). Contemporary estimates of suicide rate in BPD are also rare. Until recently, research has also been limited by relatively small sample sizes, short follow-up periods and lack of data from representative community samples, and because of the rarity of suicide, identifying risk factors in BPD has been challenging (Schaffer et al., 2015).

1.4.2 Physical comorbidity in bipolar disorder

As with studies of mortality, BPD has been relatively neglected in terms of describing and quantifying morbidity risks (McIntyre et al., 2007, Weiner et al., 2011). Studies of physical comorbidities in representative population based cohorts of individuals with BPD have only recently begun to be carried out (Crump et al., 2013, McIntyre et al., 2007, Weiner et al., 2011), but gaps in the knowledge around specific comorbidities remain. The longitudinal cohort study by Crump et al. is the only one to have examined the association between BPD and physical comorbidities in a representative sample, rather than relying on psychiatric
inpatient data (Crump et al., 2013). Patients with BPD have been found to have more physical health problems than the general population (Smith et al., 2013) and reduced access to appropriate treatment (Crump et al., 2013).

1.4.2.1 Cardiovascular disease in bipolar disorder
Evidence from a 2011 review suggests MI, CVE and ischemic heart disease (IHD) risk are all elevated in BPD (Weiner et al., 2011), however included studies are now dated (i.e., cover time before the introduction of atypical antipsychotic medication), and involve populations limited in size and generalisability. It is unclear if the high prevalence of CVD is related solely to increased traditional risk factors (such as smoking, obesity and type 2 diabetes mellitus (T2DM)), or if it is related to unrecognised emerging risk factors such as inflammation, abnormal metabolism and renal insufficiency (Joynt et al., 2003, Kupfer, 2005). There is evidence for reduced appropriate treatment for individuals with BPD and CVD (Laursen et al., 2009, Smith et al., 2013), but it is unclear if this remains the case following a number of high profile healthcare interventions (Department of Health, 2011, Doran et al., 2011, Roland, 2004).

1.4.2.2 Self-harm and accidental injury in bipolar disorder
In UK based research, self-harm refers to self-poisoning or self-injury regardless of apparent intent (Kapur et al., 2013). ‘Deliberate self-harm’ superseded previously used terms for non-fatal suicidal behaviour such as ‘parasuicide’ and ‘attempted suicide’ in the 1970s, and more recently the ‘deliberate’ prefix has been dropped because motivation or intent is not always clear. In the United States, research commonly refers to non-suicidal self-injury (NSSI) and this was formalised as a distinct disorder in DSM-5 (American Psychiatric Association, 2013). Concerns have been raised about the prefix ‘non-suicidal’ as NSSI is a risk factor for suicide, and the diagnosis excludes self-poisoning (i.e., all self-poisoning is
assumed to have suicidal intent) (Kapur et al., 2013). This thesis therefore uses the UK definition of self-harm. Self-harm is a major cause of morbidity in BPD (Singhal et al., 2014), and is strongly associated with suicide (Owens et al., 2002). As with suicide, rates have been poorly defined, relative to the general population and other SMI.

Risk of accidental injury has also been understudied in BPD, despite deaths from unintentional injury being approximately six times higher in BPD than in the general population (Hoang et al., 2011). Although unintentional injuries are often recorded in drug trials of BPD, they are rarely reported as important outcomes (Matson et al., 2006). Observational studies (i.e., studies where the exposure allocation is not under the control of the researcher) of accidental injury are rare (Khalsa et al., 2008). Accidental injury is thought to be associated with mania or hypomania, whereas self-harm may be related to depressive episodes (Khalsa et al., 2008).

1.5 Maintenance treatment in bipolar disorder

“Maintenance treatment” is not well defined in the literature (Goodwin et al., 2016, Grunze et al., 2013, National Institute for Health and Care Excellence, 2014). NICE state maintenance treatment means “long-term” treatment (National Institute for Health and Care Excellence, 2014), and there is no guidance on when to stop treatment. As such, it is often taken to mean anything from post-remission to lifetime prescribing (depending on if you are planning an RCT or treating a patient clinically). Guidelines suggest maintenance treatment should be considered after each mood episode, and especially for all patients experiencing a manic episode (including their first manic episode) (Goodwin et al., 2016, National Institute for Health and Care Excellence, 2014). Therefore, the large majority of patients with bipolar I and a number with bipolar II should receive maintenance treatment. The aim of maintenance treatment is to prevent re-emergence of symptoms, including
subsyndromal symptoms, and to limit illness progression (Grunze et al., 2013). Until September 2014 NICE recommended three drugs as possible first-line treatments for BPD maintenance: lithium, valproate and olanzapine, with quetiapine being recommended in predominantly depressive BPD (National Institute for Health and Care Excellence, 2006). In the updated guideline (National Institute for Health and Care Excellence, 2014), lithium is now first-line. This change was based on expert consensus and a number of trials conducted since 2006 (Amsterdam and Shults, 2010, Geddes et al., 2010, Licht et al., 2010, Weisler et al., 2011). This change is also supported by two recent meta-analyses (Miura et al., 2014, Severus et al., 2014) and more recently by the updated British Association for Psychopharmacology guidelines for treating BPD (Goodwin et al., 2016). In Chapter 5 I complete a systematic review and network meta-analysis of all maintenance treatment trials that include comparisons of lithium, valproate, olanzapine, quetiapine and placebo.

NICE continue to highlight the need for further evidence about effective maintenance medication (National Institute for Health and Care Excellence, 2014), and evidence regarding effectiveness and safety of long-term maintenance treatment of BPD remains sparse (Goodwin, 2009, Goodwin et al., 2016). Head-to-head comparisons of maintenance medications are rare. There is also no clear evidence that any one of the recommended medications will prove more effective for a particular patient with a particular illness pattern.

1.5.1 Prescription rates for bipolar disorder in the UK

Lithium, valproate, olanzapine and quetiapine are the most commonly prescribed in the UK (Hayes et al., 2011) and are the four maintenance medications investigated in Chapter 6, Chapter 7 and Chapter 8. Patterns of prescribing up to 2009 reflected previous NICE guidance (National Institute for Health and Care Excellence, 2006). After their introduction,
there was a rapid increase in the use of valproate, olanzapine and quetiapine, whereas lithium has remained relatively stable (Figure 1.iv).

Figure 1.iv Prescribing for bipolar disorder in the UK: Proportion of time spent on treatment with lithium, valproate and any second generation antipsychotic

Increasing age and an earlier date of starting treatment were the only factors that were associated with being treated with lithium rather than other drugs, suggesting that individuals with new prescribing for BPD in recent years will most likely receive valproate or an antipsychotic medication. Guidelines around avoiding valproate in women of childbearing potential were not being followed, with almost one in three being prescribed this drug (Hayes et al., 2011). This guidance was reinforced by NICE and the Medicines and Healthcare products Regulatory Agency in 2016.
1.5.2 Lithium for the treatment of bipolar disorder

1.5.2.1 Effectiveness of lithium as a maintenance mood stabiliser

Lithium was introduced by John Cade in 1949 and has been in clinical use for more than 50 years (Abou-Saleh and Coppen, 1986). Lithium is widely used in the long-term treatment of affective disorders and its use has been considered well established (Geddes and Miklowitz, 2013, Licht, 2012). However, its efficacy has been repeatedly questioned (Blackwell and Shepherd, 1968, Moncrieff, 1997), there are few trials that compare it with an alternative active treatment (Miura et al., 2014), and its benefits are restricted by its side effect profile and a narrow therapeutic window (Geddes and Miklowitz, 2013). Because of these factors, alternative maintenance treatments have become commonly used (Hayes et al., 2011). Overall, RCT evidence with up to 2 years follow-up suggests lithium is superior to placebo in treating both manic and depressive polarities of relapse, but it is poorly tolerated (Goodwin et al., 2016, Miura et al., 2014).

1.5.2.2 Adverse effects from long-term use of lithium

A number of long-term adverse effects of lithium have been identified since its use as a mood stabiliser became established in the 1970s (Bech, 2006), but it is only recently that they have begun to be characterised and quantified (Clos et al., 2015, Close et al., 2014, Kessing et al., 2015, McKnight et al., 2012, Murrur et al., 2015, Shine et al., 2015). Lithium’s adverse effects include renal, thyroid, and parathyroid dysfunction. Lithium is also recognised to cause weight gain, but the risk of weight gain relative to other potential maintenance therapies has not been widely investigated (McKnight et al., 2012).

1.5.2.3 Renal effects of long-term lithium use

In 2012, McKnight et al. identified 30 studies (no RCTs) which investigated lithium effect on estimated glomerular filtration rate (eGFR) or urine concentrating ability. Only nine case-
control studies could be used in the meta-analysis, the remainder being uncontrolled cohort studies with no data on within-patient changes over follow-up. The pooled estimate from these studies was a 6.22 mL/min (95% CI -14.65 to 2.20) reduction in eGFR for people prescribed lithium (which is not statistically or clinically significant). Since publication of this meta-analysis, large cohort and case-control studies have suggested that whilst lithium use appears to result in a decline in eGFR, clinically significant severe or end-stage chronic kidney disease (CKD) remains rare (Aiff et al., 2015, Castro et al., 2015, Clos et al., 2015, Close et al., 2014, Shine et al., 2015). Table 1.i describes the main characteristics of these studies.

Often these studies have been limited by lack of active comparator groups, confounding by indication (that is; the chance of receiving treatment is related to severity and other baseline factors), surveillance biases (follow-up is differential and is related to exposure or outcome), and little information on potential confounders. It has been argued by others that modern lithium treatment regimens, involving regular and frequent monitoring and individually adjusted dosing, have eradicated lithium induced renal failure by avoiding toxicity, and keeping plasma levels below 0.8 mmol/L (Aiff et al., 2014b). Lithium has a very narrow therapeutic range where the majority of patients will experience toxic effects above 1.5 mmol/L (Goodwin et al., 2016). The theory that toxicity alone causes renal failure is based on evidence that periods of toxicity are related to decline in eGFR (Table 1.i) and end-stage renal disease (Aiff et al., 2014a).
### Table 1: Studies of lithium’s effects on eGFR published since 2012

<table>
<thead>
<tr>
<th>Authors</th>
<th>N</th>
<th>Comparison group</th>
<th>Follow-up</th>
<th>Controlled for confounding</th>
<th>Key limitations</th>
<th>Statistically significant decline in eGFR compared to comparison</th>
<th>Toxicity related to decline in eGFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aiff et al. (2015)</td>
<td>630 lithium exposed</td>
<td>Cohort/case-series – no comparator group</td>
<td>Max=30 years, min=10 years</td>
<td>No</td>
<td>No adjustment, no comparison group, missing exposure data</td>
<td>Decline observed</td>
<td>Yes</td>
</tr>
<tr>
<td>Castro et al. (2015)</td>
<td>1145 cases, 4306 controls</td>
<td>Case-control – CKD stage 3 vs. no CKD stage 3</td>
<td>Max=9 years, min=1 day (median 178 days)</td>
<td>Age, sex, ethnicity, Charlson index, insurance, hypertension, smoking, diabetes, schizophrenia</td>
<td>Case-control, no report of change in eGFR, short follow-up</td>
<td>Not reported</td>
<td>Yes</td>
</tr>
<tr>
<td>Clos et al. (2015)</td>
<td>305 lithium exposed, 815 comparator drugs</td>
<td>Cohort – patients prescribed mood stabiliser or antipsychotic</td>
<td>Max=12 years (mean=55 months)</td>
<td>Age, sex, baseline eGFR, genitourinary disease, hypertension, diabetes, NSAID, β-blocker, ACEI, toxicity</td>
<td>Potential confounding by indication, surveillance bias, young cohort (&lt;65), short follow-up</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Close et al. (2014)</td>
<td>2496 lithium exposed, 3864 unexposed</td>
<td>Patients with BPD with no lithium exposure</td>
<td>Max=18 years (median 5.4 years)</td>
<td>Age, sex, alcohol, smoking, BMI, IMD, CVD, liver failure, diabetes, cancer, hypertension, antipsychotics, mood stabilisers, β-blocker, diuretics, ACEI, paracetamol, NSAID</td>
<td>Potential confounding by indication, surveillance bias</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Shine et al. (2015)</td>
<td>4678 lithium exposed, 689228 unexposed</td>
<td>All individuals receiving eGFR tests not taking lithium</td>
<td>Max=28 years (median 3 years)</td>
<td>Age, sex, diabetes</td>
<td>Potential confounding by indication, surveillance bias, heterogeneous population, limited adjustment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
1.5.2.4 Endocrine effects of long-term lithium use

Regarding the effects of lithium on thyroid function, four RCTs with thyroid outcomes were identified by McKnight et al. (2012). These reported that, during up to a year of follow-up, 4% of patients prescribed lithium developed hypothyroidism, compared to none taking placebo. Their meta-analysis of case-controls studies found an odds ratio (OR) of 5.78 (95% CI 2.00 to 16.67) for individuals prescribed lithium compared to placebo. The quality of evidence is limited by lack of active comparator groups and low power. Observational studies had similar limitations to those examining renal outcomes. Studies published since 2012 have been equally limited. There has been one major high-quality US EHR study which found similarly elevated rates of hypothyroidism across all commonly used mood stabiliser medication, including lithium, at 4 years treatment duration (Lambert et al., 2016). Four case-control studies reported increased thyroid function in those taking lithium, but meta-analysis of all studies measuring TSH showed no difference between lithium and placebo (McKnight et al., 2012).

Lithium may also affect parathyroid function and calcium levels, but again the evidence is limited, inconsistent and tends to be of poor quality. McKnight et al. identified 4 cohort studies, 14 case-control studies and 6 cross-sectional studies. Meta-analysis of case-control studies suggested calcium and parathyroid hormone were increased by 10% in those taking lithium, however results were highly heterogeneous (McKnight et al., 2012). The source of this heterogeneity was not investigated statistically, but studies included mixed groups in terms of diagnosis and control group health status. These studies were also relatively small (maximum 142 cases). Of the cohort studies, sample size was small (maximum 53 people) and follow-up short (maximum 24 months). To date, there have been two additional cohort studies. One which found no association between lithium and adjusted calcium
concentration (Shine et al., 2015), the limitations of which are described in the renal effects section. The other found new hypothyroidism occurred at a rate of 12.9% in those taking lithium, however the cohort was small, had no comparison group and did not account for potential confounders (Albert et al., 2015).

1.5.2.5 Other physical health effects of lithium treatment

Treatment with lithium may induce weight gain. From 14 RCTs, clinically significant weight gain (>7%) was more common in those receiving lithium compared to placebo (OR 1.89, 95% CI 1.27 to 2.82) but lower than olanzapine (OR 0.32, 95% CI 0.21 to 0.49) (McKnight et al., 2012). Head-to-head comparisons of weight gain with mood stabiliser medications are rare and of short duration (one trial of lithium vs. valproate (Bowden et al., 2010), two trials of lithium vs. olanzapine (Niufan et al., 2008, Tohen et al., 2005), one trial of lithium vs. quetiapine (Sachs et al., 2004)). Additionally, none of these trials were powered (or set up) to test whether there were differences in weight gain. Other adverse effects that may be associated with weight gain, such as CVD, hypertension, and T2DM have not been widely investigated in lithium treated patients, and RCTs will not last long enough for these adverse outcomes to develop. Small, poor quality cohort studies have suggested that patients taking lithium have lower CVD mortality than other patients with BPD, and similar CVD mortality to the general population (Ahrens et al., 1995, Bocchetta et al., 2007). One cross-sectional study of 40 patients taking lithium monotherapy found no relationship between duration of treatment and hypertension risk (Klumpers et al., 2004). I am aware of only one RCT examining metabolic abnormalities, which found no association with lithium (McIntyre et al., 2011). A cross-sectional study found that patients taking lithium had lower metabolic syndrome prevalence than those taking SGAs (Yumru et al., 2007). A meta-analysis of case-control and cohort studies of patients with BPD found that prevalence of metabolic syndrome was higher than the general population and highest in those exposed
to antipsychotic medication, however it did not address the issue of lithium (Vancampfort et al., 2013a). Table 1.ii summarises these adverse effects.

Table 1.ii Summary of recognised adverse effects from long-term use

<table>
<thead>
<tr>
<th>Adverse effect</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Renal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyponatremia</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal failure</td>
<td>+</td>
<td>+?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Renal replacement therapy</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hepatic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased transaminase</td>
<td>+</td>
<td>++</td>
<td>+?</td>
<td>+?</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td><strong>Endocrine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>++</td>
<td>+</td>
<td></td>
<td>+?</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
<td>+?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypercalcemia</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperprolactinemia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Metabolic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight gain</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Hypertension</td>
<td>?</td>
<td>?</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>T2DM</td>
<td>+?</td>
<td>+?</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>CVD</td>
<td>+?</td>
<td>+?</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>Neurological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tremor</td>
<td>++</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Gastrointestinal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancreatitis</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Haematological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leukopenia</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sexual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased libido/function</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Infertility</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Teratogenic</td>
<td>+</td>
<td>++</td>
<td>+?</td>
<td>+?</td>
</tr>
</tbody>
</table>

* effect present, ++ strong effect, - protective effect, ? unclear effect. Adapted from (Calkin et al., 2013, Correll et al., 2015, Goodwin et al., 2016, Lambert et al., 2016, McKnight et al., 2012, McLaren and Marangell, 2004, Murru et al., 2015, Shine et al., 2015, Terrana et al., 2015).
Other adverse effects are likely to be important to patients and modify treatment adherence, but are unlikely to be well recorded by clinicians. For example, studies suggest that around one third of those taking lithium experience sexual dysfunction (Elnazer et al., 2015, Grover et al., 2014), however it is unclear how rates compare to those in individuals with untreated bipolar depression or treated with antidepressants. The teratogenic risk associated with lithium is low in absolute terms, but is elevated relative to placebo (Nguyen et al., 2009).

1.5.2.6 Suicide risk in patients taking lithium

Because RCTs often exclude those with a history of suicidal behaviour, drug effects on self-harm and suicide have been difficult to quantify due to low event rates (Perlis, 2011). The findings of a meta-analysis suggested that suicide was less likely in individuals prescribed lithium than placebo (4 trials, including no suicides in those treated with lithium and 6 in those treated with placebo – OR 0.13; 95% CI 0.03 to 0.66) (Cipriani et al., 2013a). However lithium did not reduce suicide compared to active comparator groups, and there was no difference in self-harm rates (Cipriani et al., 2013a). The results of observational studies have suggested that lithium use may reduce fatal and non-fatal self-harm compared with maintenance treatment alternatives, most commonly anticonvulsant medication (Baldessarini et al., 2006b, Goodwin et al., 2003, Schou, 1998, Smith et al., 2009, Søndergård et al., 2008). Baldessarini and colleagues used data from a number of sources to demonstrate that individuals with BPD not treated with lithium had 10 times the risk of suicide and almost 4 times the risk of suicide attempts, than those treated with lithium (Baldessarini et al., 2006a). However, these studies are similarly limited by low event rates, confounding by indication, and have often been designed to investigate the effects of medication adherence on suicide, rather than comparing individual drugs (Smith et al., 2009, Søndergård et al., 2008). In addition, the findings have not always been consistent,
with a number of studies suggesting there is no difference between mood stabilisers (Ahearn et al., 2013, Bowden et al., 2000, Marangell et al., 2008). A recent large Swedish cohort study (unpublished PhD thesis (Song, 2017)) found a reduced rate of suicide-related events (self-harm and suicide) in people prescribed lithium, compared to those prescribed valproate. The study also found a reduced rate during periods on lithium compared to off-lithium, such that (if the association were causal) 12% (95% CI 4-20) of suicide-related events could be prevented if all patients had taken lithium for the entire follow-up period.

1.5.3 Valproate

1.5.3.1 Effectiveness as a maintenance mood stabiliser

Valproate has become increasingly used over the last two decades in the UK as prophylaxis against further affective episodes (Hayes et al., 2011), despite a lack of robust evidence for its effectiveness (National Institute for Health and Care Excellence, 2014). Although RCT meta-analysis suggests valproate is favoured over placebo for prevention of a composite outcome of “any new mood episode”, this is not the case when individual manic or depressive relapses are the outcome of interest (Miura et al., 2014). This is potentially due to low powered studies examining these specific outcomes. It is also not superior to placebo in terms of tolerability or acceptability measures (Miura et al., 2014). There is more positive evidence from naturalistic studies that compare hospital admission rates on and off treatment over a number of years. These suggest valproate is associated with more hospitalisations than lithium, but fewer than other maintenance mood stabiliser options (Goodwin et al., 2016). Combination treatment with lithium plus valproate has been shown to be better than lithium monotherapy (Geddes et al., 2010), however this ignores the potential for additive untoward effects. The risk of teratogenic effects has meant that valproate has been contraindicated in women of child-bearing age since the mid-2000s.
(Vajda et al., 2004) but this is not reflected in the prescribing trends (Figure 1.iv) (Hayes et al., 2011).

1.5.3.2 Adverse effects from long-term use of valproate

A number of adverse physical health events have been reported during treatment with valproate. Studies of these effects are rarer than those investigating the potential adverse outcomes for people taking lithium. Whilst, to some extent, the adverse effects of valproate have been compared with other antiepileptic medication in individuals with epilepsy, comparisons with other mood stabilisers in BPD are lacking (Greenwood, 2000). The most commonly reported adverse effects include gastrointestinal disturbance, tremor, weight gain and transaminase abnormalities (see Table 1.ii).

1.5.3.3 Weight gain during maintenance treatment with valproate

Weight gain appears to be a particular issue with valproate relative to other anticonvulsants (Cramer et al., 2010). RCTs suggest incident weight gain during valproate treatment may be as high as 59% – 12 times higher than in those treated with carbamazepine (Richens et al., 1994). A systematic review of cohort studies suggests up to 71% of adults with epilepsy treated with valproate will experience weight gain, with 47% gaining greater than 10% of baseline bodyweight (Jallon and Picard, 2001). Among trials in BPD populations, weight gain appears to be lower in people taking valproate than those taking antipsychotics, but higher than in lithium-treated patients. For example, a 12 week trial of manic patients found 10% of valproate-treated patients gained weight compared to 25% of olanzapine-treated patients (Zajecka et al., 2002). However, in an extension of this RCT to 47 weeks it became apparent that this was an early effect of olanzapine and after week 19 there was no difference between groups in terms of weight gain from baseline (Tohen et al., 2003). In a large maintenance treatment RCT comparing valproate, lithium
and placebo weight gain occurred in 21%, 13% and 7% respectively (P=0.004 for the
difference between valproate and placebo)(Bowden et al., 2000). The physiological
mechanism behind valproate associated weight gain remains unclear, but is potentially
related to hyperinsulinemia (Rakitin et al., 2015).

1.5.3.4 Hepatic Effects of valproate
Asymptomatic elevation of transaminases occurs in approximately 40% of individuals
prescribed valproate (Murru et al., 2015). Severe liver toxicity is rare, but fatal hepatic
failure occurs in 1 in 10,000 exposures, and appears to be non-dose-related (Perucca and
Gilliam, 2012).

1.5.3.5 Other effects of long-term valproate use
Given the association between valproate and weight gain, we might expect increases in
metabolic syndrome, hypertension, CVD and T2DM. Comparative studies of these longer-
term outcomes in BPD are rare or non-existent. I could identify three cross-sectional
studies of metabolic syndrome and hypertension; two found no association between
metabolic syndrome or hypertension and valproate (Correll et al., 2008, Elmslie et al.,
2009), one found no difference in risk between valproate and lithium (Yumru et al., 2007).
However, these studies were of small numbers of patients, were not longitudinal in nature
and took place in potentially non-representative samples, without appropriate comparison
groups. No studies looked specifically at CVD or T2DM. Other notable adverse effects
include encephalopathy symptoms, platelet disorders, pancreatitis and teratogenicity
(including a 1% to 3% risk of neural tube defects) (Perucca, 2002, Perucca and Gilliam,
2012). Unlike lithium, there is no evidence that valproate is associated with renal or
thyroid abnormalities.
1.5.3.6 Suicide risk in patients taking valproate

In 2008, a warning was issued from the US Food and Drug Administration (FDA) (US Food and Drug Administration, 2009) that anticonvulsant medications carry an increased risk of suicidal self-harm. The FDA conducted a meta-analysis of 199 RCTs (27,863 patients) and found four suicides in those treated with anticonvulsants vs. none treated with placebo. Interestingly, the overall risk of suicidal ideation or behaviour was 3.5 times higher than placebo in epilepsy trials and 1.5 times placebo in psychiatric trials. Of these RCTs, 14 were of valproate and found an OR for suicidal behaviour or ideation of 0.72 (95% CI 0.29 to 1.84). Since then a number of studies have investigated this issue in BPD. A meta-analysis of RCTs of valproate (Redden et al., 2011) and several observational studies that included only patients with BPD (Arana et al., 2010, Gibbons et al., 2009, Leon et al., 2014, Reid, 2011) did not support this concern. A recent Swedish cohort study suggests that self-harm and suicide occur at similar rates on and off valproate and more frequently than in lithium-treated patients (Song, 2017).

1.5.4 Olanzapine and Quetiapine

1.5.4.1 Effectiveness as maintenance mood stabilisers

The antipsychotic olanzapine has been shown to be one of the most potent treatments for acute mania, with a standardised mean difference compared to placebo of -0.43 (95% CI -0.54 to -0.32) (Cipriani et al., 2011). In many clinical situations, it would seem reasonable to continue this after remission from the acute episode (National Institute for Health and Care Excellence, 2006). However, there are few long-term trials of olanzapine as maintenance treatment, most use enrichment designs (i.e., continuing treatment in those that have shown an initial response), and none have the same degree of independent replication of efficacy as lithium (Miura et al., 2014, Severus et al., 2014). However this lack of evidence
has not prevented its inclusion in guidelines as a first-line treatment (National Institute for Health and Care Excellence, 2006) and a rapid increase in use (Hayes et al., 2011).

Quetiapine is the only monotherapy recommended as first-line treatment for acute bipolar depression (National Institute for Health and Care Excellence, 2014). Evidence for this recommendation comes from a meta-analysis of 11 trials which found a number needed to treat of 6 (95% CI 5 to 8) compared to placebo (Chiesa et al., 2012). Again, it would seem to be a rational treatment approach to continue this following mood stabilisation. This rationale is supported by one RCT with an active comparator group (lithium versus quetiapine (Weisler et al., 2011)). This trial suggested that individuals responding to quetiapine had an increased time to reoccurrence compared to those switched to lithium or placebo. This study design strongly favours quetiapine and was considered very low quality evidence by NICE (National Institute for Health and Care Excellence, 2014). Beyond this, RCTs of quetiapine as a longer-term maintenance treatment have not been carried out.

1.5.4.2 Adverse effects of olanzapine and quetiapine

Antipsychotic adverse effects have mostly been examined in the context of treatment for schizophrenia. Weight gain is a common adverse effect of antipsychotic medication and is associated with several diseases including CVD, hypertension, T2DM, respiratory problems and 11 different cancers (Kyrgiou et al., 2017, Must et al., 1999).

1.5.4.3 Weight gain during treatment with olanzapine or quetiapine

As can be seen, each drug in Figure 1.4 is associated with more than twice as many patients experiencing >7% weight gain from baseline than individuals taking placebo, and weight gain with olanzapine and quetiapine appears to be particularly problematic. A recent meta-analysis of weight gain in RCTs found, within a median duration of 3 months, olanzapine
and quetiapine were associated with a 2.4kg and 1.1kg weight gain respectively (Domecq et al., 2015). As with some of the other outcomes studied in this thesis, only short-term weight gain has been well studied, so the trajectory of weight gain in individuals taking these drugs long-term is speculative. In addition, these are not head-to-head comparisons of olanzapine and quetiapine, so the relative weight gain on each drug is not known. Yearlong RCTs in BPD patients, with head-to-head comparisons of olanzapine vs. lithium (mean weight gain 1.8kg vs. -1.4kg; P<0.001) and olanzapine vs. valproate (mean weight gain 2.8kg vs. 1.2kg; P<0.001) suggest olanzapine may be particularly associated with weight gain (Tohen et al., 2005, Tohen et al., 2003). This is also reflected in the literature on schizophrenia treatment with SGAs. Risk of dyslipidaemia with SGAs appears to mirror their propensity to cause weight gain (De Hert et al., 2012, Haddad and Sharma, 2007).

Figure 1.v Percentage of patients with clinically significant weight gain (>7%) in short-term (3-8 weeks) placebo-controlled RCTs.

Adapted from (Haddad, 2005).
1.5.4.4 Metabolic and cardiac effects of olanzapine and quetiapine

As stated, weight gain and dyslipidaemia are risk factors for adverse metabolic and cardiac outcomes. Studies of these longer-term outcomes are limited in number. Meta-analyses have found that SGAs are more strongly associated with T2DM than first generation antipsychotics (FGAs) (Smith et al., 2008). Observational studies have found that both olanzapine and quetiapine increase T2DM rates, relative to haloperidol (Lambert et al., 2006) and untreated patients (Buse et al., 2003). However, these studies were in cohorts with schizophrenia and were not able to show differences between olanzapine and quetiapine because of the small numbers of people developing T2DM. One previous Danish population based cohort study has examined rates of CVD in patients prescribed olanzapine or quetiapine for all indications (Pasternak et al., 2014). This study found no difference between quetiapine and olanzapine in CVD rates, after adjustment for a number of sociodemographic and health variables (HR 0.86; 95% CI 0.50–1.48).

1.5.4.5 Other effects of olanzapine and quetiapine treatment

Mild and transiently abnormal liver function tests (LFTs) are common in patients receiving antipsychotics (median 32% abnormal (Marwick et al., 2012)), but clinically significant elevations are rare (Correll et al., 2015). There is currently no clear evidence that either olanzapine or quetiapine convey a particular risk (see Table 1). T2DM is associated with renal disease, and therefore theoretically olanzapine and quetiapine have the potential to increase renal failure via this route. However, I am aware of no studies that specifically examine CKD incidence.

1.5.4.6 Suicide risk in patients taking olanzapine and quetiapine

There are sparse data on the association between antipsychotic medication use and self-harm or suicide in BPD. Small retrospective cohorts have shown no difference in self-harm
in patients taking olanzapine or quetiapine (Koek et al., 2012), but have demonstrated higher rates of suicide attempts in those prescribed SGAs (such as olanzapine or quetiapine) compared with lithium or valproate (Ahearn et al., 2013, Yerevanian et al., 2007). In contrast, in schizophrenia, observational studies have shown that suicide risk is lower in those taking SGAs compared to individuals taking FGAs (Altamura et al., 2003). The findings of the study by Altamura et al. are limited by small sample size, potential bias and confounding. In RCTs, individuals prescribed olanzapine have been found to have lower suicide attempt rates than those taking risperidone (Tran et al., 1997) and haloperidol (Beasley et al., 1998, Glazer, 1998). However, these studies all have limited follow-up and were not powered to examine self-harm.

1.5.5 Head-to-head comparisons of lithium, valproate, olanzapine and quetiapine as maintenance mood stabilisers in bipolar disorder

As discussed, head-to-head comparisons of lithium, valproate, olanzapine and quetiapine in terms of effectiveness, tolerability, self-harm, accidental injury, suicide and adverse physical health effects when used long-term for mood stabilisation in BPD are rare. Comparisons of this nature, accounting as well as possible for potential confounders, are vital to inform clinical decision-making.

From the potential adverse effects of these medicines (Table 1.iii), I chose to examine those which appear to be strongly related to a particular drug, commonly occurring (and well recorded in EHR), and/or having greatest impact on patient wellbeing and function. These can be divided into four types of adverse events: renal (chronic kidney disease), hepatic (hepatotoxicity), endocrine (thyroid disease, and hypocalcaemia) and metabolic (weight gain, hypertension, T2DM and CVD).
1.6 Limitations of randomised controlled trials in bipolar disorder

Although RCTs are the gold standard for demonstrating the efficacy of medications, their applicability to individuals with BPD may be compromised by the complex, labile symptomatic presentations of the illness, patients’ tendency to deny illness or reject treatment, and diagnostic heterogeneity in routine practice. These concerns have been raised in other areas of medicine regarding RCTs where the disorder under investigation is chronic and relapsing-remitting and when the exclusion criteria of RCTs can often mean that external validity is low (i.e., people included in trials do not represent a real life sample of individuals receiving a bipolar disorder diagnosis) (Lancet Editors, 1992). Therefore applying existing RCTs results to managing a lifelong illness of unpredictable course is not straightforward (Black, 1996). By definition, as soon as a clinical treatment lasts longer than the RCT that informed it, the treatment is no longer “evidence-based”. Necessary trials are also costly and difficult to run for sufficient periods in relation to the time-course of BPD (Hayes and Osborn, 2011).

1.6.1 Problems with trial designs in bipolar disorder

Many RCTs of maintenance treatment have been criticised because of enriched designs, which select patients who have responded to the treatment used in the acute phase. The network meta-analysis by Miura et al. found this to be the case in 19 of 33 included trials (Miura et al., 2014). It is recognised that this may be particularly problematic in the case of lithium, where withdrawal of the drug can prompt relapse (Moncrieff, 1997). However this is also true of valproate and olanzapine (Goodwin, 2009). The enrichment design can answer questions about the continued benefits of the investigational medicine, but is not a fair test of the comparator agent.
Of all existing trials, the study design for the ‘BALANCE’ RCT (Geddes et al., 2010) aimed to maximise the generalisability of the findings to a clinical population, by allowing clinicians to prescribe as they wished (beyond allocation to either lithium, valproate or lithium plus valproate). However, limitations inevitably remained in terms of including patients who had shown a differential previous response to either lithium or valproate, diagnostic heterogeneity within the sample population, and frequency of comorbidity compared with the general population.

1.6.2 Recruitment problems in pharmacological trials for bipolar disorder

Recruitment to trials of maintenance BPD medications has proved difficult and costly. Again, using the ‘BALANCE’ RCT (Geddes et al., 2010) as an example: it took seven years to recruit its 345 randomised subjects, compared to the original target of 3,000. This also forced the investigators to change the primary outcome to provide sufficient statistical power (Geddes and Goodwin, 2001, Geddes et al., 2010). Other trials have had similar problems (Charlson and Horwitz, 1984, Jones et al., 2006, March et al., 2014, Nolen et al., 2007). As such, further large comparative efficacy trials of “established” off-patent medications are unlikely to be funded. Therefore, alternative study designs are need to advance the evidence-base, such as quasi-experimental designs in routine cohorts.

1.6.3 Length of follow-up in pharmacological trials for bipolar disorder maintenance treatment

In the recent meta-analyses of maintenance treatment trials (including 33 studies) the longest follow-up was found to be 3.3 years, with the majority having only one year of follow-up and many of these risking bias because of high attrition rates (Miura et al., 2014). It is difficult to extrapolate conclusions of these RCTs to the longer-term maintenance treatment. Notably, given the recognised kindling effect of recurrent affective episodes (i.e., each new mood episode increases the risk of further episodes) it is unlikely that the
risk-benefit profile of a medication will be constant over time. In addition, none of the recognised long-term side effects of these medications will have the opportunity to develop. Much longer follow-up is needed to determine whether there is a timepoint when the benefits of the treatment cease or when the risks become unacceptable.

1.6.4 Choosing meaningful outcomes for maintenance treatment trials

BPD trials have tended to define their primary outcome for efficacy/effectiveness in terms of time to relapse or recurrence, measured by a mood disorder scale, meeting diagnostic criteria for a new mood disorder episode, hospitalisation, need for a change in medication or add-on of medication or ECT, or the investigator judging that medication was stopped due to a new mood episode. Often, these studies will include a measure of tolerability represented by time to stopping medication: i.e., staying on the drug means that the patient has no untoward side effects. Sometimes trials have used a combined efficacy and tolerability measure as their primary outcome which becomes more difficult to interpret, but may more accurately reflect real life use of the drugs (Hayes and Osborn, 2011).

In the real world, treatment choice has multiple aims, including reducing hospital admissions, but also improving mood stability, reducing inter-episode residual symptoms, improving functioning and improving quality of life. Trials have struggled to assess these outcomes.

1.6.5 Lack of trials with active comparator groups in bipolar disorder maintenance treatment

Both placebo and active comparator trials are important for the full assessment of efficacy, effectiveness, tolerability and safety of medications. Active comparator studies have been identified as particularly important when 1) the medication may be associated with safety concerns for mortality or morbidity, and 2) a medication with inferior efficacy may
conceivably lead to significant, long-term or irreversible harm (Committee for Medicinal Products for Human Use, 2011). Both of these issues are true in the case of lithium, valproate, olanzapine and quetiapine.

1.6.6 Overcoming the limitations of trials of maintenance treatment

In summary, trials are short in follow-up, limited in number, have few participants who are unrepresentative of the broader population with bipolar disorder diagnoses, and often biased because of their design. The following sections explain how this thesis aims to address the issues outlined above using observational data from primary care EHRs.

1.7 Addressing the gaps in our understanding about long-term outcomes in bipolar disorder with observational data

1.7.1 Data Source: Primary care electronic health records

The data source for each of the studies is The Health Improvement Network (THIN) an anonymised EHR primary care database in the United Kingdom. It is described in detail in Section 2.3. At the time the cohorts were extracted for this thesis, THIN contained records of over 11 million people, with a median follow-up time of approximately six years (http://www.epic-uk.org/). Patients included in the database have been shown to be broadly representative of the UK population, and General Practitioners (GPs) contributing data are representative in terms of consultation and prescribing statistics (Blak et al., 2006, Blak et al., 2011). Approximately 98% of the population is registered with a GP (Lis and Mann, 1995). The reliability of THIN for research purposes has been validated against experimental and other observational evidence (Langley et al., 2010, Lewis et al., 2007). The validity of diagnoses of SMI (schizophrenia, BPD and other psychotic illnesses) in primary care has been shown (Nazareth et al., 1993) and the incidence rate of BPD in the database is similar to other European estimates (Hardoon et al., 2013). Although diagnosis
of BPD is mainly made in secondary psychiatric services, many patients will primarily be
cared for by their GP (National Institute for Health and Care Excellence, 2014). This makes
THIN an ideal data source for conducting representative, valid and generalisable research
into the adverse outcomes and effects of drug treatment in BPD.

1.7.2 Mortality and morbidity in bipolar disorder
From the systematic review (Chapter 3), particular areas were identified for further
investigation and are addressed in Study 1 (Chapter 4) of the thesis. Study 1 reports rates
of all-cause mortality and mortality from cardiovascular events, and suicide in BPD,
schizophrenia and a frequency matched general population comparator group. It also
examines rates of CVD diagnoses, and self-harm in the three groups. It is hypothesised that
all-cause mortality, CVD mortality, suicide, CVD diagnosis and self-harm rates are elevated
relative to the general population, and in line with rates for schizophrenia.

1.7.3 Effectiveness and tolerability of maintenance medication
In the past, commentators have criticised observational studies of medications, suggesting
they would overestimate treatment effects and that they therefore provide little value in
assessing therapies; however, comparative studies with RCTs, across various branches of
medicine, have now challenged this claim (Black, 1996). Study 2 (Chapter 6) is based on a
similar complementary approach; attempting to reconfirm findings from RCTs, with larger
patient numbers, over longer follow-up periods, and with patients who are more
representative of a wider cross-section of those with a BPD diagnosis. This study compares
the time to monotherapy treatment failure (defined as stopping the drug, swapping to a
new drug or requiring add-on of a mood stabiliser, antipsychotic, antidepressant or
benzodiazepine) in individuals prescribed lithium, valproate, olanzapine or quetiapine. This
outcome was chosen as a proxy for combined effectiveness and tolerability, because it has
been used in a number of drug trials in BPD (Miura et al., 2014). This study uses a quasi-experimental design using a propensity score (PS) approach to balance baseline characteristics of patients prescribed different drugs. The PS approach is further described in Sections 2.9.2 and 2.11. It is hypothesised that all drugs will show similar rates of monotherapy treatment failure.

1.7.4 Adverse events associated with maintenance medication in bipolar disorder

Follow-up time in RCTs is usually short, relative to the time it may take to develop many drug side effects during long-term maintenance treatment. RCTs are also designed with sufficient sample size and statistical power to detect efficacy, rather than the power to determine the safety of the study medications (Wahab et al., 2013). Current methods of reporting chronic adverse events in routine surveillance, either by health professionals or patients are not felt to be adequate (Blenkinsopp et al., 2007). Use of routine EHRs is likely to be a more efficient way of identifying and quantifying the risk of adverse events, as they are likely to be complete for severe events, and do not require extra work of reporting events by the clinician (Honigman et al., 2001, von Euler et al., 2006).

Lithium, valproate, olanzapine and quetiapine have been widely used across the world and so there should be many years of patient data available for side effect profiling. Despite this the risks of chronic adverse effects have only recently begun to be quantified and appraised (Close et al., 2014, McKnight et al., 2012, Shine et al., 2015) and there are no existing studies making head-to-head comparisons of adverse events associated with these drugs. Study 3 (Chapter 7) is a comparison of the major adverse events associated with lithium, valproate, olanzapine and quetiapine using a PS approach. These can be divided into four types of adverse events: renal (chronic kidney disease), hepatic (hepatotoxicity), endocrine (thyroid disease, and hypocalcaemia) and metabolic (weight gain, hypertension, T2DM and
CVD). It is hypothesised that lithium will be associated with more renal and endocrine adverse events, valproate with more hepatic adverse events, and olanzapine and quetiapine with more metabolic adverse events. However, given that there is some evidence that each of these drugs can potentially have hepatic, renal and metabolic effects, not all differences may be clinically significant.

1.7.5 Self-harm rates with different maintenance medications

Preventing self-harm and suicide are key aims in the effective management of BPD. Lithium is the only drug which has been consistently found to reduce self-harming behaviour in both observational (Baldessarini and Tondo, 2003, Baldessarini et al., 2003, Goodwin et al., 2003) and trial data (Cipriani et al., 2013a). However, existing cohort studies have tended to lack active comparator groups and results from RCTs are secondary analyses with small numbers of events. **Study 4 (Chapter 8)** is therefore a comparison of rates of self-harming behaviour in patients prescribed lithium, valproate, olanzapine or quetiapine using a PS approach. The study also examines accidental injury rates to test hypotheses about potential mechanisms of action. It is hypothesised that patients prescribed lithium will have the lowest rates of self-harm.

1.8 Objectives of the studies included in this thesis

1.8.1 Long-term outcomes in bipolar disorder

i) To summarise previous observational studies of long-term prognosis in individuals with a diagnosis of BPD by examining all-cause and cause specific mortality via systematic review and meta-analysis (**Chapter 3**)

ii) To calculate recent time trends in all-cause mortality in the UK in individuals with BPD compared to individuals with schizophrenia and the general population (**Study 1: Chapter 4**)

54
iii) To determine relative rates of i) CVD deaths, ii) suicide, iii) CVD diagnoses, iv) self-harm in individuals diagnosed with BPD or schizophrenia compared to the general population, while accounting for sociodemographic factors (Study 1: Chapter 4)

1.8.2 Effectiveness and adverse effects of maintenance treatments

iv) To determine relative efficacy of the four most commonly used maintenance mood stabiliser medications (lithium, valproate, olanzapine and quetiapine) via a network meta-analysis of all head-to-head and placebo controlled RCTs (Chapter 5)

v) To assess comparative effectiveness and tolerability of the four most common mood stabilisers by calculating rates of time to cessation of treatment, or add-on of another psychotropic medication in individuals prescribed lithium, valproate, olanzapine and quetiapine, while accounting for propensity to be prescribed one of these mood stabilisers (Study 2: Chapter 6)

vi) To calculate rates of adverse events on these four mood stabilisers, specifically chronic renal, hepatic, endocrine and metabolic effects, accounting for propensity to be prescribed one of these mood stabilisers (Study 3: Chapter 7)

vii) To determine rates of self-harm, unintentional injury and suicide on these four mood stabilisers, while accounting for propensity to be prescribed one of these mood stabilisers (Study 4: Chapter 8)
Chapter 2  Methods – justification and overview

2.1  Summary

This chapter provides a rationale for decisions made about how to compete the systematic reviews and meta-analyses and the planning of the five completed cohort studies. THIN was identified as a suitable data source for the studies in this thesis, for the reasons outlined. The steps taken in the developing the hypotheses to be tested, the study design used and the analysis plans for each of the studies (with the aim of minimising confounding and bias and maximising the possibility of drawing conclusions related to causality) are described. The limitations of the chosen approach are considered in the discussion section of each chapter and more generally in Chapter 9.
2.2 Systematic review and meta-analysis

2.2.1 Summarising studies of mortality in people with bipolar disorder

There are a number of methods of reporting premature mortality; including death rates, life-expectancy, and measures of years of life lost. However, the most commonly used comparative measure is the SMR. The SMR is an indirect method of standardisation calculated by the ratio of observed deaths in the study group to expected deaths in the general population. Studies reporting SMR were the focus of the review to be reported in Chapter 3 because the large number of available SMR estimates allowed a meta-analysis and updated the highly cited work of Harris and Barraclough (Harris and Barraclough, 1998).

2.2.2 Network meta-analysis for comparing the efficacy and tolerability of maintenance mood stabilisers

A key limitation of traditional meta-analysis techniques to examine treatment effects is that they compare only two interventions at a time (Cipriani et al., 2013b). This is a particular problem if several treatment options need to be compared, as the result will be a number of pairwise meta-analysis comparisons. This is an issue in the assessment of maintenance mood stabilisers for BPD, where multiple drugs may be effective. Network meta-analysis (NMA) is a method to summarises trial results of all treatments (Caldwell et al., 2005). NMA synthesises data from a network of trials about more than two competing interventions. The integration of direct evidence (from studies directly comparing interventions) with indirect evidence (information about two treatments derived via a common comparator) increases the precision in the estimates and can produce a relative ranking of all treatments for the studied outcome (Bucher et al., 1997, Salanti et al., 2011). As such, the relative effectiveness of two treatments can be estimated, even if no trials compare them directly.
However, the validity of the conclusions drawn relies on a number of assumptions about heterogeneity, transitivity and consistency (Chaimani et al., 2013).

Chapter 5 of this thesis is a NMA of trials comparing any combination of lithium, valproate, olanzapine, quetiapine and placebo. These treatments are better connected in the NMA than all the drugs included in Miura et al. (Miura et al., 2014) and comparisons are less likely to violate the transitivity assumption (this is discussed in the introduction to Chapter 5). Comparisons with findings from the NMA of Miura et al. are made in the discussion section of Chapter 5.

2.3 Data Source for cohort studies – primary care electronic health records

2.3.1 Comparison of The Health Improvement Network with other primary care electronic health records

Three large primary care EHR systems exist in the UK: Clinical Practice Research Datalink (CPRD), QResearch, and THIN (Shephard et al., 2011). Both CPRD and THIN are descendants of the Value Added Information System (VAMP) which began extracting pseudonymised medical records from the Vision software platform (a frontend user EHR) in 1987 (Williams et al., 2012). Currently there is an overlap of data held in CPRD and THIN, with approximately 60% of GP practices contributing to both, because of their use of Vision (Carbonari et al., 2015). QResearch extracts from another commonly used EHR software; Egton Medical Information System (EMIS) (Coupland et al., 2016). A number of studies carried out across these three databases have suggested high levels of similarity (Hippisley-Cox et al., 2014, Reeves et al., 2014). I chose THIN as the data source for the studies because of the extensive expertise that exists within UCL (in both the Division of Psychiatry and the Department of Primary Care and Population Health) and because UCL holds a licence to access all THIN data.
2.3.2 Coverage of The Health Improvement Network database

The studies in this thesis use data up until the end of 2014, at which time, THIN covered 641 GP practices, representing 6% of the UK population (13,816,680 total patients) (Personal correspondence with IMS Health, 2016). Patients in the database have been shown to be broadly representative of the UK population, and GPs contributing data have been shown to be representative in terms of consultation and prescribing statistics (Blak et al., 2006, Blak et al., 2011). Approximately 98% of the UK population is registered with a GP (Lis and Mann, 1995).

2.3.3 Data recording in The Health Improvement Network

Information in the THIN database is generated through routine record keeping of GPs for the purposes of clinical management of their patients. GPs use Read codes, a hierarchical coding system, to record information in THIN (Chisholm, 1990). These codes include diagnoses (which map onto ICD-10 codes), symptoms, examination findings, health indicators (for example smoking, alcohol intake, body mass index (BMI)), referrals, test results and information from hospital specialists, creating a longitudinal record for each patient (Davé and Petersen, 2009). GPs are responsible for issuing drug prescriptions if treatment is ongoing, and this information is available and essentially complete in THIN (Health and Social Care Information Centre, 2012). Missing prescribing data may occur when prescriptions are handwritten or when a patient is in hospital for a prolonged period. Some drugs, such as clozapine and some depot antipsychotic medications are prescribed in secondary care but are not relevant to the studies in this thesis. Over the counter medications are also not included, but similarly are not important in the context of these studies.
THIN also contains a record of patient ethnicity. This was poorly recorded prior to 2006, but
GPs were incentivised, via the Quality and Outcomes Framework (QOF), to add it to records
from 2004. By 2010 over 90% of newly registered primary care patients were having
ethnicity recorded (Mathur et al., 2014). There is no patient level measure of
socioeconomic status (SES) or educational attainment recorded in THIN, but there is an
area level measure which is associated with individual SES; the Townsend deprivation
index. The Townsend score incorporates four variables: unemployment (as a percentage of
those aged 16 and over who are economically active), non-car ownership (as a percentage
of all households), non-home ownership (as a percentage of all households), and household
overcrowding (Townsend, 1987). In THIN a score is assigned to each lower super output
area (an area of approximately 650 households) and is expressed in quintiles of Townsend
score based on 2001 UK census data.

2.3.4 Ethical approval for the cohort studies

The scheme for THIN to provide anonymous patient data to researchers was approved by
the National Health Service South-East Multicentre Research Ethics Committee in 2003, and
scientific approval for the studies in this thesis was obtained from Cegedim Strategic Data
Medical Research’s Scientific Review Committee in March 2015. A copy of the scientific
approval application is included in Appendix 1.

2.4 Code lists for defining exposures, outcomes and covariates

Code lists were developed for each variable used in the studies using the method described
by Davé and Petersen (Davé and Petersen, 2009). These were either Read code lists or drug
code lists, but the essential technique for code list preparation is the same. Researchers at
the Department of Primary Care and Population Health (UCL) and the Division of Psychiatry
(UCL), have previously developed a large number of code lists, and further lists are held at https://clinicalcodes.rss.mhs.man.ac.uk/ (Springate et al., 2014). However, on reviewing the Read code lists it became apparent that each study would require unique lists, often due to the sensitivity and specificity of the lists. Code lists for exposure and outcome variables would often need to be more sensitive than those used for confounder adjustment. Where possible, existing code lists were used as the starting point for study specific code list development. Code lists were developed via an iterative process of refinements involving all supervisors. Full code lists for BPD and schizophrenia are included in Appendix 2 as examples. Code lists for outcomes and covariates are available on request. These have not been included as code lists for outcomes alone were greater than 400 pages long.

2.5 Inclusion criteria for the cohort studies

2.5.1 Primary care practice inclusion criteria

Two measures of data quality have been developed for use with THIN: acceptable mortality reporting (AMR) (Maguire et al., 2009) and acceptable computer usage (ACU) (Horsfall et al., 2013). AMR is a measure comparing the standardised annual rates of all-cause mortality reporting for a GP practice with expected deaths using Office for National statistics (ONS) death data, accounting for the age and sex distribution of patients in the practice (Maguire et al., 2009). The ACU is the date at which a practice first entered an average of two or more therapy records, or one medical record and one additional data item into each patient record per year (Horsfall et al., 2013). For the studies in this thesis, practices are only included after they meet both these criteria. Combining ACU and AMR has been found to produce incident time trends consistent with external data sources (Horsfall et al., 2013). Furthermore, GP practices had to contribute at least three years of data to THIN, have a list
size of more than 2000 patients and have a measure of area level deprivation for at least 80% of registered patients. These additional criteria have been used previously to ensure only high quality patient records are used for analysis.

2.5.2 Bipolar disorder and schizophrenia inclusion criteria

The code lists used as a starting point for BPD and schizophrenia case identification have been used previously (Hardoon et al., 2013, Osborn et al., 2011, Osborn et al., 2015). For BPD, 97 codes were identified. These codes denoted a range of episode severity and chronicity of disorder. Unfortunately, it was not possible to separate out subtypes of BPD using Read codes. However, patients with BPD diagnoses who attended GPs and were coded as having depression and or depression symptoms were identified. For schizophrenia, 92 codes were identified. A conservative approach was taken to code list generation to increase specificity. For example; individuals with codes suggestive of schizoaffective illness were excluded, as this disorder is an intermediate phenotype, which shares diagnostic and endophenotypic features with both BPD and schizophrenia. Individuals with schizotypal diagnoses were only included if they eventually were coded as schizophrenia. Other non-organic psychotic illness diagnoses were also excluded, as their aetiology was uncertain.

Individuals were included if they were 16 years of age or over. Diagnosis start date was considered the first date at which any SMI or depression diagnosis code was entered into the patient record, in the case of BPD, and the first date at which any SMI code was entered, in the case of schizophrenia. If an individual received more than one SMI diagnosis in their longitudinal EHR, the most recent (latest) one was used. Most recent diagnosis was considered most accurate, as the clinician would have access to information about the patient’s full illness history. Previous research has shown that BPD and schizophrenia
diagnoses in THIN are relatively stable between first record of SMI and diagnosis ultimately assigned (Table 2.i) (Hardoon et al., 2013).

Table 2.i Diagnostic stability of bipolar disorder and schizophrenia

<table>
<thead>
<tr>
<th>First SMI record</th>
<th>BPD</th>
<th>Schizophrenia</th>
<th>Other SMI</th>
<th>SMI register</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPD</td>
<td>98.5</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>1.5</td>
<td>98.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other SMI</td>
<td>3.4</td>
<td>8.0</td>
<td>88.5</td>
<td>0</td>
</tr>
<tr>
<td>SMI register</td>
<td>5.3</td>
<td>3.4</td>
<td>5.6</td>
<td>85.8</td>
</tr>
</tbody>
</table>

Adapted from (Hardoon et al., 2013).

2.5.3 General population frequency matching for the study of mortality and morbidity

For each individual with BPD or schizophrenia, up to six people without BPD or schizophrenia (but potentially with other mental health problems) were selected at random from the same GP practice. There are diminishing gains in increasing the ratio of unexposed (general population) to exposed (BPD or schizophrenia) in the study, and a ratio beyond 3:1 or 4:1 is commonly seen as unnecessary (Strom, 2011). However, because of the size of THIN a ratio of 6:1 is feasible. Individuals from the general population were matched on sex, age (in five-year age bands) and index GP appointment attendance date (to ensure individuals in the comparator group were active in healthcare seeking in the same period as those with BPD and schizophrenia). This strategy for generating a comparison group partially removes potential confounding by age, sex, area level differences and cohort effects. By only excluding those with BPD or schizophrenia, it avoids rendering the comparison group an abnormally healthy sample, therefore biasing results.
2.6 Data extraction from The Health Improvement Network database

Data extraction for the studies in this thesis was carried out on two separate occasions. The effectiveness/tolerability study (Study 2: Chapter 6), the adverse effects study (Study 3: Chapter 7) and the suicide/self-harm study (Study 4: Chapter 8) were completed first, and used data from 1 January 1995 until 31 December 2013. The mortality/morbidity study (Study 1: Chapter 4) used a later THIN extraction of data from 1 January 2000 until 31 December 2014 (Figure 2.i). Data extraction was completed using Stata version 13 or 14 (StataCorp, 2013). I wrote or modified code to extract the required cohort and covariates of interest. All data cleaning and new variable generation was also completed using Stata (StataCorp, 2013).

2.7 Definition of exposures, outcomes and other covariates

Study specific exposures, outcomes and covariates are defined in each chapter, with code lists in Appendix 2 and available from the author on request.
Figure 2.1 Flow diagram of included patients for each study

All practices with acceptable AMR and ACU

- Bipolar disorder diagnosis with follow-up from 01.01.95-31.12.14
- Schizophrenia diagnosis with follow-up from 01.01.95-31.12.14
- No bipolar disorder or schizophrenia diagnosis with follow-up from 01.01.95-31.12.14

Exclusions
- Aged <16
- Final diagnosis not bipolar disorder or schizophrenia

Bipolar disorder or schizophrenia, with follow-up from 01.01.00-31.12.14

Matched by:
- Sex
- Age group
- GP practice index appointment date

Cohort used for mortality, morbidity Studies (Study 1; Chapter 4)

Bipolar disorder medication studies

Bipolar disorder diagnosis at any time, follow-up from 01.01.95-31.12.13

Treated with any mood stabiliser at any time during follow-up

Treated with lithium, valproate, olanzapine, quetiapine

Also prescribed mood stabiliser (other than lithium, valproate, olanzapine or quetiapine), antidepressant or benzodiazepine at start of follow-up

Cohort used for adverse effects studies (Study 3; Chapter 7 and Study 4; Chapter 8)

Cohort used for monotherapy prescribing study (Study 2; Chapter 6)
2.8 Benefits of using The Health Improvement Network for bipolar disorder outcomes and medication response research

2.8.1 Size, variable recording, reliability, validity and generalisability of available data

THIN has been used extensively to examine outcomes for people with SMI (Petersen et al., 2016), predict risk in SMI (Osborn et al., 2015) examine prescribing trends (Hayes et al., 2011, McCrea et al., 2015), monitor health trends (Hardoon et al., 2016) and perform clinical effectiveness studies (Blackburn, 2016). Previous recent work has identified a cohort of over 10,000 individuals with BPD (Osborn et al., 2015) and the cohort used in this thesis was larger than this: benefiting from extra years of data and additional GP practices joining THIN (at which point historical records are added to the database). This makes the cohort of BPD patients in THIN comparable in size to population based registries, such as those in Denmark (Medici et al., 2015) or Sweden (Ösby et al., 2016).

The incidence rate of BPD in THIN has been shown to be similar to other European cohorts (Hardoon et al., 2013) and validity of severe mental illness diagnoses held in primary care has been established (Nazareth et al., 1993). NICE guidance recommends that any patient with suspected BPD should be referred to a psychiatrist for diagnosis and treatment planning (National Institute for Health and Care Excellence, 2014). As such, individuals in this cohort are considered to fulfil ICD-10 criteria for BPD. A number of the health outcomes and drug related adverse events examined in this thesis have also been validated in THIN, including mortality, suicide, CKD, thyroid disease, T2DM, hypertension, and CVD (Blak et al., 2011, Hall, 2009, Khan et al., 2010). THIN has also shown itself to be suitable for pharmacoepidemiological research, with a number of studies showing comparability between UK EHR prescribing and national dispensing data (Langley et al., 2010, Lewis et al., 2007, Walley and Mantgani, 1997).
THIN offers an opportunity to examine a number of outcomes experienced by individuals with BPD. The coverage of THIN should mean that findings are generalisable to individuals with BPD living in the UK. It should be noted that the approximately 1.5% of the population missing from any study of primary care might have higher rates of SMI, including BPD, because prisoners, illegal immigrants and homeless people are less likely to be registered with a GP. However, the majority of this unregistered group is likely to be young healthy men (Harvey et al., 2012). Since 2004, the QOF has meant that GPs are incentivised to record particular information about chronic health problems, including SMI (Roland, 2004). There is evidence that this has improved the quality and accuracy of GP EHR (Doran et al., 2011). With regards to SMI, GPs are remunerated for keeping a register of individuals with SMI and carrying out an annual review (Osborn et al., 2011). Particular to individuals prescribed lithium, QOF indicators exist for having lithium levels checked six monthly and serum creatinine and thyroid stimulating hormone checked every 15 months. Beyond SMI, there are a number of measures of health (hypertension, T2DM, for example) that are used in this thesis and are well recorded due to QOF (Doran et al., 2011). Another particular strength of THIN for BPD research is the complete and accurate prescribing data available. Prescribing data are better quality and are available for longer, than prescribing data from the previously mentioned Nordic registers. The other advantage over these registers is the recording of blood tests and illness symptoms, which are vital for the studies proposed in this thesis.

2.9 Potential problems with the use of THIN for bipolar disorder outcomes and medication response research

2.9.1 Causal inference in observational data

The limitations of RCTs in the area of BPD have been discussed in Section 1.6. The limitations of observational data and approaches to managing these must also be
considered. One particular problem is the potential to make causal inference in non-experimental settings. Causal inference based on counterfactual models has become increasingly standard in medical and epidemiological studies (Höfler, 2005). In its simplest form the counterfactual model considers the potential outcomes of the same individual, at the same time point, assigned to both treatment and no treatment. Clearly only one of these two states can be observed, be it in an RCT or observational study. Therefore, it is not feasible to examine the effects of a given treatment on a single individual, but estimating the average effect across equivalent individuals who receive different treatments is possible.

2.9.2 Using propensity scores to achieve conditional exchangeability

If we consider a perfectly designed and executed RCT (one that is adequately powered, appropriately randomised, no loss to follow-up), randomisation ensures exchangeability; that is, participants are equivalent because treatment assignment is not associated with the counterfactual outcome. Under these circumstances, the causal effect is not dependent on which group receives the intervention and the differences observed between treatment and control group are the same as the average difference in potential outcomes. Some RCTs, rather than randomly assigning on a one-to-one basis will randomise accounting for a particular baseline characteristic. In this case, the RCT is conditioned or stratified on a baseline variable, and whilst exchangeability is not maintained between treatment and control groups, there will be exchangeability within a particular strata (this is called conditional exchangeability). This provides a starting point to consider the design of observational studies, which aim to fulfil criteria for conditional exchangeability. The use of a PS is an approach that has been suggested to achieve conditional exchangeability. PS approaches are further discussed in Section 2.11.1 and are used in Studies 2 (Chapter 6), 3 (Chapter 7) and 4 (Chapter 8).
2.9.3 Confounding by indication in observational studies

A traditional concern about the validity of findings from epidemiologic studies, and one that is linked to lack of exchangeability, is the possibility of bias from uncontrolled confounding. Case-control and cohort studies compare outcomes between groups with different exposures, and confounding arises when the groups under comparison differ in other ways than the exposure alone. These differences may include demographic factors, behaviours, clinical characteristics, medical conditions, or treatments. Some exposures are more liable to confounding than others. When the outcome is an unintended or unanticipated effect of the exposure, for example rare adverse effects, confounding is less likely to occur than when the outcome is an intended effect of the exposure (Miettinen, 1983). The potential problem of intended effects is likely to arise when the exposure of interest is a medication or a medical procedure, and it is often called confounding by indication (Walker, 1996). Confounding by indication may arise when a drug treatment serves as a marker for a clinical characteristic or medical condition that triggers the use of the treatment and that, at the same time, increases the risk of the outcome under study. Confounding by indication is not conceptually different from confounding by other factors, and the approaches to control for confounding by indication are the same: matching, stratification, restriction, and multivariable adjustment. However confounding by indication remains an often-intractable threat to validity in observational studies (Freemantle et al., 2013).

2.9.4 Other types of bias in studies using routine data

Structural biases, such as the bias due to left truncation or censoring, can occur when the inclusion criteria for a study are related to a variable of primary study interest, either directly or indirectly (Cain et al., 2011). This is unlikely to be an issue for Study 1, which examines time to death, cause specific death and morbidity. In this study, the comparison
group is matched on index GP appointment attendance date and age to avoid this problem. However, it is a potential problem in the maintenance mood stabiliser studies (Studies 2, 3 and 4). Individuals included in these studies may have previously been exposed to the mood stabiliser, but also (by definition) they would not be included if they had already experienced an outcome event of interest. For these studies, the possibility of working with an incident cohort was explored, but numbers of individuals taking each drug became too small to examine rare outcomes. Often information about previous use of mood stabilisers and BPD diagnosis before the start of follow-up was available and this was included in the statistical models used to minimise this problem. There is also a strong argument that although studies of prevalent exposures can involve left truncation, studies of incident exposures may involve right censoring and therefore may not be able to adequately assess the long-term effects of exposure (Vandenbroucke and Pearce, 2015). RCTs of these medications invariably face the same problem of prevalent exposure. In fact, a mix of prevalent and incident exposures might be the ideal situation as this allows exploration of the effects of different durations of exposure and enhances information on long exposure durations (Vandenbroucke and Pearce, 2015).

2.9.5 Lack of recording of important covariates

Because the EHR exists for the purposes of clinical management, rather than to run perfectly designed cohort studies, data on important confounders may not be recorded, and data at necessary time points may be missing.

2.9.5.1 Proxy variables

Proxy variables are known/measured variables that are associated with an unmeasured variable of interest. In the case of THIN, a proxy for body fat percentage (unrecorded) may be BMI (recorded), a proxy for individual level SES (unrecorded) may be Townsend score
(recorded) (Pickett and Pearl, 2001). Potential proxies for clinician prescribing choice and patient preference are discussed in Section 2.10.2.

2.9.5.2 Missing data

A number of approaches have been developed to manage missing data. Multiple imputation is a favoured approach (Schafer, 1999) and techniques have been specifically developed in THIN to manage missingness in longitudinal datasets (Welch et al., 2014). Before designing the studies presented in this thesis, a lot of thought was given to the analytical approach. Analysis using PS was favoured over other ways of managing confounding in the studies comparing mood stabiliser medication (for example instrumental variables, self-controlled case series). Currently there is no reliable strategy for incorporating multiple imputation and PS approaches (Mitra and Reiter, 2016) and thus other methods of managing missingness were considered. It was decided that the best approach was to consider the record as complete and perform sensitivity analyses as appropriate to each study. There are a number of examples where complete-case analysis is superior, or equal to, to multiple imputation. In particular, complete-case analysis has negligible bias when data is missing completely at random and when missingness is independent of outcome (White and Carlin, 2010).

There are situations in EHR where variables are incompletely recorded. In some circumstances, this is not a major problem, for example if an individual does not have a record of T2DM; it is unlikely that they have T2DM. Blood tests that would have led to the diagnosis (in this case abnormal HbA1C results) may augment diagnostic codes, and this approach is used in the adverse effects study (Study 3). Obviously, this would not be a problem in a specially designed, prospectively followed up cohort, where the researchers may confirm that a participant did not have T2DM, but the EHR reflects real world events.
Therefore, in all studies, outcomes should be regarded as having the prefix “GP recorded diagnosis/symptoms of…”. This approach may be more prone to error for other important variables, namely smoking status, alcohol intake and ethnicity. Potentially these variables are missing not at random (i.e., the missing observations are related to the values of the unobserved data) and if they are not included in a patients registration consultation with a GP they are less likely to be recorded. There is no method of data analysis that can fully account for missingness if this is the case (Sterne et al., 2009). My approach is set out below.

In the studies in this thesis, smoking status is coded as current smoker, ex-smoker, or non-smoker (if there is no record of being a smoker at any time). Smoking status is recorded for 84% of patients within a year of them registering with a THIN GP practice (Marston et al., 2014). If the EHR states that an individual is a smoker, this is likely to be correct (low risk of false positive), but if there is no smoking record it cannot be guaranteed that this person truly doesn’t smoke (risk of false negative) (Lewis and Brensinger, 2004). Despite this concern, previous research has supported this approach with evidence that the vast majority of those with missing smoking records were ex- or non-smokers (Lewis and Brensinger, 2004, Marston et al., 2010, Marston et al., 2014). Alcohol consumption is coded as none (if there is no record of being a drinker at any time) /low or moderate/heavy/dependent. Of all newly registered patients, 76% have an entry for alcohol intake (Khadjesari et al., 2013). Comparisons with British Regional Heart Study and ONS Omnibus Survey on Drinking in Adults suggest that a large number of individuals consuming alcohol within UK recommended limits have missing alcohol consumption records, and that heavy, hazardous and dependent drinking is underreported to GPs (Khadjesari et al., 2013, Marston et al., 2010). Therefore, this approach to categorising alcohol consumption may reduce between group differences. For both smoking and alcohol
intake use of multiple imputation has been shown to fail because of violation of key assumptions, so there is no perfect solution to this problem (Marston et al., 2010). As noted above, ethnicity recording in GP EHRs has been found to be 78% complete after 2006 (Mathur et al., 2014). For the studies in this thesis ethnicity was coded as White, Black, Asian, other (including mixed). Missing ethnicity was coded as White, in line with previous research using primary care data EHR, which suggests that more than 93% of individuals without ethnicity recorded are from a White ethnic group (Hippisley-Cox et al., 2008).

In the adverse effects study (Study 3: Chapter 7) a number of adverse physical outcomes relied on the presence of blood test results. Approaches to managing missing data for these outcomes are discussed in Chapter 7.

### 2.10 Study design to minimise bias and confounding

#### 2.10.1 Mortality and morbidity studies

Study 1 (Chapter 4) compares BPD with schizophrenia; another SMI which has traditionally been considered more severe and resulting in worse outcomes (Kraepelin, 1921). It also compares these groups with a frequency matched comparison group. All-cause mortality has been identified as being well recorded in THIN, with 99.6\% positive predictive value and 99.7\% sensitivity (Hall, 2009). However cause of death is less well recorded (Hall, 2009). To manage this, a number of algorithms for assigning cause of death have been developed. For example, studies of suicide using THIN have defined a suicide death as a code for self-harm, followed by a code for death in the following month and a final date of any activity in the patient record within 6 months (Arana et al., 2010). This approach has a positive predictive value of 97\% for suicide (Arana et al., 2010). Because of concerns about potential under identification, cause-specific mortality hazard ratios (HRs) in this study are likely to be
accurate, but rates may potentially be underestimates compared to the UK population (this is investigated in Chapter 4 and discussed in Chapters 4 and 9).

As discussed in Chapters 1 and 3, the SMR typically accounts for only the age and sex distribution of the population. Study 1 (Chapter 4) examines all-cause mortality HRs in BPD and schizophrenia compared to the UK general population. The simplest model accounts only for age, sex and calendar period. Estimates derived from this model are therefore comparable with studies reporting SMR in the systematic review and meta-analysis (Chapter 3). Additionally, models adjust for sociodemographic factors (ethnicity and area level SES), median number of GP contacts per year and use a robust sandwich estimator for standard errors to account for clustering within GP practice. In the case of CVD diagnosis and mortality an adjunct model examining the potential confounding effects of smoking, BMI, hypercholesterolemia, hypertension and T2DM diagnosis was run.

2.10.2 Maintenance mood stabiliser studies

A key missing variable in the EHR is a measure of illness severity. The starting point for the design of Studies 2 (Chapter 6), 3 (Chapter 7) and 4 (Chapter 8) was the assumption that lithium, valproate, olanzapine and quetiapine were all similarly likely to be prescribed to people with BPD, who reached a threshold for maintenance treatment (i.e., likely to be of similar severity). Beyond this, the majority of reasons for differences in prescribing will be recorded in the EHR or represented by proxy variables. There was no untreated BPD comparison group as this was likely to produce highly biased estimates of treatment effect because the untreated population are likely to be a very different group in terms of a number of factors including illness severity and medication adherence.
During the bulk of the study period NICE recommended lithium, valproate or olanzapine as first-line maintenance mood stabilisers (for both polarities), and quetiapine in the case of bipolar depression (as discussed in Section 1.5). This suggests clinical equipoise between these drugs. Beyond this, clinicians prescribing choices are likely to depend on baseline patient characteristics (including potential for adverse effects), clinician preference and patient preference (Elwyn et al., 2003). Many of the baseline characteristics that would influence drug choice will be recorded in the EHR. For example avoiding drugs known to cause weight gain in those that are already obese, avoiding prescribing in people with a pre-existing contraindication to a given drug, or avoiding drugs that interact with alcohol if a patient is known to be a heavy drinker. Clinician preference (beyond that based on a risk/benefit analysis based on what the treating clinician knows about the patient’s illness and general health) is unmeasured in the EHR, but proxies may exist. For example, most prescribing for BPD will be commenced in secondary care, each GP practice will be served by only one NHS Mental Health Trust, and therefore a limited number of psychiatrists, consequently including an identifier for GP practice will account for some of the variability in prescribing choice. Similarly, patient preference will not be explicitly recorded in the EHR, but it is likely to be associated with variables that are. For example, weight gain, a side effect of olanzapine in particular, might be less acceptable to women who are already overweight, than to men who are not (Regitz-Zagrosek et al., 2006); weight and sex will be well recorded in THIN.

2.11 Approach to analysis to minimise bias and confounding

Analyses are described in each chapter. The following is a summary and discussion of the main approaches used:
2.11.1 Propensity score estimation

2.11.1.1 The propensity score: An approach to managing confounding in observational studies

The concept of PS adjustment was developed by Rosenbaum & Rubin in 1983 with the intent of addressing residual confounding by simulating a randomised environment (Rosenbaum and Rubin, 1983). In non-randomised observational studies, treatment assignment is not arbitrary; therefore, direct comparisons of outcomes in different treatment groups will be misleading. This problem is traditionally minimised by design (for example, matched sampling) or analysis (for example, stratification or covariance adjustment). These methods are limited because they can only employ a limited number of covariates for adjustment. However, PSs provide a scalar summary of the covariate information, and do not have this limitation. This is particularly relevant to pharmacoepidemiology where the focus of the study is often rare outcomes that occur in patients with multiple risk factors and many possible indications and contra-indications for drug use. For instance, it has been found that with fewer than eight events per confounder, analysis based on PSs yields less biased, more robust, and more precise estimates than a regression approach based on logistic regression (Cepeda et al., 2003).

Therefore, the PS is an estimate of the likelihood of an individual receiving a particular treatment calculated using their covariate information (d’Agostino, 1998). It has even been argued that if unmeasured variables are associated with observed variables the approach can reduce bias from these unknown covariates (Austin et al., 2005, Joffe and Rosenbaum, 1999). PS methods derive from a formal model for causal inference, the potential outcomes framework, so that causal questions can be well defined, explicitly specified and not conflated with the modelling approach, as they are with traditional regression approaches.
2.11.1.2 Calculating the propensity score

The PS approach, initially developed for studies comparing two treatment options (Rosenbaum and Rubin, 1983) has been generalised to address the issue of comparing multiple treatments. This approach was initially described by Imbens (Imbens, 2000), with practical examples provided by Spreeuwenberg et al. (Spreeuwenberg et al., 2010). The multiple treatments PS is defined as the conditional probability of receiving a particular treatment, given a set of observed pre-treatment variables. The score is calculated via multinomial logistic regression, where the probability of receiving each treatment is estimated. It has been shown that the multiple treatments PS is a balancing score, like the two treatments version and therefore can be used to correct for initial differences at baseline, and leads to valid estimates in multiple treatment comparisons (Imai and Van Dyk, 2012, Imbens, 2000).

2.11.1.3 Variable selection for the propensity score

Clear guidance on variable selection for PS estimation is sparse (Brookhart et al., 2006). In the studies in this thesis, as recommended, specification of the model was guided by clinical knowledge. The following recommendations were followed in constructing the PS:

i) Include all variables thought to be related to outcome, regardless of whether it is expected that they are related to the exposure, the inclusion of these variables will increase the precision of the estimated exposure effect without increasing bias (Rubin and Thomas, 1996),

ii) Statistically non-significant associations between covariate and exposure are important in PS models, again these variables can increase precision without additional bias (Brookhart et al., 2006),
iii) Addition of a variable unrelated to the outcome, but related to the exposure will increase the variance of an estimated exposure effect, without reducing bias (Brookhart et al., 2006), however;

iv) In medium/large studies covariates related to exposure should not be excluded unless it was known *a priori* that they are not related to outcome (Brookhart et al., 2013),

v) Only variables unaffected by participation should be included in the model; that is variables should be fixed over time or measured at baseline (Caliendo and Kopeinig, 2008),

vi) Covariates included in the model should have 8-10 (exposure) events per variable (Weitzen et al., 2004),

vii) Over-parameterised models should be avoided; this will not bias the score or make it inconsistent, but it may increase variance (Caliendo and Kopeinig, 2008),

viii) Collinearity is not an issue, as this will only affect the precision of the estimated coefficients (and will not result in bias) (Harrell, 2002),

ix) Interaction terms should be included if there is improvement in the resulting balance between treatment groups (Weitzen et al., 2004).

With these points in mind, the PS was built using a structured, non-parsimonious, iterative approach similar to that described by Rosenbaum and Rubin (Rosenbaum and Rubin, 1984). I only considered age and sex as potential interaction terms. The PSs therefore varied slightly for the effectiveness/tolerability, adverse effects and suicide/self-harm studies (Studies 2, 3 and 4).
2.11.1.4 Assessing the propensity score

Plotting PSs can help to understand their distributions and areas of common support (Caliendo and Kopeinig, 2008). Figure 2.ii shows a simplified version of this for two drug treatments. This highlights the fact that there are some individuals who, given their PS, will always be prescribed drug A, or will always be prescribed drug B. There is a range of PSs where individuals have similar chances of receiving drug A or drug B.

Figure 2.ii Potential distribution of propensity scores for two drugs

Adapted from (Schneeweiss, 2010).

The performance of the PS can be assessed by evaluating the balance of covariates (Brookhart et al., 2013). However, thorough methods have not been developed for multiple treatments (McCaffrey et al., 2013). In this thesis, balance was checked following stratification of the score into deciles. For each drug ranges, mean (and standard deviation) and median (and interquartile range) of multiple PSs were compared (Caliendo and Kopeinig, 2008). The distribution of individual covariates within each decile for each drug was then compared (Spreeuwenberg et al., 2010).
2.11.1.5 Using the propensity score

There are four common approaches to using the PS: adjustment, stratification, matching and inverse probability of treatment weighting (IPTW) (Austin, 2011). Adjusting for PS is the most commonly used approach in the clinical literature and is the method most similar to traditional regression modelling (Shah et al., 2005). Simulation studies have suggested that stratification by quintiles of PS will remove over 90% of the bias in each of the covariates that contributes to the PS (Cochran, 1968, d’Agostino, 1998, Rosenbaum and Rubin, 1984).

Both adjusting and stratification techniques use all available data and as such make full use of the generalisability of the dataset. Stratification on the PS can be conceived as a set of quasi-RCTs, which are then combined, weighted by the proportion of subjects in each strata (Austin, 2011). Matching has been described as a more accurate way of estimating treatment effects than stratification or adjustment, because it compares patients with similar observed characteristics. However, it may produce a non-representative sample of patients receiving treatment, because of the patients that are unmatched and therefore dropped from the analysis (d’Agostino, 1998, Rosenbaum and Rubin, 1985). Again, there is no consensus on which matching technique is superior (Caliendo and Kopeinig, 2008). IPTW uses PSs to form a weight, which creates a pseudo or synthetic population in which the covariates and treatment assignment are independent of each other. This approach is similar to the use of survey sampling weights so that results are representative of specific populations (Morgan and Todd, 2008).

Studies 2 (Chapter 6), 3 (Chapter 7) and 4 (Chapter 8) report results adjusted for PS. In each case, this was shown to be preferable to stratification by quintiles or deciles of PS, tested formally using Akaike information criteria and Bayesian information criteria. In the effectiveness/tolerability study (Study 2) and the suicide/self-harm study (Study 4), a matching technique is used. This was not used for the adverse events study (Study 3).
because a number of the outcomes were too rare. Given the aim of estimating treatment effects, and that results of matching were compared with adjustment, a strict matching regimen was employed. Patients taking each other drug were matched with lithium on a one-to-one basis, with their closest possible match, with a caliper (maximum permitted difference between matched subjects) of 0.01. This caliper was used as it falls below the upper limit of 0.25 PS standard deviations recommended by Cochran and Rubin (Cochran and Rubin, 1973). Inverse probability of treatment weighting was not used, apart from as a sensitivity analysis, because there is no evidence that it is superior to other methods (Austin and Stuart, 2015). In addition, adjustment and matching were considered more transparent approaches to data analysis that could be easily communicated to the target audience of these studies: clinicians, patients and key stakeholders.

2.11.2 Survival analyses regression techniques

Survival analysis is a set of techniques for analysing data where the outcome is time until an event of interest. A survival approach is more appropriate than other regression methods, such as linear or logistic regression, because censoring can be handled appropriately.

2.11.2.1 Cox Proportional Hazards model

A commonly used regression model for the analysis of survival data is the Cox proportional hazards regression model (Cox, 1972). It allows testing for differences in survival times of two or more groups of interest, while allowing for adjustment of confounders. The Cox regression model is a semiparametric model, which makes fewer assumptions than typical parametric methods. In particular, it makes no assumptions about the shape of the so-called baseline hazard function. The Cox regression model provides information regarding the relationship of the hazard function to predictors (Cleves, 2008). While a nonlinear relationship between the hazard function and the predictors is assumed, the HR comparing
any two observations is in fact constant over time in the setting where the predictor variables do not vary over time. This assumption is called the proportional hazards assumption and checking if this assumption is met is an important part of a Cox regression analysis.

2.11.2.2 Kaplan-Meier curves & assessing the proportional hazards assumption

For unadjusted survival analysis, Kaplan–Meier analyses are often applied, and provide a way of displaying results graphically. The Kaplan–Meier method estimates the probability of survival up until a certain time point in the presence of censored cases. For subjects whose data are censored, either because they left the cohort or because they reached the end of the study period without an outcome event, all information until their time of censoring is included in the analysis (Bland and Altman, 1998). Results can then be reported as survival probabilities (for example median survival, or 1-, 2-, and 5-year cumulative survival). However, when using the Kaplan–Meier method, the effect size cannot easily be quantified (Jager et al., 2008).

The proportional hazards assumption can be informally checked via Kaplan-Meier curves: plotting survival function against survival time, the shape of the curves should be essentially the same and the separation between curves should remain proportional across analysis time (Hess, 1995). Formal tests of proportionality can be completed by plotting a scaled version of the Schoenfeld residuals (Grambsch and Therneau, 1994, Schoenfeld, 1982).

2.11.2.3 Competing risks regression

In some circumstances, competing risks can be an important problem (Noordzij et al., 2013). A competing risk is an event that hinders observation of the outcome of interest or modifies the likelihood that this event will occur, for example in the adverse events study
(Study 3: Chapter 7); death competes with the physical health outcomes of interest. In this study cumulative incidence competing risk methods are used to display results graphically and provide survival probabilities, rather than traditional Kaplan-Meier plots, which are likely to be inaccurate in this circumstance and cannot adequately account for potential confounders. However, this approach generates sub-distribution HRs, rather than HRs, which cannot be interpreted in the same way as HRs (Noordzij et al., 2013). Therefore, alongside competing risk methods, HRs from Cox regression models are presented.

2.11.2.4 Data analysis

All data analysis was completed using Stata versions 13 and 14 (StataCorp, 2013).
Chapter 3  A systematic review and meta-analysis of mortality in individuals with bipolar disorder

3.1  Summary

3.1.1  Objective

To summarises previous observational studies examining all-cause and cause specific mortality in individuals with a diagnosis of BPD via systematic review and meta-analysis.

3.1.2  Method

Cause-specific mortality was grouped into natural and unnatural causes. These subgroups were further divided into circulatory, respiratory, neoplastic, infectious causes, and suicide and other violent deaths. Summary SMRs were calculated using random-effects meta-analysis. Heterogeneity was examined via subgroup analysis and meta-regression.

3.1.3  Results

Systematic searching found 31 studies meeting inclusion criteria. Summary all-cause SMR was 2.05 (95% CI 1.89 to 2.23) but heterogeneity was high ($I^2=96.2\%$). This heterogeneity could not be accounted for by date of publication, cohort size, mid-decade of data collection, population type or geographic region. Unnatural death summary SMR was 7.42 (95% CI 6.43 to 8.55) and natural death 1.64 (95% CI 1.47 to 1.83). Specifically, suicide SMR was 14.44 (95% CI 12.43 to 16.78), other violent death SMR 3.68 (95% CI 2.77 to 4.90), deaths from circulatory disease SMR 1.73 (95% CI 1.54 to 1.94), respiratory disease SMR 2.92 (95% CI 2.00 to 4.23), infection SMR 2.25 (95% CI 1.70 to 3.00) and neoplasm SMR 1.14 (95% CI 1.10 to 1.21).
3.1.4 Conclusion

Despite considerable heterogeneity, all summary SMR estimates and a large majority of individual studies showed elevated mortality in BPD compared to the general population. This was true for all causes of mortality studied.

A modified version of this chapter was published as Hayes JF, Miles J, Walters K, King M, Osborn DP. A systematic review and meta-analysis of premature mortality in bipolar affective disorder. Acta Psychiatrica Scandinavica. 2015; 131: 417-25
3.2 Introduction

As described in Section 1.4.1, relative to major depression and schizophrenia, there is a limited understanding of the premature mortality associated with BPD. In 1998, Harris and Barraclough reviewed mortality in all mental disorders. Six studies totalling a population of 4547 people contributed to their meta-analysis of mortality in BPD (Harris and Barraclough, 1998). The pooled all-cause SMR was 2.02 (95% CI 1.88 to 2.17). Unnatural deaths SMR was 9.18 (95% CI 8.01 to 10.46) and natural deaths 1.50 (95% CI 1.37 to 1.64). These SMRs estimates are all more elevated than those for schizophrenia from the same systematic review. The schizophrenia mortality meta-analysis was made up of 20 studies including over 35,000 patients (Harris and Barraclough, 1998), highlighting the relative paucity of BPD research. A more recent review, published in 2009, included 13 studies of death by natural causes (Roshanaei-Moghaddam and Katon, 2009). However, this review only searched one database and included patients without a clear diagnosis of BPD (such as mixed unipolar/bipolar groups, schizoaffective disorder or affective psychoses diagnoses). In addition, all-cause mortality and SMRs for unnatural deaths were not investigated. Of the included studies, five produced precise estimates due to reasonable sample size (greater than 2500 individuals). The authors did not perform a meta-analysis but concluded “higher mortality from natural causes among patients with bipolar spectrum disorders ranged from 35% higher than a comparison group to twofold higher”. Since these publications, a number of large cohort studies have reported SMR estimates. Given the paucity of studies examining only BPD mortality before 2009, the systematic review and meta-analysis of these studies is likely to be an important addition to the evidence base.

There is likely to be considerable heterogeneity among SMR estimates, because of both BPD mortality data source, and the data used to generate the expected number of deaths.
Overreliance on either inpatient data or community-based samples is an important limitation that will influence observed death recording. For example, sole use of inpatient data may potentially result in bias and poor generalisability by including only more severe cases, whereas community-based samples may be limited by insufficient sample sizes or loss to follow-up. Heterogeneity may also be introduced by period effects (that is; comparing estimates from different time periods, when different services or treatments were available), and by comparing treated and untreated groups. The choice of data source used to generate the expected number of deaths is also important, and will influence the effect estimate. Of particular impact is whether this is an internal comparison estimated from within the same data source, or data from another source. Internal comparisons may be more valid in terms of important confounders, but may produce an overly healthy comparator group. These factors and other reported study level characteristics will be considered in assessing the heterogeneity of SMR estimates.

3.3 Methods

Existing studies of SMR in BPD were systematically reviewed to examine the association between BPD and all-cause and cause-specific mortality. Cause-specific mortality was grouped into natural and unnatural causes. These subgroups were further divided into suicide and other violent deaths, and deaths from circulatory, respiratory, neoplastic, and infectious causes. Heterogeneity was assessed by geographic region, population type, cohort size, mid-decade of cohort data collection (to account for cohort effects and potential changes in treatment) and decade of publication. I closely followed the guidance provided by the PRISMA statement and MOOSE proposal for reporting (Moher et al., 2009, Stroup et al., 2000). Quality checks were completed by a collaborator Dr Joseph Miles (JM).
3.3.1 Identification of studies

To identify all studies examining mortality in BPD, the Medical Subject Heading (MeSH) terms and keywords for BPD and mortality were searched in PsychINFO, Medline and EMBASE. MeSH terms were: bipolar disorder, mortality, life expectancy, death, death and dying. Keywords searched were: bipolar illness, manic depression, bipolar disorder, bipolar affective disorder, life expectancy, mortality, death (see Appendix 3.1). All databases were searched from their inception until 30 July 2014. JM and I performed the searches individually and then compared results. The abstracts of potentially relevant articles were reviewed by JM and myself. Additional articles and conference papers including primary data were identified from citations in relevant studies and reviews, the Cochrane database of systematic reviews and Google Scholar. Emails were then sent to senior authors of articles that met inclusion criteria to attempt to identify all missing studies. One extra published study was identified by this method, no further unpublished data were made available (Figure 3.i).

3.3.2 Inclusion and exclusion criteria

Included studies met all of the following a priori defined criteria:

i) Published between 1 January 1960 and 1 July 2014,

ii) Reported deaths of individuals diagnosed with BPD; studies were included if BPD was diagnosed by any criteria,

iii) Individuals included in the study were 16 years or older,

iv) Primary data on all-cause mortality or cause specific mortality were included; specific sub-categories of mortality were: natural deaths, unnatural deaths, suicide, other violent deaths, infection, neoplasm, respiratory and circulatory system disease,
v) Reported data on observed and expected deaths, or SMR allowing the number of observed and expected deaths to be calculated.

Figure 3.1 Flow diagram of the published articles evaluated for inclusion in this meta-analysis

Studies were excluded if they:

i) Involved cohorts that could not be defined as having BPD (i.e., studies which grouped together affective disorders),

ii) Included a cohort of less than 50 patients (to avoid including cohorts in which there were no observed deaths),

iii) Were not standardised by age,
iv) Reported mortality in a particular subgroup of the population with BPD (i.e., prison population),

v) Reported duplicate data, or datasets from overlapping time periods at the same site, (if this occurred the most informative paper was then used as the representative mortality estimate for inclusion in the meta-analysis. I.e., larger samples and longer time periods were preferred).

3.3.3 Data extraction
Once a study was included, data were extracted and entered into a database that included the following variables: authors, country, year of publication, years of data collection, length of follow-up, which covariates the mortality was standardised by, the site of collection (i.e., multiple site or population level), population type (i.e., recruited from inpatient or community), number of men and women in the cohort, deaths from all causes and specific causes for both men and women, and population level estimates of expected deaths. JM and I individually extracted data used in the analysis using a standardised form I designed. If disagreements arose, these were resolved by consensus.

3.3.4 Statistical methods
The SMR gives the ratio of death in BPD compared to the general population. For each cause of death, SMRs and their 95% CIs were extracted from each publication or calculated (observed deaths/expected deaths).

The statistical significance of the SMR is based on the Poisson distribution (two-tailed) using 95% CIs. The SMR is significantly raised when the lower CI is greater than 1.00. For each study 95% CIs were calculated using the Rothman–Greenland method (Rothman et al., 2008). Pooled SMRs with 95% CI for all-cause and cause-specific mortality were calculated
using the DerSimonian and Laird method, a random-effects model that incorporates both between-study and within-study variation (DerSimonian and Laird, 1986). Using this method assumes that significant heterogeneity exists between studies (Veroniki et al., 2015).

Statistical heterogeneity was assessed in a number of ways. First, the $I^2$ index and $\chi^2$ test were used to investigate differences among studies with respect to SMRs. Additionally, meta-regression was performed for heterogeneity of the all-cause SMR because of decade of publication, cohort size, geographic region, mid-decade of cohort data collection and population type (i.e., inpatient or community). Subgroup analyses were performed to assess potential sources of heterogeneity separately as a result of the following available patient-level and study-level factors: geographic region of study, patient population type, and decade of the middle year of patient observation. These were considered the key sources of potential bias in the included studies. Funnel plots and Egger’s regression were used to assess for publication and small-study bias in groups containing 10 or more studies (Sterne et al., 2008). All analysis was completed using metan and associated commands in Stata (StataCorp, 2013).

### 3.4 Results

2007, Osby et al., 2001, Saku et al., 1995, Schneider et al., 2001, Sharma and Markar, 1994, Tsuang et al., 1980, Vestergaard and Aagaard, 1991, Weeke et al., 1987, Westman et al., 2013) (Figure 3.i, Table 3.i). Of these, 64% were studies where patients were recruited from inpatient settings. A large number (45%) of studies came from Scandinavian countries (Norway, Sweden, Denmark and Finland). Overall, there were more than 83,919 individuals with a diagnosis of BPD (3 studies did not provide total numbers of patients (Hiroeh et al., 2001, Hiroeh et al., 2008, Hoang et al., 2011)). Data collection ranged from 1935-2010.

The reported SMRs for all-cause mortality in patients with BPD ranged from 1.24 (95% CI 0.83 to 1.17) to 4.65 (95% CI 1.27 to 11.91). Within the 26 individual studies assessing all-cause mortality, all SMR point estimates were elevated, but 4 out of 26 had CIs that overlapped one (Ahrens et al., 1995, Dutta et al., 2007, Schneider et al., 2001, Tsuang et al., 1980), these were all relatively small studies (N<440). The all-cause summary SMR for BPD was 2.05 (95% CI 1.89 to 2.23). There was significant heterogeneity between these studies ($I^2=96.2\%$, 95% CI 95.6 to 96.7, $P<0.001$) (Figure 3.ii).

Sex-specific all-cause mortality showed similarly elevated summary estimates, but again the studies were highly heterogeneous (Table 3.ii). This was also true for mortality grouped as natural and unnatural. Studies had heterogeneous SMR estimates for suicide, other violent deaths, circulatory and respiratory disease mortality. SMR estimates for, infectious and neoplastic deaths were more homogenous. Summary estimates suggested increased rates of death from all causes in individuals with BPD (Table 3.ii).
Table 3. Studies include in the meta-analysis

<table>
<thead>
<tr>
<th>Country</th>
<th>Years of collection</th>
<th>Total N</th>
<th>N Men</th>
<th>N Women</th>
<th>Standardised by</th>
<th>Site of collection</th>
<th>Population type</th>
<th>Mortality outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bratfos &amp; Haug (1968)</td>
<td>Norway</td>
<td>1950-1963</td>
<td>207</td>
<td>0/0</td>
<td>Age, Sex</td>
<td>Single site</td>
<td>Inpatient</td>
<td>All-cause, suicide</td>
</tr>
<tr>
<td>Kay &amp; Petterson (1977)</td>
<td>Sweden</td>
<td>1961-1970</td>
<td>192</td>
<td>84</td>
<td>Age, sex</td>
<td>Multiple site</td>
<td>Inpatient</td>
<td>All-cause</td>
</tr>
<tr>
<td>Tsuang et al. (1980)</td>
<td>United States</td>
<td>1935-1974</td>
<td>100</td>
<td>45</td>
<td>Age, sex</td>
<td>Single site</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, infectious, neoplasm, circulatory</td>
</tr>
<tr>
<td>Weeke et al. (1987)</td>
<td>Denmark</td>
<td>1969-1976</td>
<td>417</td>
<td>185</td>
<td>Age, sex, time at risk</td>
<td>Inception cohort</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, suicide, other violent, neoplasm, circulatory</td>
</tr>
<tr>
<td>Newman &amp; Bland (1991)</td>
<td>Canada</td>
<td>1976-1985</td>
<td>1429</td>
<td>0/0</td>
<td>Age, sex, time at risk</td>
<td>Multiple site</td>
<td>Inpatient &amp; community</td>
<td>All-cause, suicide</td>
</tr>
<tr>
<td>Vestergaard &amp; Aagaard (1991)</td>
<td>Denmark</td>
<td>1981-1988</td>
<td>133</td>
<td>0/0</td>
<td>Age</td>
<td>Multiple site</td>
<td>Inpatient</td>
<td>All-cause, suicide, circulatory, respiratory</td>
</tr>
<tr>
<td>Jorgensen &amp; Mortensen (1992)</td>
<td>Denmark</td>
<td>1970-1988</td>
<td>18293</td>
<td>0/0</td>
<td>Age, sex</td>
<td>Inception cohort</td>
<td>Inpatient</td>
<td>All-cause</td>
</tr>
<tr>
<td>Sharma et al. (1994)</td>
<td>United Kingdom</td>
<td>1970-1987</td>
<td>472</td>
<td>0/0</td>
<td>Age</td>
<td>Population</td>
<td>Inpatient &amp; community</td>
<td>All-cause, suicide, circulatory, respiratory</td>
</tr>
<tr>
<td>Ahrens et al. (1995)</td>
<td>Denmark, Germany, Canada, Austria</td>
<td>1962-1992</td>
<td>440</td>
<td>189</td>
<td>Age, sex</td>
<td>Multiple site</td>
<td>Inpatient &amp; community</td>
<td>All-cause, suicide, circulatory</td>
</tr>
</tbody>
</table>

Continued overleaf
<table>
<thead>
<tr>
<th>Country</th>
<th>Years of collection</th>
<th>Total N</th>
<th>N Men</th>
<th>N Women</th>
<th>Standardised by</th>
<th>Site of collection</th>
<th>Population type</th>
<th>Mortality outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hiroeh et al. (2001)</td>
<td>Denmark</td>
<td>1973-1993</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Inpatient Unnatural, suicide, other violent</td>
</tr>
<tr>
<td>Schneider et al. (2001)</td>
<td>Germany</td>
<td>1983-1988</td>
<td>74</td>
<td>24</td>
<td>50</td>
<td>Age, sex</td>
<td>Single site</td>
<td>Inpatient All-cause, unnatural</td>
</tr>
<tr>
<td>Angst et al. (2002)</td>
<td>Switzerland</td>
<td>1959-1997</td>
<td>220</td>
<td>N/A</td>
<td>N/A</td>
<td>Age</td>
<td>Single site</td>
<td>Inpatient All-cause, unnatural, natural, suicide, other violent, neoplasm, circulatory, respiratory</td>
</tr>
<tr>
<td>Amaddeo et al. (2007)</td>
<td>Italy</td>
<td>1982-2001</td>
<td>278</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Community All-cause</td>
</tr>
<tr>
<td>Dutta et al. (2007)</td>
<td>United Kingdom</td>
<td>1966-1999</td>
<td>135</td>
<td>102</td>
<td>133</td>
<td>Age, sex</td>
<td>Inception cohort</td>
<td>Community All-cause, natural, suicide, neoplasm</td>
</tr>
<tr>
<td>Osborn et al. (2007)</td>
<td>United Kingdom</td>
<td>1987-2002</td>
<td>10742</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Community Circulatory</td>
</tr>
<tr>
<td>Hiroeh et al. (2008)</td>
<td>Denmark</td>
<td>1973-1993</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Inpatient Natural, infectious, neoplasm, circulatory, respiratory</td>
</tr>
<tr>
<td>Osborn et al. (2008)</td>
<td>United Kingdom</td>
<td>1987-2002</td>
<td>10742</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Community Suicide</td>
</tr>
<tr>
<td>Chang et al. (2010)</td>
<td>United Kingdom</td>
<td>2007-2009</td>
<td>2700</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Inception cohort</td>
<td>Community All-cause</td>
</tr>
<tr>
<td>Hoang et al. (2011)</td>
<td>United Kingdom</td>
<td>1999-2006</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Age, sex</td>
<td>Population</td>
<td>Inpatient All-cause, unnatural, natural</td>
</tr>
<tr>
<td>Nordentoft et al. (2011)</td>
<td>Denmark</td>
<td>1970-2006</td>
<td>5927</td>
<td>2571</td>
<td>3356</td>
<td>Age, sex</td>
<td>Inception cohort</td>
<td>Inpatient &amp; community Suicide</td>
</tr>
</tbody>
</table>

Continued overleaf
<table>
<thead>
<tr>
<th>Country</th>
<th>Years of collection</th>
<th>Total N</th>
<th>N Men</th>
<th>N Women</th>
<th>Standardised by</th>
<th>Site of collection</th>
<th>Population type</th>
<th>Mortality outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ajetunmobi et al. (2013)</td>
<td>United Kingdom</td>
<td>1986-2010</td>
<td>3839</td>
<td>N/A</td>
<td>N/A</td>
<td>Inception cohort</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, suicide, infectious, neoplasm, circulatory</td>
</tr>
<tr>
<td>Hoang et al. (2013)</td>
<td>United Kingdom</td>
<td>2006-2008</td>
<td>14017</td>
<td>N/A</td>
<td>N/A</td>
<td>Population</td>
<td>Inpatient</td>
<td>All-cause</td>
</tr>
<tr>
<td>Laursen et al. (2013)a</td>
<td>Sweden</td>
<td>2000-2007</td>
<td>18355</td>
<td>7367</td>
<td>10988</td>
<td>Population</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, circulatory</td>
</tr>
<tr>
<td>Laursen et al. (2013)b</td>
<td>Finland</td>
<td>2000-2007</td>
<td>9919</td>
<td>4489</td>
<td>5430</td>
<td>Population</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, circulatory</td>
</tr>
<tr>
<td>Laursen et al. (2013)c</td>
<td>Denmark</td>
<td>2000-2007</td>
<td>11101</td>
<td>4280</td>
<td>6821</td>
<td>Population</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, circulatory</td>
</tr>
<tr>
<td>Westman et al. (2013)</td>
<td>Sweden</td>
<td>1987-2006</td>
<td>17101</td>
<td>8208</td>
<td>8893</td>
<td>Population</td>
<td>Inpatient</td>
<td>All-cause, unnatural, natural, circulatory</td>
</tr>
</tbody>
</table>

N/A, not available; Hiroeh et al. (2001) and Hiroeh et al. (2008) presented person-years at risk (PYAR) rather than individuals: men = 155,337 PYAR, women 309,639 PYAR, Hoang et al. (2011) reported number of hospital discharges (100,851) but did not identify number of patients that this represented
In univariable meta-regression all-cause mortality was not significantly associated with decade of publication (P=0.63), cohort size (P=0.75), geographic region (P=0.55) mid-decade of cohort data collection (P=0.89) or population type (P=0.65). After accounting for all of these possible explanatory variables in multivariable meta-regression, residual variation due to heterogeneity amongst all-cause mortality SMRs remained ($I^2=88.3\%$, 95% CI 84.6-90.7, P<0.001).
### Table 3.ii Summary standardised mortality ratios for all-cause and cause specific mortalities

<table>
<thead>
<tr>
<th></th>
<th>No. of studies</th>
<th>Number of individuals</th>
<th>Summary SMR (95% CI)</th>
<th>I² (95% CI)</th>
<th>Het P</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause</td>
<td>26</td>
<td>220,134</td>
<td>2.05 (1.89–2.23)</td>
<td>96.2 (95.6–96.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>15</td>
<td>34,636</td>
<td>2.17 (2.01–2.34)</td>
<td>83.6 (74.8–88.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>15</td>
<td>46,075</td>
<td>2.11 (1.93–2.31)</td>
<td>84.4 (83.4–91.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Natural</td>
<td>12</td>
<td>159,495</td>
<td>1.64 (1.47–1.83)</td>
<td>98.0 (98.0–98.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>9</td>
<td>34,220</td>
<td>1.72 (1.54–1.93)</td>
<td>94.5 (92.5–95.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>9</td>
<td>45,598</td>
<td>1.74 (1.51–1.99)</td>
<td>97.6 (97.1–98.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unnatural</td>
<td>12</td>
<td>159,434</td>
<td>7.42 (6.43–8.55)</td>
<td>95.6 (93.9–96.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>9</td>
<td>34,142</td>
<td>7.89 (7.05–8.81)</td>
<td>82.4 (68.1–88.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>9</td>
<td>45,515</td>
<td>9.23 (7.14–11.94)</td>
<td>96.6 (95.6–97.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Suicide</td>
<td>15</td>
<td>46,756</td>
<td>14.44 (12.43–16.78)</td>
<td>87.0 (80.4–90.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>9</td>
<td>12,325</td>
<td>13.31 (10.62–16.69)</td>
<td>87.8 (78.7–91.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>9</td>
<td>16,698</td>
<td>15.74 (12.84–19.31)</td>
<td>81.7 (63.4–88.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Other violent</td>
<td>5</td>
<td>22,641</td>
<td>3.68 (2.77–4.90)</td>
<td>89.5 (77.2–93.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>4</td>
<td>9,463</td>
<td>3.06 (2.19–4.03)</td>
<td>86.2 (57.6–92.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>4</td>
<td>12,958</td>
<td>5.53 (4.60–19.14)</td>
<td>99.0 (98.7–99.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Circulatory</td>
<td>14</td>
<td>153,948</td>
<td>1.73 (1.54–1.94)</td>
<td>95.2 (93.9–96.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>9</td>
<td>34,041</td>
<td>1.81 (1.61–2.05)</td>
<td>90.3 (85.1–93.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>9</td>
<td>45,396</td>
<td>1.72 (1.46–2.03)</td>
<td>96.0 (94.8–96.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Respiratory</td>
<td>5</td>
<td>22,609</td>
<td>2.92 (2.00–4.23)</td>
<td>94.14 (89.8–96.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>men</td>
<td>4</td>
<td>9,278</td>
<td>2.73 (1.76–4.24)</td>
<td>90.2 (75.9–94.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>women</td>
<td>4</td>
<td>12,726</td>
<td>2.72 (1.78–4.20)</td>
<td>91.3 (80.0–95.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Infection</td>
<td>5</td>
<td>22,895</td>
<td>2.25 (1.70–3.00)</td>
<td>45.4 (0.0–78.5)</td>
<td>0.12</td>
</tr>
<tr>
<td>men</td>
<td>4</td>
<td>9,323</td>
<td>2.76 (1.94–3.92)</td>
<td>43.1 (0.0–80.0)</td>
<td>0.15</td>
</tr>
<tr>
<td>women</td>
<td>4</td>
<td>12,781</td>
<td>1.77 (1.30–2.40)</td>
<td>20.9 (0.0–4.1)</td>
<td>0.29</td>
</tr>
<tr>
<td>Neoplasm</td>
<td>10</td>
<td>27,693</td>
<td>1.14 (1.10–1.21)</td>
<td>20.7 (0.0–61.9)</td>
<td>0.25</td>
</tr>
<tr>
<td>men</td>
<td>7</td>
<td>9,729</td>
<td>1.11 (1.05–1.17)</td>
<td>0.0 (0.0–58.5)</td>
<td>0.47</td>
</tr>
<tr>
<td>women</td>
<td>7</td>
<td>13,214</td>
<td>1.19 (1.05–1.37)</td>
<td>56.3 (0.0–79.3)</td>
<td>0.03</td>
</tr>
</tbody>
</table>

SMR, standardised mortality ratio; CI, confidence interval; I², index of heterogeneity; Het P from X² test; T not including individuals in studies Hiroeh et al. (2001), Hiroeh et al. (2008) and Hoang et al. (2011)

Subgroup analyses were performed stratified by geographic region, population type and mid-decade of study (Table 3.iii). Stratifying by these covariates had little effect on heterogeneity in summary estimates for all-cause, natural and unnatural death SMRs, which remained high.
Table 3.iii Summary SMRs by subgroup for all-cause mortality, and natural and unnatural death

<table>
<thead>
<tr>
<th></th>
<th>No. of Studies</th>
<th>Summary SMR (95% CI)</th>
<th>I² (95% CI)</th>
<th>Het P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-cause</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scandinavia</td>
<td>10</td>
<td>2.20 (2.01–2.41)</td>
<td>96.5 (95.2–97.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>North America</td>
<td>3</td>
<td>1.77 (1.24–2.53)</td>
<td>81.4 (42.7–94.0)</td>
<td>0.005</td>
</tr>
<tr>
<td>UK</td>
<td>8</td>
<td>2.06 (1.74–2.34)</td>
<td>91.7 (86.0–95.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Other European</td>
<td>4</td>
<td>1.67 (1.18–2.38)</td>
<td>51.7 (0.0–84.1)</td>
<td>0.1</td>
</tr>
<tr>
<td>Japan</td>
<td>1</td>
<td>1.93 (1.44–2.52)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Population type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>17</td>
<td>2.05 (1.86–2.25)</td>
<td>97.3 (96.9–97.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Inpatient &amp; community</td>
<td>6</td>
<td>2.21 (1.79–2.73)</td>
<td>80.6 (49.1–89.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Community</td>
<td>3</td>
<td>1.85 (1.16–2.94)</td>
<td>77.6 (0.0–91.1)</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Mid-point of study</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950s</td>
<td>2</td>
<td>1.86 (0.88–3.94)</td>
<td>90.9*</td>
<td>0.001</td>
</tr>
<tr>
<td>1960s</td>
<td>3</td>
<td>1.85 (1.58–2.17)</td>
<td>0.0 (0.0–72.9)</td>
<td>0.7</td>
</tr>
<tr>
<td>1970s</td>
<td>6</td>
<td>1.93 (1.67–2.24)</td>
<td>82.5 (60.8–89.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1980s</td>
<td>6</td>
<td>2.30 (1.84–2.87)</td>
<td>84.1 (62.3–90.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1990s</td>
<td>3</td>
<td>2.12 (1.42–3.14)</td>
<td>99.3 (99.0–99.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2000s</td>
<td>5</td>
<td>2.13 (1.90–2.39)</td>
<td>95.7 (93.9–96.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Natural</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scandinavia</td>
<td>6</td>
<td>1.79 (1.56–2.05)</td>
<td>98.8 (98.6–99.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>North America</td>
<td>2</td>
<td>1.34 (0.90–2.00)</td>
<td>64.9*</td>
<td>0.09</td>
</tr>
<tr>
<td>UK</td>
<td>3</td>
<td>1.42 (0.95–2.12)</td>
<td>98.7 (98.1–99.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Other European</td>
<td>1</td>
<td>1.40 (1.17–1.66)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Population type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>10</td>
<td>1.67 (1.48–1.88)</td>
<td>98.5 (98.3–98.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Inpatient &amp; community</td>
<td>1</td>
<td>1.79 (1.68–1.91)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Community</td>
<td>1</td>
<td>1.03 (0.71–1.44)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mid-point of study</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950s</td>
<td>1</td>
<td>1.10 (0.79–1.49)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1960s</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1970s</td>
<td>3</td>
<td>1.36 (1.23–1.51)</td>
<td>0.0 (0.0–72.9)</td>
<td>0.43</td>
</tr>
<tr>
<td>1980s</td>
<td>3</td>
<td>1.51 (1.14–2.00)</td>
<td>99.1 (98.8–99.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1990s</td>
<td>2</td>
<td>2.06 (2.00–2.12)</td>
<td>0.0*</td>
<td>0.62</td>
</tr>
<tr>
<td>2000s</td>
<td>3</td>
<td>1.75 (1.50–2.03)</td>
<td>97.1 (95.8–97.9)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Continued overleaf*
No. of Studies | Summary SMR (95% CI) | $I^2$ (95% CI) | Het $P$  
--- | --- | --- | ---  
**Unnatural**  
Geographic region  
Scandinavia | 6 | 8.15 (7.20–9.24) | 94.0 | <0.001  
North America | 2 | 3.12 (1.97–4.96) | 0* | 0.75  
UK | 2 | 7.90 (3.16–19.73) | 99.2* | <0.001  
Other European | 2 | 5.60 (2.42–12.95) | 31.8* | 0.23  
Population type  
Inpatient | 11 | 7.56 (6.52–8.77) | 95.4 (94.0–96.3) | <0.001  
Inpatient & community | 1 | 6.05 (5.14–7.12) | - | -  
Community | 0 | - | - | -  
Mid-point of study  
1950s | 1 | 2.93 (1.52–5.13) | - | -  
1960s | 0 | - | - | -  
1970s | 3 | 5.46 (3.03–9.85) | 84.3 (15.7–93.0) | 0.002  
1980s | 3 | 8.75 (6.39–11.97) | 97.3 (95.4–982) | <0.001  
1990s | 2 | 6.93 (3.60–13.34) | 98.9* | <0.001  
2000s | 3 | 8.26 (6.53–10.44) | 94.9 (91.6–96.6) | <0.001  

*SMR, standardised mortality ratio; CI, confidence interval; $I^2$, index of heterogeneity; Het $P$ from $X^2$ test. *95% CI cannot be calculated due to 1 degree of freedom.

For all-cause mortality SMR Eggers test did not suggest significant publication bias ($P=0.63$).

The same was true for SMR of unnatural deaths ($P=0.55$) and suicide ($P=0.40$). However, given the high heterogeneity, publication bias cannot be ruled out with certainty in these groups. It is even more likely to be present in studies of natural deaths ($P=0.05$), and circulatory disease ($P=0.17$) where $P$-values are closer to the established threshold of $P<0.05$.

### 3.5 Discussion

**3.5.1 Main findings**

This review of mortality in patients with BPD highlights the increased risk of death from all causes. Summary SMR estimates from random effects meta-analysis showed that all-cause mortality in BPD is double that expected in the general population. Natural deaths are over 1.5 times greater in BPD than the general population; these natural deaths are made up of
an almost double risk of deaths from circulatory illnesses (heart attacks, strokes, etc.) and 3 times risk of deaths from respiratory illness (COPD, asthma, etc.). Unnatural deaths are around 7 times more common, with increased risk of suicide of around 14 times and other violent deaths (accidents, homicide, etc.) almost 4 times as likely. Deaths by all causes were similarly elevated in both men and women. It is particularly concerning that having BPD in the 2000s has the same mortality risk, compared to the general population, as it did in the 1950s. With the increased use of SGAs in these cohorts (Alexander et al., 2011, Hayes et al., 2011) and associated elevated risk of CVD; the failure of smoking cessation policy to address the needs of the severely mentally ill, relative to the general population (Cassidy et al., 2001, Hert et al., 2011); and the continued lack of equality in access to healthcare for individuals with BPD (Hert et al., 2011), this gap in deaths from medical illness may widen unless it is directly addressed. In terms of modifying the increased rate of unnatural deaths, particular attention needs to be paid to comorbid substance misuse, exploitation, receipt and perpetration of violence, and suicidal ideation (Cassidy et al., 2001, Gonda et al., 2012, Khalifeh and Dean, 2010).

Heterogeneity across studies was high (less so for cancer, and infection deaths). Heterogeneity in all-cause mortality, natural and unnatural death SMRs could not be accounted for by year of publication, study size, mid-point of data collection, geographical region or population type. Whilst it is possible that some of these factors were imperfectly adjusted for in the analysis (for example some cohorts spanned many decades), the results suggest that there are unidentified factors that led to differences in outcomes for different cohorts of patients with BPD. It has been found that patients with BPD in the United States have worse physical health and greater comorbidity than those in Germany and the Netherlands (Post et al., 2014), however it may be the case that there are even more localised differences in BPD outcomes. This meta-analysis suggests that within the US or
Europe, mortality estimates for BPD are not homogenous. Adjusting for cohort size did little to reduce the heterogeneity, despite the increased accuracy in SMR estimates larger studies should provide. Stratifying by decade of data collection did not affect heterogeneity, suggesting that the differences are not down to improvements in treatment over time. Studies of inpatient populations and community cohorts were also heterogeneous, suggesting these differences are not down to engagement with services or severity of illness.

3.5.2 Potential limitations

There are several limitations of studies included in this review. SMRs were often only age and sex adjusted; therefore, other characteristics of the study populations such as illness duration, treatment and lifestyle factors, may have contributed to the significant heterogeneity. For example, it was not possible to assess whether current or former smoking contributed to excess respiratory mortality. Disease severity was not assessed in all of the included studies, and therefore, it is not possible to assess heterogeneity in the overall mortality by severity. It has been recognised that patients with BPD accumulate numerous medical risk factors including smoking, poor nutrition, use of alcohol and other illicit drugs, prescribed medication and comorbid anxiety and eating disorders that lead to earlier disease onset, poor engagement with healthcare and poor long-term outcomes (Kilbourne et al., 2004). These risk factors may be assigned differentially, both geographically and temporally. In many of the included studies, cause of death was ascertained from death certificates and therefore may be subject to potential misclassification bias. Having a mental health diagnosis has been shown to increase the risk of a coroner’s verdict of suicide rather than accidental or undetermined death (Rosenberg et al., 1988) but may also reduce diagnosis of terminal illness leading to miscoding of physical cause of death (Hert et al., 2011).
It has been argued that in the event of high heterogeneity, meta-analysis is inappropriate as the pooled effect estimate represents the mean distribution of SMRs from included studies, rather than a potentially “true” effect estimate (Hedges and Vevea, 1998). One aims of this analysis was to attempt to identify reasons for this heterogeneity and so a random-effects approach was taken. A random-effects meta-analysis model involves an assumption that the effects being estimated in the different studies are not identical, but follow some distribution. The model represents the lack of knowledge about why real, or apparent, SMRs differ by considering the differences as if they were random.

Additionally, it was not possible to fully assess study quality. I considered assessing methodological quality via a tool such as the Newcastle–Ottawa Scale (Stang, 2010). However, many of the points included in this tool are not applicable to the studies in question, for example the definition of outcome does not vary; it is always death, nor does the definition of control population; it is always overall mortality in the general population. I also considered adding that studies needed to have explicit inclusion/exclusion criteria and operationalised diagnostic criteria to be considered ‘high quality’. However, this tended to make scores more similar; single site studies of inpatient populations score well on these criteria, whereas population-based cohorts score badly (despite population-based cohorts clearly being more generalisable). Universally reported factors that could introduce biased estimates were assessed, these included the following: country of study, year of publication, years of data collection, factors for standardisation, site of data collection (population, single site and multisite), population type (inpatient or community). I did not feel combining these factors into a ‘score’ would have improved the analysis or its interpretation.
3.5.3 Conclusions

This meta-analysis highlights differential mortality in patients with BPD and the general population. Similarly to schizophrenia, patients with BPD have over twice the all-cause mortality (Saha et al., 2007). Mortality from all physical conditions and unnatural causes is elevated. Variation in all-cause mortality is considerable across time and place. There is no evidence that all-cause mortality for patients with BPD has improved over time relative to the general population.

3.5.4 Implications of the findings from this systematic-review and meta-analysis

Generalisable and timely measures of mortality in BPD will become more available with the development of EHRs. The numbers of patients in these datasets will provide sufficient power to analyse mortality and other negative outcomes in BPD, whereas previously opportunities have been limited because of the low prevalence of the disorder and the tendency to focus only on follow-up of inpatient samples. Subgroup analysis, by treatment received or illness severity will also be possible. Study 1 (Chapter 4) partially addresses this research need using a population-based cohort from THIN, which is representative of the UK population, and examines all-cause mortality HRs and cause specific HRs for CVD deaths and suicide.
Chapter 4  Mortality and morbidity in individuals with bipolar disorder and schizophrenia, compared to the general population

4.1  Summary

4.1.1  Objective

To calculate trends in all-cause mortality in the UK in individuals with BPD compared to individuals with schizophrenia and the general population, and to calculate rates of i) CVD deaths, ii) suicide, iii) CVD diagnoses, iv) self-harm in individuals diagnosed with BPD or schizophrenia compared to the general population, while accounting for sociodemographic factors.

4.1.2  Method

A longitudinal cohort study conducted in a nationally representative UK sample using primary care EHR data collected between January 1, 2000, and December 31 2014. All patients diagnosed as having BPD or schizophrenia and a frequency matched comparison group of the general population were included. The primary outcome was all-cause mortality. Secondary outcomes were cardiovascular deaths, CVD diagnoses, suicide and self-harm.

4.1.3  Results

Among 17,341 individuals with BPD and 22,497 with schizophrenia, 1,266 and 2,061 respectively died during follow-up. Individuals with BPD had an all-cause mortality 1.79 times (95% CI 1.67 to 1.88) and those with schizophrenia 2.08 times (95% CI 1.98 to 2.19) that of the general population, accounting for sociodemographic characteristics. Adjusted
HRs were stable in BPD between 2000 and 2006, and then increased by 0.14 per year (95% CI 0.10 to 0.19). HRs for schizophrenia fell until 2004 (-0.29 per year; 95% CI -0.48 to -0.10), increased gradually between 2004 and 2010 (0.11 per year; 95% CI 0.04 to 0.17) and increased more rapidly after 2010 (0.34 per year; 95% CI 0.18 to 0.49). Cardiovascular mortality was elevated in those with schizophrenia (HR 1.39; 95% CI 1.12 to 1.73) and greatly so in those aged 50 years and under (HR 3.20; 95% CI 1.62 to 6.31), but not in BPD. The HR for CVD diagnosis suggests under-recording in those under 50 with schizophrenia. Suicide rates were elevated in both BPD (HR 12.66; 95% CI 7.79 to 20.58) and schizophrenia (HR 7.21; 95% CI 4.26 to 12.19) as were rates of self-harm (HR 25.24; 95% CI 22.37 to 28.29 and HR 22.14; 95% CI 19.58 to 25.03 respectively).

4.1.4 Conclusion

Despite falling mortality rates in individuals with BPD and schizophrenia, the gap between BPD and schizophrenia mortality, and mortality in the general population became wider between the mid-2000s and 2014. Death from cardiovascular disease is markedly elevated in those younger than 50 with schizophrenia. Death from suicide is similarly elevated in BPD and schizophrenia relative to the general population.

A modified version of this chapter was published as Hayes JF, Marston L, Walters K, King M, Osborn DP. Widening mortality gap for people with bipolar disorder and schizophrenia: UK based cohort study 2000-2014. British Journal of Psychiatry. 2017; DOI: 10.1192/bjp.bp.117.202606
4.2 Introduction

As discussed in Chapter 1, death rates are increased in people with severe SMI relative to the general population, and this translates to around 20-years of premature mortality (Laursen, 2011, Laursen et al., 2013). This has been found in a number of longitudinal studies (Hoang et al., 2013, Hoang et al., 2011, Saha et al., 2007, Tiihonen et al., 2009). It has been reported that the mortality gap has narrowed or plateaued since the mid-1990s (Bushe et al., 2010), but there are no studies examining this using UK data.

Since the turn of the millennium, a number of strategies aimed at reducing the mortality gap between people with SMI and the general population have been implemented in the UK NHS (Colton and Manderscheid, 2006, Department of Health, 2011, 2014, Doran et al., 2011, Edwards and McGorry, 2002, Roland, 2004, Schizophrenia Commission, 2012, Swinson et al., 2007). Additionally, the UK age-standardised mortality rate in the general population has declined. Between 2000 and 2014 the annual age-standardised mortality rate fell by approximately 20% (Office for National Statistics, 2016). Therefore, any intervention targeted at mortality in people with SMI would have to reduce mortality at a rate greater than this to reduce the mortality gap. Few previous studies cover SMI mortality during this period. It is therefore timely to review all-cause and cause-specific mortality rates in individuals diagnosed with BPD and schizophrenia relative to the general population. Schizophrenia has tended to be the “target” diagnosis in SMI, so it is important to understand how BPD mortality compares to a group that has been more commonly the focus of policy and research. I used CVD mortality and suicide as exemplars of natural and unnatural causes of death that have been targets of mental health policy.

Evidence suggests that CVD is the leading cause of death in individuals with SMI (Laursen et al., 2013, Roshanaei-Moghaddam and Katon, 2009, Weiner et al., 2011). However, large
representative longitudinal studies remain limited in number. For example, the only previous population based study, of which I am aware, used data from 1987-2002 and found that people aged under 50 with schizophrenia had over three times the rate of CVD mortality (Osborn et al., 2007). Data from this study are now over 15 years old; it predates the increase in SGA use and the efforts to reduce mortality in the SMI population. Osborn et al. also found that HRs for CVD incidence were smaller than HRs for CVD death. This suggests that those with SMI may present less, get diagnosed less or receive correct treatment less frequently (Osborn et al., 2007). Although lack of access to appropriate treatment has also been reported elsewhere (Kurdyak et al., 2012, Laursen et al., 2009, Newcomer and Hennekens, 2007, Smith et al., 2013), an alternative has been postulated: individuals with SMI may truly be at greater risk of unheralded coronary events, potentially due to different pathology (e.g., more plaque rupture) (Leboyer et al., 2012).

In the UK a number of initiatives have been targeted at improving physical health of those with SMI, particularly CVD, which are intended to lead to earlier diagnosis of CVD (Roland, 2004), reducing CVD risk factors (Cormac, 2009, McCreadie, 2003) and effective treatment implementation (Miller, 2009). However, studies are yet to investigate if CVD deaths in people with SMI have fallen because of these interventions, and if inequalities in access to care have reduced. It is unclear if age and sex differences remain the same as previous studies. Also, as far as I am aware the contribution of lifestyle and presence of cardiovascular risk factors as an explanation for the potential increase in CVD mortality and CVD diagnosis has not been explored.

As shown in Chapter 3, Suicide is the cause of death that is most elevated in individuals with BPD and this is also true in schizophrenia, compared to the general population (Brown, 1997). It remains unclear if suicide is elevated in BPD compared to schizophrenia, as
previous studies have given inconsistent results (Mortensen et al., 2000, Osborn et al., 2008, Tidemalm et al., 2008). Up-to-date estimates of suicide rates are important markers of the success of psychiatric care (Swinson et al., 2007). Self-harm is the major risk factor for suicide, with a large number of those dying by suicide having a history of self-harm (Owens et al., 2002). Self-harm is also a marker of quality of life and emotional distress in individuals with SMI (Singhal et al., 2014).

This study compares rates of all-cause mortality, CVD death and CVD, suicide and self-harm in people with BPD, and schizophrenia and a general population comparator group from 2000 to 2014.

4.3 Methods

4.3.1 Study design and setting

This cohort study was completed using pseudonymised primary care EHRs from THIN, as discussed in Section 2.3. Data were included from January 1 2000, until December 31 2014 (Figure 2.i, Figure 4.i).
4.3.2 Participants

All individuals aged 16 or over, ever receiving a diagnosis of BPD or schizophrenia were included in the cohort. If individuals had multiple diagnoses they were classified by the diagnosis most recently assigned. Patients with schizoaffective disorders and unipolar depression were excluded. The validity of diagnoses of BPD and schizophrenia in primary care records is discussed in Section 2.5.2. Individuals with BPD and schizophrenia were frequency matched with up to six individuals without these diagnoses to create a comparator group. The comparator group was matched on age (in 5-year age bands) and sex, from within the same primary care practice (Section 2.5.3).
4.3.3 Outcomes

CVD was defined as any entry of myocardial infarction (MI), ischemic heart disease (IHD) or cerebrovascular event (CVE) in the longitudinal EHR. CVD diagnoses have previously been validated in THIN with positive predictive values greater than 90% (Hammad et al., 2008, Ruigómez et al., 2010). In line with other studies CVD mortality was defined as a death code with a CVD cause, or a CVD code, followed by a death code in the following sixty days and a final date of any activity in the EHR within six months (Ogdie et al., 2014).

The definition for self-harm events included Read codes for intentional poisoning, intentional self-injurious behaviour, and self-harm acts of uncertain intent. This unitary definition of self-harm, where there is no distinction made between non-suicidal self-harm and self-harm with suicidal intent is consistent with UK research norms (Haw et al., 2015). The positive predictive value of this outcome in THIN has been shown to be 97% (Arana et al., 2010). Suicide was defined as a death code identified as suicide or a self-harm code followed by a death code in the following thirty days and a final date of any activity in the EHR within six months, in line with previous research (Arana et al., 2010).

4.3.4 Statistical analyses

Rates of all-cause, cause specific mortality (CVD deaths, suicide) and morbidity (CVD, self-harm) were calculated. Annual rates were calculated for all-cause mortality. In order to assess trends in rates over time, a segmented regression analysis using joinpoint models was performed (Kim et al., 2000, Wagner et al., 2002). To complete this analysis the Surveillance Research Program of the United States National Cancer Institute Joinpoint software was used (Version 4.3.1.0 (Statistical Methodology and Applications Branch, 2016)). This analysis identifies time points where there is a change in the linear slope of the trend. The optimum number of linear slopes and joinpoints is assessed using modified
Bayesian Information Criteria (Zhang and Siegmund, 2007). The models incorporate estimated variation for each data point using the standard error of the rate. After identification of a change in trend, segmented regression can be fitted and annual percentage change in rate (with 95% CIs) can be calculated.

Cox proportional hazards regression analyses were conducted comparing rates of all-cause mortality, in individuals with BPD, schizophrenia and the matched comparison group. The assumption of proportional hazards was tested by analysis of Schoenfeld residuals (Schoenfeld, 1982). A number of multivariable models were tested. Firstly, the association between diagnosis and all-cause mortality was assessed, adjusting for age, sex and calendar year and clustering within primary care practices, then additionally adjusting for area level deprivation (defined as quintiles of Townsend score – a proxy for SES (Townsend, 1987), based on the patients lower super output area), and ethnicity (categorised as White British, White other, Black, Asian, mixed and other). This fully adjusted model was stratified by sex, and age (16-50 years, over 50 years old). As with the rate, annual adjusted HRs were calculated for all-cause mortality and trends in HRs were assessed using joinpoint regression. In this instance, a change in HR per year (with 95% CIs) was calculated.

The associations between diagnosis and CVD death, CVD diagnosis, suicide and self-harm were assessed using the same multivariable approach, additionally adjusting for area level deprivation, ethnicity and average number of visits to the physician per year of follow-up (to account for the likelihood of having a diagnostic code recorded in the EHR). For each outcome these models were stratified by sex, age (16-50 years, over 50 years old) and calendar period (start of 2000 until end of 2004, start of 2005 until end of 2009, start of 2010 until end of 2014) to examine potential effect modification.
A further model additionally including smoking status (worst ever of: never, ex, current smoker), BMI (worst ever of: healthy weight, overweight, obese), and diagnoses of hypercholesterolemia (defined as total cholesterol ≥5.2mmol/L (Ford et al., 2003)), hypertension (defined as code for hypertension or two consecutive records of systolic blood pressure>140mmHg (National High Blood Pressure Education Program, 2004)) and T2DM during follow-up was tested to see if these covariates explained elevated rates of CVD death and diagnoses in individuals with BPD and schizophrenia relative to the comparison group.

4.4 Results

4.4.1 Clinical and demographic features

17,314 people had a diagnosis of BPD and 22,497 had a diagnosis of schizophrenia with active records between the start of 2000 and the end of 2014. These were matched with 219,387 individuals who never received BPD or schizophrenia diagnoses. There were 1,266 deaths in total in the group with a BPD diagnosis, 2,061 in those with schizophrenia and 6,279 in the comparison group (Table 4.i).

Table 4.i Cohort characteristics

<table>
<thead>
<tr>
<th></th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>219,387</td>
<td>17,341</td>
<td>22,497</td>
</tr>
<tr>
<td>Female, N (%)</td>
<td>102,037 (46.51)</td>
<td>10,202 (58.83)</td>
<td>8,499 (37.78)</td>
</tr>
<tr>
<td>Age at start of f/u years, median (IQR)</td>
<td>41.35 (31.75–54.20)</td>
<td>42.76 (32.59–56.43)</td>
<td>42.51 (32.34–56.51)</td>
</tr>
<tr>
<td>Age at first mention diagnosis years, median (IQR)</td>
<td>- (28.14–50.47)</td>
<td>38.00 (23.12–50.47)</td>
<td>30.00 (23.12–50.47)</td>
</tr>
<tr>
<td>Follow-up years, median (IQR)</td>
<td>2.00 (0.77–4.32)</td>
<td>2.32 (0.93–5.12)</td>
<td>2.47 (0.94–5.53)</td>
</tr>
<tr>
<td>Died, N (%)</td>
<td>6,279 (2.74)</td>
<td>1,266 (7.30)</td>
<td>2,061 (9.16)</td>
</tr>
<tr>
<td>Primary care contacts per year, median (IQR)</td>
<td>7.19 (2.39–18.89)</td>
<td>14.59 (7.86–27.13)</td>
<td>11.36 (5.98–20.84)</td>
</tr>
</tbody>
</table>

Continued overleaf
<table>
<thead>
<tr>
<th>Ethnicity, N (%)</th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td>White British</td>
<td>180,690 (82.36)</td>
<td>15,024 (86.64)</td>
<td>18,648 (82.89)</td>
</tr>
<tr>
<td>White Other</td>
<td>20,792 (9.48)</td>
<td>1,465 (8.45)</td>
<td>1,753 (7.79)</td>
</tr>
<tr>
<td>Black</td>
<td>5,091 (2.32)</td>
<td>235 (1.36)</td>
<td>981 (4.36)</td>
</tr>
<tr>
<td>Asian</td>
<td>8,464 (3.86)</td>
<td>330 (1.90)</td>
<td>635 (2.82)</td>
</tr>
<tr>
<td>Other or mixed</td>
<td>4,350 (1.98)</td>
<td>287 (1.66)</td>
<td>480 (2.13)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social deprivation at baseline, quintiles of UK Townsend score, N (%)</th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [least deprived]</td>
<td>38,404 (17.51)</td>
<td>2,696 (15.55)</td>
<td>2,142 (9.52)</td>
</tr>
<tr>
<td>2</td>
<td>38,862 (17.71)</td>
<td>2,932 (16.91)</td>
<td>2,780 (12.36)</td>
</tr>
<tr>
<td>3</td>
<td>45,679 (20.82)</td>
<td>3,742 (21.58)</td>
<td>4,127 (18.34)</td>
</tr>
<tr>
<td>4</td>
<td>47,633 (21.71)</td>
<td>4,128 (23.80)</td>
<td>5,932 (26.37)</td>
</tr>
<tr>
<td>5 [most deprived]</td>
<td>41,971 (19.13)</td>
<td>3,416 (19.70)</td>
<td>6,695 (29.76)</td>
</tr>
<tr>
<td>missing</td>
<td>6,838 (3.12)</td>
<td>427 (2.46)</td>
<td>821 (3.65)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Record during follow-up of:</th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking, N (%)</td>
<td>55,531 (25.31)</td>
<td>6,503 (37.50)</td>
<td>10,406 (46.26)</td>
</tr>
<tr>
<td>Obesity, N (%)</td>
<td>45,447 (20.72)</td>
<td>5,911 (34.09)</td>
<td>7,801 (34.68)</td>
</tr>
<tr>
<td>Hypercholesterolemia, N (%)</td>
<td>33,867 (15.44)</td>
<td>3,800 (21.91)</td>
<td>14,398 (19.55)</td>
</tr>
<tr>
<td>Hypertension, N (%)</td>
<td>71,194 (32.45)</td>
<td>6,494 (37.45)</td>
<td>8,047 (35.77)</td>
</tr>
<tr>
<td>Diabetes Mellitus, N (%)</td>
<td>14,061 (6.41)</td>
<td>1,851 (10.67)</td>
<td>2,986 (13.27)</td>
</tr>
</tbody>
</table>

### 4.4.2 All-cause mortality

The rate of all-cause mortality in individuals with BPD was 210.34 per 10,000 person years at risk (PYAR) (95% CI 199.07 to 222.25). Trends in BPD mortality rate suggested a reduction over follow-up time. Joinpoint regression fitted a linear model with a significant annual percentage change (APC) in rate of -4.1% (95% CI -5.1 to -3.1) (Figure 4.ii). In individuals with schizophrenia the mortality rate was 248.57 per 10,000 PYAR (95% CI 238.06 to 259.53). Similarly, the rate of mortality in schizophrenia reduced between 2000 and 2014, with joinpoint regression fitting a model with no joinpoints, and an APC of -2.0% (95% CI -3.0 to -0.9). In the comparison population the mortality rate was relatively stable between 2000 and 2003 (APC 4.2; 95%CI –7.6 to 17.6) and then decreased until the end of the study period (APC –8.1; 95%CI –9.6 to –6.5).
Figure 4.ii All-cause mortality rate in bipolar disorder and schizophrenia 2000-2014
Mortality was elevated in those with BPD (HR 1.77; 95% CI 1.67 to 1.88) and schizophrenia (HR 2.08; 95% CI 1.98 to 2.19) relative to the comparison group, after adjustment for age, sex, calendar year, area level deprivation and ethnicity (Table 4.ii). Stratification by sex suggested that men and women with BPD had similarly elevated mortality rates (P=0.297), but in those with schizophrenia, men’s morality rate (HR 2.50; 95% CI 2.32 to 2.69) was more elevated than women’s (HR 1.78; 95% CI 1.66 to 1.90, test for interaction P<0.0001). Mortality rates in those aged 50 and under were more elevated, relative to the comparison group, than those over 50 (test for interaction in both groups p<0.0001). Individuals with BPD aged 50 or under had an adjusted HR of 3.22 (95% CI 2.77 to 3.75), individuals with schizophrenia had an adjusted HR of 4.69 (95% CI 4.16 to 5.29).

Table 4.ii All-cause mortality

<table>
<thead>
<tr>
<th></th>
<th>General population comparison&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-cause mortality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaths, n</td>
<td>6279</td>
<td>1266</td>
<td>2061</td>
</tr>
<tr>
<td>PYAR (10,000s)</td>
<td>64.88</td>
<td>6.02</td>
<td>8.29</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>96.79 (94.42–99.21)</td>
<td>210.34 (199.07–222.25)</td>
<td>248.57 (238.06–259.53)</td>
</tr>
<tr>
<td>Age, sex, calendar period adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.79 (1.69–1.90)</td>
<td>2.14 (2.03–2.25)</td>
</tr>
<tr>
<td>&lt;sup&gt;b&lt;/sup&gt;Sociodemographics adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.77 (1.67–1.88)</td>
<td>2.08 (1.98–2.19)</td>
</tr>
<tr>
<td>Stratified fully&lt;sup&gt;b&lt;/sup&gt;adjusted model HR (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (reference)</td>
<td>1.85 (1.68–2.04)</td>
<td>2.50 (2.32–2.69)</td>
</tr>
<tr>
<td>Women</td>
<td>1 (reference)</td>
<td>1.73 (1.61–1.87)</td>
<td>1.78 (1.66–1.90)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td></td>
<td>P=0.297</td>
<td>P&lt;0.0001</td>
</tr>
<tr>
<td>16-50</td>
<td>1 (reference)</td>
<td>3.22 (2.77–3.75)</td>
<td>4.69 (4.16–5.29)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>1.60 (1.50–1.70)</td>
<td>1.80 (1.71–1.91)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td></td>
<td>P&lt;0.0001</td>
<td>P&lt;0.0001</td>
</tr>
</tbody>
</table>

<sup>a</sup>General population comparison group of up to 6 individuals without BPD or schizophrenia matched for sex, age group, and primary care practice, <sup>b</sup>age, sex, calendar period, area level deprivation, ethnicity

There was evidence of a change in HRs over the study period for groups with both BPD and schizophrenia diagnoses (Figure 4.iii). Joinpoint analysis fitted a model with one joinpoint.
at 2006 for BPD. This model suggested that the HR decreased by 0.08 (95% CI -0.18 to 0.02) per year between 2000 and 2006, and significantly increased by 0.14 (95% CI 0.10 to 0.19) per year from 2006 to 2014 (Figure 4.iii). For schizophrenia, joinpoint regression fitted three linear models (two joinpoints); from 2000 to 2004 the HR reduced significantly by 0.29 (95% CI -0.48 to -0.10) per year, from 2004 to 2010 the HR increased by 0.11 (95% CI 0.04 to 0.17) per year, and from 2010 to 2014 increased by 0.34 (95% CI 0.18 to 0.49) (Figure 4.iii).

Figure 4.iii All-cause mortality adjusted hazard ratio for bipolar disorder and schizophrenia compared to the general population 2000-2014

Continued overleaf
4.4.3 Cardiovascular disease mortality and cardiovascular disease diagnoses

There was evidence of an elevated HR for cardiovascular deaths in schizophrenia (HR 1.39; 95% CI 1.12 to 1.73) after accounting for age, sex, calendar year, area level deprivation, ethnicity and average number of visits to the physician during follow-up (Table 4.iii). Following additional adjustment for smoking, hypercholesterolemia, hypertension, BMI and T2DM, there was no evidence that CVD deaths were elevated in people with schizophrenia relative to the general population (HR 1.22; 95% CI 0.98 to 1.52), suggesting that the excess CVD deaths are explained by increases in traditional CVD risk factors. The HR for BPD was not elevated, relative to the comparison group and there was no evidence that this differed by sex or age (Table 4.iii). Stratification by five year periods suggested that cardiovascular deaths were elevated in people with BPD, relative to the general population after 2010 (HR 1.92; 95% CI 1.24 to 2.98), but not before this. Amongst those individuals with
schizophrenia, sex was not an effect modifier (P=0.068). The HR for CVD mortality in people with schizophrenia aged 50 or under was 3.20 (95% CI 1.62 to 6.31), whereas in those over 50 it was 1.29 (95% CI 1.03 to 1.63, test for interaction P=0.013) (Table 4.iii).

Rates of new CVD were elevated in both BPD and schizophrenia groups compared to the comparison group after adjustment for age, sex, calendar year, area level deprivation, ethnicity and average number of visits to the physician during follow-up (HR 1.41; 95% CI 1.26 to 1.58, and HR 1.36; 95% CI 1.24 to 1.50 respectively) (Table 4.iii). Records of smoking, hypercholesterolemia, hypertension, BMI, and T2DM in the patient notes did not explain the increased rates, though it did attenuate them (HR 1.26; 95% CI 1.12 to 1.41 in BPD, and HR 1.22; 95% CI 1.11 to 1.35 in schizophrenia after adjustment for these confounders).

Men were significantly more likely to receive a CVD diagnosis than women in both BPD and schizophrenia groups (HR 1.55; 95% CI 1.31 to 1.82, and HR 1.37; 95% CI 1.20 to 1.57 respectively). Women with either SMI diagnosis did not have elevated rates relative to women in the general population (Table 4.iii). Increased rates of CVD mortality in those with schizophrenia aged 16-50 were not reflected in equally increased rates of CVD diagnosis (HR 1.66; 95% CI 1.29-2.15).
<table>
<thead>
<tr>
<th></th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiovascular disease death</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>437</td>
<td>59</td>
<td>106</td>
</tr>
<tr>
<td>PYAR (10,000s)</td>
<td>64.88</td>
<td>6.02</td>
<td>8.29</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>6.74 (6.13–7.40)</td>
<td>9.80 (7.59–12.65)</td>
<td>12.78 (10.57–15.46)</td>
</tr>
<tr>
<td>Age, sex, calendar period adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.11 (0.85–1.46)</td>
<td>1.43 (1.15–1.76)</td>
</tr>
<tr>
<td>Sociodemographics adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.10 (0.84–1.46)</td>
<td>1.39 (1.12–1.73)</td>
</tr>
<tr>
<td><strong>Stratified a adjusted model HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (reference)</td>
<td>1.28 (0.83–1.97)</td>
<td>1.74 (1.26–2.39)</td>
</tr>
<tr>
<td>Women</td>
<td>1 (reference)</td>
<td>1.00 (0.70–1.43)</td>
<td>1.16 (0.86–1.57)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.394</td>
<td>P=0.068</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>1.13 (0.34–3.71)</td>
<td>3.20 (1.62–6.31)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.960</td>
<td>P=0.013</td>
<td></td>
</tr>
<tr>
<td>16-50</td>
<td>1 (reference)</td>
<td>1.10 (0.83–1.45)</td>
<td>1.29 (1.03–1.63)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>1.06 (0.64–1.77)</td>
<td>1.61 (1.12–2.30)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.009</td>
<td>P=0.323</td>
<td></td>
</tr>
<tr>
<td>2000-2004</td>
<td>1 (reference)</td>
<td>1.06 (0.42–1.14)</td>
<td>1.13 (0.80–1.60)</td>
</tr>
<tr>
<td>2005-2009</td>
<td>1 (reference)</td>
<td>1.92 (1.24–2.98)</td>
<td>1.57 (1.01–2.43)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.009</td>
<td>P=0.323</td>
<td></td>
</tr>
<tr>
<td>2010-2014</td>
<td>1 (reference)</td>
<td>1.10 (0.34–3.71)</td>
<td>3.20 (1.62–6.31)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.960</td>
<td>P=0.013</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>1.13 (0.84–1.46)</td>
<td>1.39 (1.12–1.73)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.009</td>
<td>P=0.323</td>
<td></td>
</tr>
<tr>
<td><strong>Cardiovascular disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>2445</td>
<td>382</td>
<td>551</td>
</tr>
<tr>
<td>PYAR (10,000s)</td>
<td>61.59</td>
<td>5.60</td>
<td>7.76</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>39.70 (38.15–41.30)</td>
<td>68.19 (61.69–75.39)</td>
<td>70.97 (65.28–77.15)</td>
</tr>
<tr>
<td>Age, sex, calendar period adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.46 (1.31–1.63)</td>
<td>1.47 (1.34–1.61)</td>
</tr>
<tr>
<td>Sociodemographics adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.41 (1.26–1.58)</td>
<td>1.36 (1.24–1.50)</td>
</tr>
<tr>
<td><strong>Stratified b adjusted model HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (reference)</td>
<td>1.55 (1.31–1.82)</td>
<td>1.37 (1.20–1.57)</td>
</tr>
<tr>
<td>Women</td>
<td>1 (reference)</td>
<td>1.00 (0.86–1.17)</td>
<td>1.04 (0.90–1.20)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P&lt;0.0001</td>
<td>P=0.018</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>2.30 (1.78–2.99)</td>
<td>1.66 (1.29–2.15)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.670</td>
<td>P=0.002</td>
<td></td>
</tr>
<tr>
<td>2000-2004</td>
<td>1 (reference)</td>
<td>1.22 (0.96–1.56)</td>
<td>1.45 (1.12–1.69)</td>
</tr>
<tr>
<td>2005-2009</td>
<td>1 (reference)</td>
<td>1.26 (1.03–1.54)</td>
<td>1.46 (1.24–1.71)</td>
</tr>
<tr>
<td>2010-2014</td>
<td>1 (reference)</td>
<td>1.57 (1.34–1.83)</td>
<td>1.35 (1.16–1.57)</td>
</tr>
<tr>
<td>Test for interaction</td>
<td>P=0.118</td>
<td>P=0.762</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>1.26 (1.12–1.41)</td>
<td>1.22 (1.11–1.35)</td>
</tr>
</tbody>
</table>

1General population comparison group of up to 6 individuals without BPD or schizophrenia matched for sex, age group, primary care practice, bAdjusted for age, sex, calendar period, area level deprivation, ethnicity, number of primary care contacts, cAdjusted for age, sex, calendar period, area level deprivation, ethnicity, smoking, high cholesterol, high blood pressure, BMI, diabetes mellitus and primary care contacts.
4.4.4 Suicide and self-harm

After accounting for age, sex, calendar year, area level deprivation, ethnicity and average number of visits to the physician per year of follow-up, the rate of suicide in those with BPD was 12.66 (95% CI 7.79 to 20.58) times that of the comparison group (Table 4.iv). The similarly adjusted HR in the group with schizophrenia was 7.21 (95% CI 4.26 to 12.19). Increased suicide rates were observed in both BPD and schizophrenia irrespective of sex, age or calendar period. However, whilst there were some differences in the point estimates, confidence intervals were wide and there was no evidence of significant differences by sex, age group or time period (Table 4.iv).

Self-harm rates were elevated in both BPD (HR 25.24; 95% CI 23.63 to 29.96) and schizophrenia (HR 22.14; 95% CI 19.58 to 25.03) after adjusting for sociodemographic characteristics and physician visits (Table 4.iv). Rates of self-harm did not differ by sex in either BPD (P=0.096) or schizophrenia (P=0.735). Self-harm was dramatically elevated in those aged 50 or under with a diagnosis of BPD (HR 55.74; 95% CI 45.35 to 68.52) and schizophrenia (HR 52.07; 95% CI 42.43 to 63.92) and still increased, but to a lesser extent in those aged over 50 years old (Table 4.iv).
Table 4.iv Suicide and self-harm

<table>
<thead>
<tr>
<th></th>
<th>General population comparison</th>
<th>Bipolar disorder</th>
<th>Schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suicide</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>33</td>
<td>36</td>
<td>33</td>
</tr>
<tr>
<td>PYAR (10,000s)</td>
<td>64.88</td>
<td>6.02</td>
<td>8.29</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>0.51 (0.36–0.72)</td>
<td>5.98 (4.31–8.29)</td>
<td>3.98 (2.83–5.60)</td>
</tr>
<tr>
<td>Age, sex, calendar period adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>12.94 (8.04–20.82)</td>
<td>7.90 (4.84–12.90)</td>
</tr>
<tr>
<td><strong>Sociodemographics adjusted HR (95% CI)</strong></td>
<td>1 (reference)</td>
<td>12.66 (7.79–20.58)</td>
<td>7.21 (4.26–12.19)</td>
</tr>
<tr>
<td><strong>Stratified fully adjusted model HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (reference)</td>
<td>11.10 (5.85–21.06)</td>
<td>6.91 (3.71–12.87)</td>
</tr>
<tr>
<td>Women</td>
<td>1 (reference)</td>
<td>15.27 (7.11–32.78)</td>
<td>7.90 (3.15–19.79)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P=0.524</td>
<td>P=0.810</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>8.79 (3.30–23.41)</td>
<td>5.47 (2.00–14.90)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P=0.478</td>
<td>P=0.499</td>
<td></td>
</tr>
<tr>
<td>2005-2009</td>
<td>1 (reference)</td>
<td>9.50 (4.74–19.07)</td>
<td>6.26 (3.05–12.88)</td>
</tr>
<tr>
<td>2010-2014</td>
<td>1 (reference)</td>
<td>18.88 (8.62–41.32)</td>
<td>7.50 (2.94–19.12)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P=0.306</td>
<td>P=0.772</td>
<td></td>
</tr>
<tr>
<td><strong>Self-harm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>390</td>
<td>950</td>
<td>1101</td>
</tr>
<tr>
<td>PYAR (10,000s)</td>
<td>64.87</td>
<td>5.99</td>
<td>8.26</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>6.01 (5.44–6.64)</td>
<td>158.66 (148.88–169.07)</td>
<td>133.33 (125.68–141.44)</td>
</tr>
<tr>
<td>Age, sex, calendar period adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>26.61 (23.63–29.96)</td>
<td>24.04 (21.37–27.07)</td>
</tr>
<tr>
<td><strong>Sociodemographics adjusted HR (95% CI)</strong></td>
<td>1 (reference)</td>
<td>25.24 (22.37–28.49)</td>
<td>22.14 (19.58–25.03)</td>
</tr>
<tr>
<td><strong>Stratified fully adjusted model HR (95% CI)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1 (reference)</td>
<td>22.21 (18.48–26.68)</td>
<td>22.32 (18.91–26.34)</td>
</tr>
<tr>
<td>Women</td>
<td>1 (reference)</td>
<td>27.31 (23.21–32.13)</td>
<td>21.43 (17.97–25.54)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P=0.096</td>
<td>P=0.735</td>
<td></td>
</tr>
<tr>
<td>16-50</td>
<td>1 (reference)</td>
<td>55.74 (45.35–68.52)</td>
<td>52.07 (42.43–63.92)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (reference)</td>
<td>13.99 (11.98–16.34)</td>
<td>9.99 (8.53–11.72)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P&lt;0.0001</td>
<td>P&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>2010-2014</td>
<td>1 (reference)</td>
<td>26.69 (22.48–31.69)</td>
<td>24.68 (20.79–29.31)</td>
</tr>
<tr>
<td><strong>Test for interaction</strong></td>
<td>P=0.424</td>
<td>P=0.030</td>
<td></td>
</tr>
</tbody>
</table>

*General population comparison group of up to 6 individuals without BPD or schizophrenia matched for sex, age group, primary care practice. *Adjusted for age, sex, calendar period, area level deprivation, ethnicity, number of primary care contacts
4.5 Discussion

4.5.1 Main findings
In this cohort of over 17,000 people with BPD and over 22,000 people with schizophrenia, I found decreasing rates of all-cause mortality for both diagnoses. Despite this, due to an even steeper decline in the mortality rate in the general population, the rate of death relative to a matched general population comparison group increased from the mid-2000s. This suggests that the improvement in health in the general population is increasing more rapidly than in those with SMI, and health inequalities are growing.

4.5.2 Comparison with existing literature
After accounting for sociodemographic characteristics, over the fifteen-year follow-up period, the rate of all-cause mortality in those with BPD was 1.77 times that of the general population (95% CI 1.67 to 1.88) and in those with schizophrenia was 2.08 times greater (1.98 to 2.19). These results are slightly lower than those from meta-analyses of the existing literature (Chapter 3 and (Saha et al., 2007)), but are consistent with population based samples (Chapter 3). Deaths in those aged 50 and under were markedly elevated for both BPD and schizophrenia. In individuals with schizophrenia, the risk of dying before 50 of CVD is strikingly elevated and CVD is infrequently diagnosed in advance of the terminal event. This is in-line with the work of Osborn and colleagues that used a cohort ending in 2002 (Osborn et al., 2007). Increased rates of CVD death in those with schizophrenia were explained by traditional risk factors (smoking, BMI, hypercholesterolemia, hypertension and T2DM). Suicide was rare in the cohort and as such, I could not state which of the SMI diagnoses has more elevated risk, relative to the general population.
4.5.3 Potential explanations for these findings

During the study period, a number of factors could have differentially influenced mortality in people with BPD and schizophrenia compared to the general population. There is evidence that addressing negative health behaviours has been more effective in the general population, for example, population level smoking cessation programmes have had less impact on people with SMI (Lawrence and Kisely, 2010). Whilst SGAs have been shown to reduce mortality overall (Tiihonen et al., 2009), polypharmacy and higher drug dosages may increase it (Weinmann et al., 2009). Polypharmacy is increasingly common in BPD (Hayes et al., 2011) and may be contributing to the worsening CVD mortality compared to the general population in the 2010-2014 period. Wahlbeck and colleagues speculated that the reducing mortality gap seen in Nordic countries up until 2006 reflected the success of deinstitutionalisation (Wahlbeck et al., 2011). Whilst deinstitutionalisation in the UK has been a success in terms of integrating people into wider society, it has been argued that there is now too little support for people living with BPD and schizophrenia in the community (Fakhoury and Priebe, 2007, Green and Griffiths, 2014) and this may be reflected in mortality rates. Research into the health effects of recession has suggested that consequences will be most severe for the poorest groups in society and will impact most where social safety-nets are lacking and public hardship grows rapidly (Cooper, 2011, Riumallo-Herl et al., 2014). Given this, I could hypothesise that policies made in the UK following the 2008 financial crash (i.e., austerity) have impacted hardest on those with SMI. A comprehensive understanding of the mechanisms by which SMI shortens life remains elusive, but potentially constitutes a syndemic including psychiatric and physical comorbidity, substance misuse, clustering of adverse social factors and lifestyle behaviours (Liu et al., 2017).
4.5.4 Strengths and limitations

A major strength of this study, beyond its size and length of follow-up, is that the results are generalisable to individuals living with BPD and schizophrenia in the UK (whether or not they have been inpatients in psychiatric hospitals), because of the representativeness of the THIN database. The exposure groups (BPD and schizophrenia) (Hardoon et al., 2013, Nazareth et al., 1993) and the outcomes (mortality, CVD deaths, CVD, suicide and self-harm) (Arana et al., 2010, Hammad et al., 2008, Ruigómez et al., 2010) have been well validated. Despite this, there is potential under-recording of cause specific deaths and morbidity, and potential for misclassification. Cause of death from death certificates would have improved the study, but this information was not available. In particular, suicide rates were lower than would be expected from ONS data (Office for National Statistics, 2016) and generally suicide deaths and self-harm events may be under recognised in the general population using EHRs (Thomas et al., 2013a). Additionally, suicide was a rare outcome and therefore the study was potentially underpowered to investigate differences by sex, age and calendar period. However, I would not expect these to be differential by diagnostic group and my HRs reflect recent standardised mortality ratio estimates from the UK (Brown et al., 2010, Hoang et al., 2013).

There is no evidence that CVD death recording would be differential by diagnostic group (Denaxas et al., 2012), and the proportion of deaths from CVD in the general population comparison group (7%) is consistent with 2014 ONS data, where in a similar age range, CVD mortality represented 6% of all deaths (Office for National Statistics, 2016). There should be minimal under-recording of all-cause mortality (Maguire et al., 2009).

I included only those individuals with a diagnosis code of schizophrenia, and excluded those with other non-affective psychosis codes. This was because previous literature on mortality
in people with SMI has tended to focus on schizophrenia, and non-affective psychosis is a highly heterogeneous group. Those receiving a schizophrenia diagnosis are likely to be at the more severe end of the psychosis spectrum – therefore it is an important finding that mortality rates are similar in both BPD and schizophrenia.

There was a small amount of missing data for both ethnicity and SES. As discussed in Section 2.9.5, ethnicity was dealt with via agreed methods. Individuals with no measure of social deprivation were dropped from the analysis, there is good evidence that variables with less than 5% missing data will not generate bias (Bennett, 2001).

The measure of social deprivation included in the multivariable model was one related to the area in which the individual lived, rather than their individual SES. As such, this may not be sensitive enough to capture the impact of socioeconomic status on the association between SMI and mortality. However, my study goes further than others in this area, which use age and sex adjusted SMRs, by adjusting for a number of other recognised confounders. There are potential confounders that I did not include in the model that aimed to explain the elevated CVD rates in SMI, for example alcohol use and other lifestyle factors. Therefore, there remains the potential for unmeasured confounding. However, the covariates included are those used in CVD risk prediction models and are likely to partially mediate the relationship between SMI and CVD (rather than being true confounders). Recording of smoking status, BMI, hypercholesterolemia, hypertension and T2DM may be incomplete and therefore there is potential for residual confounding. However, recording of a number of these CVD risk factors has been incentivised by the QOF (Doran et al., 2011), and I have attempted to minimise missingness by defining these as ever recorded during follow-up. Although this incentivisation potentially improved CVD recording after 2004,
there is no evidence that this was differential with regards to BPD, schizophrenia or the comparison group. Limitations are discussed in more detail in Section 9.5.

4.5.5 Conclusions
Mortality trends in individuals with severe mental illness are important indicators of outcome and quality of psychiatric and medical care (Bushe et al., 2010, Ösby et al., 2000). This study suggests that despite important reductions in over-all mortality since 2000, interventions to improve health outcomes for those with BPD or schizophrenia have not reduced the mortality gap.

4.5.6 Implications of the findings from this study
My results underscore how continuous monitoring of mortality and morbidity in people with BPD and schizophrenia might guide us in evaluating the impact of interventions to manage physical comorbidity, reduce inequalities in medical care provision and prevent inequalities in their background risk factors. In Study 3 (Chapter 7) I continue to explore physical outcomes (such as CVD) in BPD when I assess how commonly used maintenance treatments may influence rates of adverse effects, and in Study 4 (Chapter 8) where I examine the relative rates of suicide in people prescribed these medications. Implications for clinicians, patients, policy makers and for further research are discussed in Chapter 9.
Chapter 5  Systematic review and network meta-analysis comparing the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder

5.1 Summary

5.1.1 Objective
To summarise relative efficacy of commonly used maintenance mood stabiliser medications (lithium, valproate, olanzapine and quetiapine) via a NMA of all head-to-head and placebo controlled RCTs.

5.1.2 Method
The Cochrane Central Register of Controlled Trials was searched to identify trials of treatments for BPD (lithium, valproate, olanzapine or quetiapine), which involved head-to-head comparisons or were placebo controlled, lasted for six months or longer, and included any measure of effectiveness to prevent any mood episode and/or discontinuation. Effectiveness and tolerability were assessed using a random-effects NMA within a Bayesian framework.

5.1.3 Results
I screened 382 trials, and 18 fulfilled inclusion criteria. All active medications were significantly more effective than placebo at preventing any mood episode and had
significantly lower all-cause discontinuation. It was not possible to distinguish between lithium, valproate, olanzapine or quetiapine statistically in terms of efficacy or tolerability.

5.1.4 Conclusion

All four medications examined are superior to placebo, but NMA does not clearly identify one treatment as superior. There is better quality evidence to support the use of lithium, notwithstanding its tolerability profile.
5.2 Introduction

Recent meta-analysis (Severus et al., 2014), NMA (Miura et al., 2014) and guidelines updates (Goodwin et al., 2016, National Institute for Health and Care Excellence, 2014) concluded that lithium is the most appropriate first-line treatment for maintenance of euthymia in BPD. However, there are multiple limitations in evaluating the trials that contributed to this conclusion, and it is unclear how this recommendation should be applied clinically. Across the RCTs included in Miura et al. and Severus et al. there were a number of conceptual design differences. Inclusion criteria varied with trials including pure bipolar I or bipolar II samples, or combinations of both. It is likely that the diagnosis of BPD has been applied differently over time (trials form the early 1970s are included) and by Country (Geddes et al., 2004). Therefore, a group (at both study-level and individual-level) that appears homogenous may in reality include a range of illness severities. Further complicating this is the fact that many patients present with a similar polarity of illness during each relapse, therefore recruitment during a manic or depressive phase may obscure effectiveness of a drug for the other polarity.

In traditional pairwise meta-analyses, comparing treatments is only appropriate when trials are similar in terms of methodological and clinical characteristics. The same holds true for indirect comparisons in NMA. This extension of homogeneity to indirect comparisons in NMA is known as transitivity (Cipriani et al., 2013b). As such, it is only appropriate to perform NMA on studies that are clinically and methodologically transitive; this should include factors such as inclusion criteria, illness severity, and stage of illness. The plausibility of the transitivity assumption requires judgment to decide whether differences in the distributions of the effect modifiers across studies are large enough to make NMA invalid (Chaimani et al., 2013). Consistency is the statistical expression of transitivity. It is assessed
by statistically comparing direct and indirect summary effects in specific loops (Bucher et al., 1997). Until recently, NMA was limited by a need for advanced statistical software and computational knowledge. However, NMA is now possible through a suite of Stata commands (mvmeta) (StataCorp, 2013) and a number of authors have provided guidance on its methodology (Chaimani et al., 2013, Cipriani et al., 2013b, Salanti et al., 2011, Salanti et al., 2014).

Miura et al. completed an NWA of all maintenance treatments for BPD (Miura et al., 2014). Interventions included monotherapies: aripiprazole; carbemazapine; fluoxetine; imipramine; lithium; lamotrigine; olanzapine; oxcabazepine; paliperidone; quetiapine; risperidone injection and valproate, and combination therapies: lithium and oxcarbazepine; lithium and imipramine; lithium and valproate; valproate and aripiprazole; aripiprazole and lamotrigine; valproate and lamotrigine. Trials were parallel design with active or placebo comparator groups. Using the GRADE framework (Guyatt et al., 2011) the authors show that the quality of evidence (certainty of point estimate) was low or very low for all comparisons apart from lithium versus placebo and olanzapine versus placebo (which were both moderate). Transitivity is also a problem in this NMA. For example, a number of treatments (such as drug combinations or those given via injection) are systematically different from other comparator drugs. There are also fundamental differences in terms of illness severity, diagnostic criteria, illness polarity, enriched design, pragmatic design, outcome definition and duration.

In some trials, participants were recruited and randomised to interventions whilst they were euthymic. In others, participants were recruited during an acute episode and prescribed a particular intervention drug, then responders to this drug would be randomised to continue this treatment or switch to a placebo or active comparator.
(enriched design). This method tends to favour the investigational drug, and is more commonly found in pharmaceutical company sponsored trials (Goodwin et al., 2016). Of the trials included in Miura et al. 58% were enriched. In an attempt to overcome this limitation the authors completed a sensitivity analysis reducing the weight of these studies by 50%. This did not affect the summary estimates or conclusions, but may not have captured the full extent of the advantage given to the enriched design drugs.

Another limitation when combining trial results is the range of outcome definitions chosen to represent relapse or reoccurrence. Whilst it is clear that a reoccurrence requiring hospitalisation is a clear failure in treatment, this was rare as a primary outcome. In addition, over time as services change, there is unlikely to be a consistent threshold for hospitalisation. Relapse as measured by rating scale, or addition of new medications (especially if these are short-term prescriptions of low dose antipsychotics or benzodiazepines) may not be considered treatment failures by clinicians or, more importantly by patients themselves. Often a composite measure of all of these outcomes has been used in a bid to increase outcome event frequency and power. The range of outcomes measured and the limited number of trials means that a sensitivity analysis by specific outcome would not be possible.

The most commonly prescribed maintenance medications in the UK are lithium, valproate and olanzapine – having been recommended in the NICE guidelines from 2006 as first-line treatments (National Institute for Health and Care Excellence, 2006), and quetiapine – because of its effectiveness in bipolar depression (Calabrese et al., 2005a). It is unclear if the evidence presented in these reviews (Goodwin et al., 2016, Miura et al., 2014, National Institute for Health and Care Excellence, 2014, Severus et al., 2014) is strong enough to change prescribing practice in terms of new prescribing for patients with BPD, let alone
consider a change of prescribing in someone in which maintenance treatment is already established. Given the current clinical use of these drugs, this review and subsequent NMA aims to examine the existing evidence for preferable prescribing of one of these drugs over the other three. To do this I have examined trials that compare two of these drugs or trials of one of these drugs against a placebo control. Given the treatment aims of maintenance mood stabiliser medication (i.e., to protect against either polarity of illness) this review includes trials that used any measure of relapse or reoccurrence, with any pole (manic or depressive) as an outcome. It also examines discontinuation for any reason, as this measure most accurately reflects real world medication adherence and usage.

5.3 Methods

I systematically searched randomised controlled parallel group trials of one of the study drugs versus another as an active comparator or versus placebo.

5.3.1 Identification of studies

The Cochrane Central Register of Controlled Trials was searched to the end of June 2015. This database includes relevant studies from MEDLINE, PsycINFO and Embase. The search was completed using terms relating to BPD and maintenance treatment and lithium or valproate or olanzapine (see Appendix 3.2 for full search terms). Following this, reference lists of all identified RCTs and other relevant papers were checked for missing trials. The abstracts of potentially relevant articles were reviewed. Study-level information was filtered to identify trials reported in more than one location.

5.3.2 Inclusion and exclusion criteria

Included trials met all of the following a priori defined criteria:

i) Published between 1 January 1960 and 1 June 2015
ii) Reported trials of lithium, valproate or olanzapine versus placebo or any of the study drugs versus each other

iii) Individuals included in the trial were 16 years or older

iv) Trial follow-up of at least 6 months

v) Included any measure of effectiveness and/or discontinuation

Studies were excluded if:

i) Patients received adjunctive treatment as part of the intervention

ii) An intention-to-treat analysis could not be completed from data in the trial manuscript or additional correspondence

5.3.3 Data extraction

Once a trial was included, data were extracted and entered into a database that included trial-level variables: authors, year of publication, decade in which trial was carried out, length of follow-up, intervention and comparator drug, number of participants in each arm, diagnosis, blinding status, if the trial had an enriched design for the drug of interest (i.e., patients were assigned to this drug prior to the start of the trial), reporting of previous maintenance treatment and outcome measure. Individual-level variables were also extracted: number of individuals with relapse or reoccurrence in each arm and number of individuals who discontinued in each arm.

5.3.4 Statistical analysis

A NMA was completed using *mvmeta* commands in Stata (StataCorp, 2013). NMA synthesises data from a network of trials about more than two competing interventions. The integration of direct evidence (from studies directly comparing interventions) with indirect evidence (information about two treatments derived via a common comparator)
increases the precision in the estimates and produces a relative ranking of all treatments for the studied outcome.

5.4 Results

After duplicate removal, 105 full manuscripts were reviewed from a possible 382 trials (Figure 5.i). Eighteen trials met the inclusion criteria: three compared lithium with valproate (Bowden et al., 2000, Calabrese et al., 2005b, Geddes et al., 2010), one compared lithium with olanzapine (Calabrese et al., 2005b), one compared lithium and quetiapine (Weisler et al., 2011), one trial directly compared valproate and olanzapine for discontinuation only (Tohen et al., 2003); eleven trials compared lithium with placebo (Amsterdam and Shults, 2010, Bowden et al., 2003, Bowden et al., 2000, Calabrese et al., 2003, Cundall et al., 1972, Dunner et al., 1976, Kane et al., 1982, Melia, 1970, Prien et al., 1973a, Stallone et al., 1973, Weisler et al., 2011) (one of these had no discontinuation data (Dunner et al., 1976)), one compared valproate with placebo (Bowden et al., 2000) two olanzapine with placebo (Tohen et al., 2006) and two quetiapine and placebo (Weisler et al., 2011, Young et al., 2014). In total 4,515 individuals were included across all trials. Characteristics of included trials are shown in Table 5.i.
The mean duration of follow-up in included trials was 74.6 weeks (SD 27.3). There was considerable variation in mood state at recruitment and in treatments used to stabilise mood episodes prior to randomisation. Five trials were enriched for one of the study drugs (i.e., patients were selected who had responded acutely to that drug) (Table 5.i).
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Intervention</th>
<th>Control</th>
<th>Duration (weeks)</th>
<th>N intervention</th>
<th>N control</th>
<th>Effectiveness outcome</th>
<th>Diagnosis</th>
<th>Recruit mood</th>
<th>Blinding</th>
<th>enriched</th>
<th>Drug before</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melia et al.</td>
<td>1970</td>
<td>lithium</td>
<td>placebo</td>
<td>104</td>
<td>5</td>
<td>6</td>
<td>hospitalisation</td>
<td>BPD</td>
<td>euthymic</td>
<td>double</td>
<td>no</td>
<td>lithium</td>
</tr>
<tr>
<td>Cundall et al.</td>
<td>1972</td>
<td>lithium</td>
<td>placebo</td>
<td>26</td>
<td>8</td>
<td>5</td>
<td>combination</td>
<td>BPD</td>
<td>unknown</td>
<td>double</td>
<td>yes</td>
<td>lithium</td>
</tr>
<tr>
<td>Prien et al.</td>
<td>1973</td>
<td>lithium</td>
<td>placebo</td>
<td>104</td>
<td>101</td>
<td>104</td>
<td>combination</td>
<td>BPD</td>
<td>manic</td>
<td>double</td>
<td>yes</td>
<td>lithium</td>
</tr>
<tr>
<td>Stallone et al.</td>
<td>1973</td>
<td>lithium</td>
<td>placebo</td>
<td>121</td>
<td>25</td>
<td>27</td>
<td>combination</td>
<td>BPD</td>
<td>euthymic</td>
<td>double</td>
<td>no</td>
<td>unknown</td>
</tr>
<tr>
<td>Dunner et al.</td>
<td>1976</td>
<td>lithium</td>
<td>placebo</td>
<td>69</td>
<td>16</td>
<td>24</td>
<td>supplementary drugs</td>
<td>BPD</td>
<td>euthymic</td>
<td>double</td>
<td>no</td>
<td>unknown</td>
</tr>
<tr>
<td>Kane et al.</td>
<td>1982</td>
<td>lithium</td>
<td>placebo</td>
<td>104</td>
<td>4</td>
<td>7</td>
<td>clinical relapse</td>
<td>BP-2</td>
<td>euthymic</td>
<td>double</td>
<td>no</td>
<td>another mood stabiliser</td>
</tr>
<tr>
<td>Bowden et al.</td>
<td>2000</td>
<td>lithium</td>
<td>valproate</td>
<td>52</td>
<td>91</td>
<td>187</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>either of study drugs</td>
</tr>
<tr>
<td>Bowden et al.</td>
<td>2000</td>
<td>valproate</td>
<td>placebo</td>
<td>52</td>
<td>187</td>
<td>94</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>either of study drugs</td>
</tr>
<tr>
<td>Bowden et al.</td>
<td>2000</td>
<td>lithium</td>
<td>placebo</td>
<td>52</td>
<td>91</td>
<td>94</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>either of study drugs</td>
</tr>
<tr>
<td>Bowden et al.</td>
<td>2003</td>
<td>lithium</td>
<td>placebo</td>
<td>76</td>
<td>46</td>
<td>70</td>
<td>supplementary drugs</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>another mood stabiliser</td>
</tr>
<tr>
<td>Calabrese et al.</td>
<td>2003</td>
<td>lithium</td>
<td>placebo</td>
<td>76</td>
<td>121</td>
<td>121</td>
<td>combination</td>
<td>BP-1</td>
<td>depressive</td>
<td>double</td>
<td>no</td>
<td>another mood stabiliser</td>
</tr>
<tr>
<td>Tohen et al.</td>
<td>2003</td>
<td>valproate</td>
<td>olanzapine</td>
<td>47</td>
<td>126</td>
<td>125</td>
<td>N/A</td>
<td>BPD</td>
<td>manic</td>
<td>double</td>
<td>no</td>
<td>unknown</td>
</tr>
<tr>
<td>Calabrese et al.</td>
<td>2005</td>
<td>lithium</td>
<td>valproate</td>
<td>80</td>
<td>32</td>
<td>28</td>
<td>rating scale relapse</td>
<td>BPD</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>combination of study drugs</td>
</tr>
<tr>
<td>Tohen et al.</td>
<td>2005</td>
<td>lithium</td>
<td>olanzapine</td>
<td>48</td>
<td>214</td>
<td>217</td>
<td>rating scale relapse</td>
<td>BP-1</td>
<td>manic</td>
<td>double</td>
<td>no</td>
<td>combination of study drugs</td>
</tr>
</tbody>
</table>

*Continued overleaf*
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Intervention</th>
<th>control</th>
<th>Duration (weeks)</th>
<th>N intervention</th>
<th>N control</th>
<th>Effectiveness outcome</th>
<th>Diagnosis</th>
<th>Recruit mood</th>
<th>Blinding</th>
<th>enriched</th>
<th>Drug before</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tohen et al.</td>
<td>2006</td>
<td>olanzapine</td>
<td>placebo</td>
<td>48</td>
<td>225</td>
<td>136</td>
<td>combination</td>
<td>BP-1</td>
<td>manic</td>
<td>double</td>
<td>yes</td>
<td>olanzapine</td>
</tr>
<tr>
<td>Geddes et al.</td>
<td>2010</td>
<td>lithium</td>
<td>valproate</td>
<td>104</td>
<td>110</td>
<td>110</td>
<td>combination</td>
<td>BP-1</td>
<td>euthymic</td>
<td>open</td>
<td>no</td>
<td>combination of study drugs</td>
</tr>
<tr>
<td>Amsterdam et al.</td>
<td>2010</td>
<td>lithium</td>
<td>placebo</td>
<td>50</td>
<td>26</td>
<td>27</td>
<td>combination</td>
<td>BP-2</td>
<td>depressive</td>
<td>double</td>
<td>no</td>
<td>antidepressant</td>
</tr>
<tr>
<td>Weisler et al.</td>
<td>2011</td>
<td>lithium</td>
<td>placebo</td>
<td>104</td>
<td>364</td>
<td>404</td>
<td>combination</td>
<td>BP-2</td>
<td>any</td>
<td>double</td>
<td>yes</td>
<td>quetiapine</td>
</tr>
<tr>
<td>Weisler et al.</td>
<td>2011</td>
<td>lithium</td>
<td>quetiapine</td>
<td>104</td>
<td>364</td>
<td>404</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>yes</td>
<td>quetiapine</td>
</tr>
<tr>
<td>Weisler et al.</td>
<td>2011</td>
<td>quetiapine</td>
<td>placebo</td>
<td>104</td>
<td>404</td>
<td>404</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>yes</td>
<td>quetiapine</td>
</tr>
<tr>
<td>Vieta et al.</td>
<td>2012</td>
<td>olanzapine</td>
<td>placebo</td>
<td>78</td>
<td>131</td>
<td>135</td>
<td>combination</td>
<td>BP-1</td>
<td>any</td>
<td>double</td>
<td>no</td>
<td>another antipsychotic</td>
</tr>
<tr>
<td>Young et al.</td>
<td>2014</td>
<td>quetiapine</td>
<td>placebo</td>
<td>52</td>
<td>291</td>
<td>294</td>
<td>combination</td>
<td>BPD</td>
<td>depressive</td>
<td>double</td>
<td>yes</td>
<td>quetiapine</td>
</tr>
</tbody>
</table>

N/A, not available; BPD, any bipolar disorder diagnosis; Bowden et al. (2000) and Weisler et al. (2011) are 3 arm trials
Figure 5.ii shows the network of comparisons in the NMA for relapse and reoccurrence and Figure 5.iii shows the network for discontinuation. Each node represents a drug included in the analysis, with the node size proportional to the number of individuals who were assigned to that drug treatment. Each line represents direct comparisons between drug treatments. The width of the line is proportional to the number of trials in that comparison. All treatments were compared with at least two other treatments.

Figure 5.ii Network plot of included trials: Relapse and reoccurrence

Figure 5.iii Network plot of included trials: Discontinuation
For any mood episode relapse or reoccurrence each of the study drugs was better than placebo (Figure 5.iv), this was also true for all-cause discontinuation (Figure 5.vi). There was no evidence of a difference between direct and indirect estimates (test of inconsistency was Chi$^2$=1.45, df=5, P=0.918 for relapse/reoccurrence and Chi2=2.05, df=6, P=0.915 for all-cause discontinuation).

None of the active drugs was superior to other study drugs for either outcome. Relative risk estimates for all comparisons are shown in Table 5.ii.

Table 5.ii Relative risk of relapse or reoccurrence (orange) and all-cause discontinuation (green) according to NMA

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RR</strong></td>
<td>0.63 (0.53–0.74)</td>
<td>0.94 (0.71–1.24)</td>
<td>0.86 (0.58–1.27)</td>
<td>0.92 (0.66–1.28)</td>
<td>1.05 (0.71–1.55)</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>0.53–0.74</td>
<td>0.71–1.24</td>
<td>0.58–1.27</td>
<td>0.66–1.28</td>
<td>0.71–1.55</td>
</tr>
</tbody>
</table>

Comparisons between treatments should be read from left to right and the estimate is in the cell in common between the column-defining treatment and the row-defining treatment. For relapse/reoccurrence, risk ratios (RRs) below 1 favour the column-defining treatment. For discontinuation, RRs higher than 1 favour the column-defining treatment. Emboldened values $P<0.05$

The ranking of the study drugs for prevention of relapse or reoccurrence was 1) olanzapine, 2) quetiapine, 3) valproate 4) lithium, (with placebo ranking last). The ranking for all-cause discontinuation was 1) quetiapine, 2) olanzapine, 3) valproate, 4) lithium (with placebo ranking last) (Table 5.iii; Figure 5.vi).
Figure 5.iv Forest plot for relapse or reoccurrence
Figure 5.v Forest plot for all-cause discontinuation
Table 5.iii Probability of each treatment being ranked best-worst; for preventing relapse/reoccurrence, and for lowest all-cause discontinuation

<table>
<thead>
<tr>
<th>Rank</th>
<th>Placebo</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>0.0</td>
<td>0.6</td>
<td>12.8</td>
<td>51.3</td>
<td>35.3</td>
</tr>
<tr>
<td>2nd</td>
<td>0.0</td>
<td>7.0</td>
<td>25.1</td>
<td>32.4</td>
<td>35.6</td>
</tr>
<tr>
<td>3rd</td>
<td>0.0</td>
<td>36.8</td>
<td>33.6</td>
<td>12.0</td>
<td>17.6</td>
</tr>
<tr>
<td>4th</td>
<td>0.0</td>
<td>55.6</td>
<td>28.5</td>
<td>4.4</td>
<td>11.5</td>
</tr>
<tr>
<td>Worst</td>
<td>100</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Placebo</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>0.0</td>
<td>0.3</td>
<td>7.7</td>
<td>27.2</td>
<td>64.8</td>
</tr>
<tr>
<td>2nd</td>
<td>0.0</td>
<td>5.5</td>
<td>26.6</td>
<td>47.0</td>
<td>20.9</td>
</tr>
<tr>
<td>3rd</td>
<td>0.1</td>
<td>25.1</td>
<td>45.9</td>
<td>18.4</td>
<td>10.5</td>
</tr>
<tr>
<td>4th</td>
<td>1.4</td>
<td>68.7</td>
<td>18.8</td>
<td>7.3</td>
<td>3.7</td>
</tr>
<tr>
<td>Worst</td>
<td>98.5</td>
<td>0.4</td>
<td>1.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure 5.vi Cumulative probability of ranking

5.5 Discussion

5.5.1 Main findings

Lithium, valproate, olanzapine and quetiapine were significantly more efficacious than placebo with regards to preventing relapse or reoccurrence of any mood episode in BPD. All drugs were also superior in terms of reduced all-cause discontinuation. Ranking of the interventions suggested olanzapine may be most effective in terms of preventing new mood episodes and quetiapine may be least associated with all-cause discontinuation. This is consistent with the results of the analysis of Miura et al. (Miura et al., 2014). However, these authors suggest lithium should be first line as it shows (relative to placebo) efficacy in
any mood episode and superiority for specific trials of emergent manic and depressive mood episodes. It also has the most robust and unbiased evidence. Whereas they state that olanzapine is not better than placebo for preventing depressive episodes, and quetiapine studies (though showing superiority at both polarities compared to placebo) were biased by using enrichment designs. In this analysis, although valproate was superior to placebo for preventing any mood relapse, it was not superior for specific manic or depressive relapses (Miura et al., 2014). However, as with the current analysis, the authors failed to find a significant risk ratio for any comparison of lithium, valproate, olanzapine or quetiapine with another of these drugs.

5.5.2 Strengths and limitations

This NMA only studies the commonly used maintenance mood stabilisers, and as such displays better transitivity and connectivity than the work of Miura et al. (Miura et al., 2014). The evidence network in my NMA was relatively well connected, but often comparisons were formed of one or two trials. I did not attempt to examine specific polarities of relapse in this analysis, as this has been recently completed (Miura et al., 2014). It was also not possible to separate trials by BPD type (bipolar II or rapid-cycling, for example). There are a number of important limitations to the RCTs included in the NMA. Trials lasted two years at most – therefore any more prolonged treatment regimen (as will typically be the case for maintenance treatment in BPD) goes beyond the evidence-base. RCTs have tended to look at emergence of new mood episodes, but the prevention of subsyndromal symptoms will also be important in clinical practice and will be more challenging to measure. Many of the studies included were funded by pharmaceutical companies, raising concerns about sponsorship bias, and 5 of the 18 trials were enriched, favouring patients who had already responded to one of the study drugs. This trial design can give an advantage to the drug in question (often this occurred when newer drugs were
compared with lithium). Additionally, many trials include patients who have been previously exposed to the study drugs, so their response may be predictable. Ideally, RCTs should involve incident treatment, but this is rarely possible.

The RCT selection process discussed in **Section 5.3.1** and **Section 5.3.2**, may have potentially missed trials, because of inadequate search terms or because they remain unpublished. I may also have excluded relevant RCTs because of the ‘intention-to-treat analysis’ criteria. Studies included in the NMA by Miura et al. (Miura et al., 2014), but excluded in my NMA were a trial that only lasted for 17.3 weeks (Prien et al., 1973b) and a study which appeared to overlap other reported RCT data (Fieve et al., 1976). My NMA included one RCT which was not in Miura et al. (Stallone et al., 1973).

### 5.5.3 Conclusions

From the results of this NMA, lithium was not convincingly superior to valproate, olanzapine or quetiapine. All of the drugs considered, which were recommended by NICE (National Institute for Health and Care Excellence, 2006), are superior to placebo for relapse and reoccurrence prevention and all-cause discontinuation, but none is superior to the others in direct or indirect comparison using all available RCTs.

### 5.5.4 Implications of the findings of this network meta-analysis

Given that there are unlikely to be trials making head-to-head comparisons of these medications in the future (because they are all now off-patent), further research into which drug is most appropriate for maintenance BPD treatment will only come from indirect comparisons (where one of these established drugs is used as the control treatment trials of a new drug) or via non-trial clinical effectiveness studies. In **Chapter 6**, I undertake such a study, comparing the time to stopping the study drug, switching to an alternative medication or add-on of another mood stabiliser, antipsychotic, antidepressant or
benzodiazepine. This outcome is similar to that used in a number of RCTs in this area and represents a combination of both effectiveness and tolerability.

One of the major limitations of all trials of maintenance mood stabiliser medication considered is that not one has provided more than two years of follow-up (Miura et al., 2014, Severus et al., 2014). However, when choosing a drug for long-term if not life-long treatment, efficacy and safety are important not only during the first years, but also thereafter. The use of lithium over longer time periods is associated with risk of kidney dysfunction (McKnight et al., 2012), the use of SGAs with metabolic syndrome and an increased mortality risk due to cardiovascular problems (Bobo et al., 2011, Vieta, 2004), and the use of valproate with weight gain, liver failure and haematological abnormalities (Perucca, 2002). Although it has not yet been established how these late adverse effects have an impact on the long-term safety of these drugs, the assumption made by NICE is that over 10 or more years the safety of lithium is at least in balance with that of SGAs (National Institute for Health and Care Excellence, 2014). However, it is clear that longer-term studies of effectiveness, tolerability and safety of these drugs, reflecting real world use are necessary. To address this I designed and conducted a longitudinal study using EHR, where longer follow-up data are available (Chapter 7), to study of some of the recognised, but previously poorly quantified, adverse effects of these maintenance medications.
Chapter 6 Comparison of the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder

6.1 Summary

6.1.1 Objective
To compare the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine using rates of time to cessation of treatment, or add-on of another psychotropic medication as a proxy measure, while accounting for propensity to be prescribed one of these mood stabilisers.

6.1.2 Methods
Cohort study using a representative, anonymous UK primary care data collected 1995-2013. 5089 patients with BPD were prescribed lithium (N=1505), valproate (N=1173) olanzapine (N=1336) or quetiapine (N=1075) as monotherapy. Treatment failure was defined as time to stopping medication, switching to another study drug, or add-on of another anticonvulsant, antipsychotic, antidepressant or benzodiazepine.

6.1.3 Results
Individuals prescribed lithium had a longer time to treatment failure than those prescribed the other study drugs. This remained the case after propensity score adjustment for key predictors of treatment allocation. Compared to lithium, valproate had an elevated HR (1.19; 95% CI 1.09 to 1.31) as did olanzapine (HR 1.16; 95% CI 1.05 to 1.28) and quetiapine
(HR 1.30; 95% CI 1.18 to 1.44). This relationship remained in a propensity score matched analysis, when treatment failure was defined as stopping, swapping or add-on of an anticonvulsant or antipsychotic, and when treatment failure was restricted to greater than 3 months after commencing the study drug.

6.1.4 Conclusion
Lithium appears to be more successful as monotherapy maintenance treatment than valproate, olanzapine or quetiapine. People receiving valproate, olanzapine or quetiapine require alternative or additional treatments earlier, which may indicate worse outcomes and may cause additional side effects.

A modified version of this chapter was published as Hayes JF, Marston L, Walters K, Geddes JR, King M, Osborn DP. Lithium vs. valproate vs. olanzapine vs. quetiapine as maintenance monotherapy for bipolar disorder: a population-based UK cohort study using electronic health records. World Psychiatry. 2016; 15: 53-8
6.2 Introduction

A number of drug treatments are recommended for maintenance in BPD (as discussed in Chapter 1 and Chapter 5). In the UK, the most commonly used medications are lithium, valproate, olanzapine and quetiapine (Hayes et al., 2011). This reflects previous National Institute for Health and Care Excellence (NICE) guidance on first-line monotherapy maintenance treatment, which suggested equivalence of these drugs (National Institute for Health and Care Excellence, 2006). An update of this guidance in September 2014 suggested that lithium should be first-line (National Institute for Health and Care Excellence, 2014). Globally there is a range of prescribing advice, which includes additionally: lamotrigine, carbamazepine, oxcarbazepine, aripiprazole and other SGAs (Association and Kernberg, 2002, Goodwin et al., 2016, Mok et al., 2011, Yatham et al., 2013). Recent meta-analyses and network meta-analyses have highlighted the superiority of lithium and these results have contributed to the change in NICE guidance (Miura et al., 2014, Severus et al., 2014). However, no one randomised controlled trial (RCT) has conclusively proved the benefit of lithium over other drugs, and there are no trials that compare valproate with olanzapine, valproate with quetiapine or olanzapine with quetiapine directly (Chapter 5). The applicability of RCT results to people with BPD in the real world may be limited by diagnostic heterogeneity, diagnosis or treatment rejection, and complex, labile presentations of the illness that occur over the life-course (Baldessarini, 2002, Reed et al., 2009).

As discussed in Chapter 1 and Chapter 2, EHRs offer an opportunity to augment RCT findings with head-to-head comparison studies which include large numbers of patients, representative of real world clinical practice and long follow-up periods. Using data from
THIN, I aimed to compare rates of stopping, switch to, or add-on of, another psychotropic drug in individuals prescribed lithium, valproate, olanzapine or quetiapine as maintenance monotherapy for BPD. This outcome represents a combination of both effectiveness and tolerability of the study medication and is similar to that used in many RCTs of maintenance treatment for BPD (Miura et al., 2014, Severus et al., 2014).

6.3 Methods

6.3.1 Study design & setting

For this study, I defined, extracted and analysed a large prospective cohort of patients in THIN from January 1st 1995 and December 31st 2013 (Figure 6.i). As discussed in Chapter 2 in the UK, GPs are responsible for issuing all drug prescriptions if treatment is ongoing, following advice from a psychiatrist, and this information is well recorded in THIN, as prescriptions are issued electronically.

6.3.2 Participants

Patients with a diagnosis of BPD were included if they had 2 or more consecutive prescriptions for treatment lasting 28 days of lithium, valproate, olanzapine or quetiapine after 1 January 1995, or after the date at which the GP practice met quality assurance criteria for data entry (based on computer usage and mortality recording rates) (Horsfall et al., 2013, Maguire et al., 2009). Patients were excluded if they received a diagnosis of schizophrenia at any time. They were also excluded if they were prescribed another of the study drugs, or any other anticonvulsant, antipsychotic, antidepressant or benzodiazepine at the start of follow-up, or in the month before this. The cohort was therefore one in which the intention was to treat with lithium, valproate, olanzapine or quetiapine monotherapy. Last date of follow-up was date of death, leaving the GP practice or the end of the study period (December 31 2013).
Figure 6.1 Flow diagram of included patients

AMR, Acceptable Mortality Reporting; ACU, Acceptable Computer Usage
6.3.3 Main outcome

Patients were followed-up until they stopped the study drug, or had another study drug, an anticonvulsant, an antipsychotic, an antidepressant or a benzodiazepine added to their treatment regimen. Date of first prescription was taken as the start of exposure time, the end of the prescription was calculated from the prescription length and prescribing instructions coded by the GP. Patients were considered to have a period of continuous prescribing if another prescription for the same drug was issued within three months of the predicted end date. If this did not occur, the date of stopping the study drug was the end date of the final prescription.

6.3.4 Observed pre-treatment variables for propensity score estimation

Sociodemographic, psychiatric and physical health characteristics at baseline were extracted from each patient’s medical record. Psychiatric and physical health problems were considered present if referenced in the patient notes. If a patient had multiple entries of the same (or similar) Read codes, the start date of the condition was taken as the earliest date of entry.

As discussed in Chapter 2 a PS for each individual was estimated using variables defined *a priori*, based on existing research (Holmes, 2013, Rosenbaum and Rubin, 1984). The PS attempts to account for all of the covariates that predict receiving a particular study drug (Holmes, 2013, Rosenbaum and Rubin, 1983). The PS was then checked by comparison of covariate balance across treatments, within strata. The included variables were: sex, age at start of treatment with the study drug, year of entry to the cohort, ethnicity (grouped as White, Black, Asian, mixed, other, with missing values coded as White), physical health history at baseline (ischemic heart disease, myocardial infarction, cerebrovascular event, hypertension, renal disease, thyroid disease, liver disease, T2DM, epilepsy, history of
alcohol dependence, history of illicit drug use), smoking status (grouped as never-smoker, ex-smoker, current smoker), BMI (grouped as healthy weight, overweight (BMI 25 to 30), obese (BMI over 30)), mental health history at baseline (history of anxiety symptoms, hypomania as most proximal diagnosis code, history of depressive symptoms, sleep disturbance, previous treatment with the study drug before baseline, incident diagnosis of BPD) and clustering by GP practice. For demographic and health-related covariates, the entire medical record prior to baseline was reviewed (potentially including records anteceding 1988, when paper records were transposed to EHR). For BMI, alcohol use, and smoking status the most proximate data in the 5 years before baseline was used. These variables were selected because they represent factors influencing prescribing choice (such as risk factors for adverse effects with a particular study medication) (National Institute for Health and Care Excellence, 2014).

Although PS estimation cannot remove all bias, it has been postulated to also reduce confounding from unmeasured variables, because of their association with measured covariates (Austin et al., 2005, Joffe and Rosenbaum, 1999). Therefore in this study, for a given PS, exposure to lithium, valproate, olanzapine or quetiapine is presumed to be at random (Becker and Ichino, 2002).

6.3.5 Statistical Analysis

Cox regression analyses were conducted comparing the rates of switch to, or add-on of, another psychotropic medication in the four treatment groups. Analyses were adjusted for sex, age, ethnicity and calendar year. Time to treatment failure was summarised by Kaplan-Meier curves. The proportional hazards model was tested formally with analysis of Schoenfeld residuals (Schoenfeld, 1982). The PS was calculated using multinomial logistic regression using the covariates described as independent variables, with drug treatment as
the dependent variable. The PS was then used as a linear term in a Cox regression analysis that also included age and calendar year (d’Agostino, 1998). This model was shown to be superior to stratifying on PS using Akaike information criterion and Bayesian information criterion (Lin and Dayton, 1997), and was a more efficient use of data than PS matching, because it uses all patients.

Analysis using PS matching was then completed. As discussed in Section 2.11.1, although matched analyses may include a non-representative sample of patients receiving treatment, they may provide a more valid estimate of treatment effect as they compare patients with similar observed characteristics (d’Agostino, 1998, Rosenbaum and Rubin, 1985). Pairwise matching was performed for each patient in the valproate, olanzapine and quetiapine groups with individuals in the lithium treated group. Patients were matched on a one-to-one basis if their PS was within 0.01 of each other; all other patients were dropped from the analysis.

Supplementary analyses excluding benzodiazepine and antidepressant add-on as a source of treatment failure were carried out. A supplementary analysis excluding patients who stopped, swapped or had psychotropic medication added-on within the first three months of follow-up was also performed.

6.4 Results

A total of 14,396 individuals had a diagnosis of BPD. Of these, 5,089 were prescribed monotherapy with one of the study drugs at the start of cohort follow-up: lithium was prescribed to 1,505 people, valproate to 1,173, olanzapine to 1,336 and quetiapine to 1,075 people (Figure 6.ii). Individuals prescribed lithium tended to be older than other groups, with more years of follow-up data and fewer GP practice contacts during this period. They
were less likely to have a previous record of depression in their notes and less likely to be an incident case (Table 6.1).

Figure 6.ii Flow diagram of included patients
Table 6.i Characteristics of patients with bipolar disorder prescribed lithium, valproate, olanzapine or quetiapine monotherapy

<table>
<thead>
<tr>
<th></th>
<th>lithium</th>
<th>valproate</th>
<th>olanzapine</th>
<th>quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>1505</td>
<td>1173</td>
<td>1336</td>
<td>1075</td>
</tr>
<tr>
<td>Female, N (%)</td>
<td>860 (57.14)</td>
<td>631 (53.79)</td>
<td>733 (54.87)</td>
<td>735 (68.37)</td>
</tr>
<tr>
<td><strong>Age at entry to the cohort, median (IQR)</strong></td>
<td>44.94 (35.45–58.66)</td>
<td>41.56 (31.40–53.69)</td>
<td>40.86 (31.88–52.74)</td>
<td>38.46 (29.31–49.77)</td>
</tr>
<tr>
<td>Total years of follow-up, median (IQR)</td>
<td>4.23 (1.54–8.62)</td>
<td>3.05 (1.13–6.27)</td>
<td>3.57 (1.36–6.88)</td>
<td>2.12 (0.87–3.91)</td>
</tr>
<tr>
<td>GP practice contacts per year of follow-up, median (IQR)</td>
<td>12.07 (7.10–19.74)</td>
<td>14.81 (8.69–23.75)</td>
<td>14.33 (8.84–24.60)</td>
<td>17.95 (11.76–26.92)</td>
</tr>
<tr>
<td>Non-white ethnic background, N (%)</td>
<td>44 (2.92)</td>
<td>50 (4.26)</td>
<td>65 (4.87)</td>
<td>35 (3.26)</td>
</tr>
<tr>
<td><strong>Health at baseline, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHD, MI, CVE history</td>
<td>76 (5.05)</td>
<td>80 (6.82)</td>
<td>58 (4.34)</td>
<td>41 (3.81)</td>
</tr>
<tr>
<td>Renal disease history</td>
<td>51 (3.39)</td>
<td>36 (3.06)</td>
<td>33 (2.47)</td>
<td>42 (3.91)</td>
</tr>
<tr>
<td>Thyroid disease history</td>
<td>161 (10.70)</td>
<td>89 (7.59)</td>
<td>89 (6.66)</td>
<td>75 (6.98)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>77 (5.12)</td>
<td>87 (7.42)</td>
<td>42 (3.14)</td>
<td>71 (6.60)</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>29 (1.93)</td>
<td>82 (6.99)</td>
<td>37 (2.77)</td>
<td>34 (3.16)</td>
</tr>
<tr>
<td>Obesity (BMI&gt;30)</td>
<td>617 (41.00)</td>
<td>488 (41.60)</td>
<td>482 (36.08)</td>
<td>467 (43.44)</td>
</tr>
<tr>
<td>Previous anxiety problems</td>
<td>98 (6.51)</td>
<td>102 (8.70)</td>
<td>133 (9.96)</td>
<td>154 (14.33)</td>
</tr>
<tr>
<td>Previous alcohol dependence</td>
<td>7 (0.47)</td>
<td>3 (0.26)</td>
<td>12 (0.90)</td>
<td>7 (0.65)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>518 (34.42)</td>
<td>462 (39.39)</td>
<td>571 (42.74)</td>
<td>425 (39.53)</td>
</tr>
<tr>
<td><strong>Bipolar disorder characteristics at baseline, N (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incident diagnosis</td>
<td>318 (19.62)</td>
<td>396 (33.99)</td>
<td>543 (41.71)</td>
<td>416 (40.82)</td>
</tr>
<tr>
<td>Previous depressive episode</td>
<td>845 (56.15)</td>
<td>701 (59.76)</td>
<td>826 (61.83)</td>
<td>788 (73.30)</td>
</tr>
<tr>
<td>Hypomania as most recent diagnosis</td>
<td>234 (15.55)</td>
<td>154 (13.13)</td>
<td>238 (17.81)</td>
<td>125 (11.63)</td>
</tr>
<tr>
<td>Previous record of taking study drug</td>
<td>936 (62.19)</td>
<td>507 (43.22)</td>
<td>463 (34.66)</td>
<td>328 (30.51)</td>
</tr>
</tbody>
</table>

IQR, interquartile range; IHD, ischemic heart disease; MI, myocardial infarction; CVE, cerebrovascular event; BMI, body mass index

In unadjusted analyses, the duration of successful monotherapy was longest in individuals treated with lithium (Figure 6.iii, Table 6.ii). The median time to treatment failure (as defined by stopping, switching or add-on of medication) in the lithium monotherapy group was 0.28 years (95% CI 0.23 to 0.35) compared to 0.22 years (95% CI 0.19 to 0.27) in the valproate group, 0.24 years (95% CI 0.21 to 0.28) in the olanzapine group and 0.17 years (95% CI 0.14 to 0.21) in the quetiapine group. Treatment failure had occurred in 75% of those prescribed quetiapine by 0.76 years (95% CI 0.64 to 0.84) compared to lithium; 2.05 years (95% CI 1.63 to 2.51), valproate; 0.98 years (95% CI 0.84 to 1.18) and olanzapine; 1.13
years (95% CI 1.00 to 1.31). The differences between treatments became more apparent the longer the duration of treatment (Figure 6.iii).

Table 6.ii Rates of treatment failure by drug

<table>
<thead>
<tr>
<th>Drug</th>
<th>N events</th>
<th>PYAR (100s)</th>
<th>Rate, per 100 PYAR (95% CI)</th>
<th>Treatment failure, Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithium</td>
<td>1151</td>
<td>1570</td>
<td>73.31 (65.55–81.78)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Valproate</td>
<td>909</td>
<td>777</td>
<td>116.93 (102.95–132.39)</td>
<td>1.25 (1.14–1.37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.22 (1.11–1.34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.19 (1.09–1.31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.20 (1.10–1.32)</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>977</td>
<td>893</td>
<td>109.36 (96.30–123.67)</td>
<td>1.19 (1.08–1.30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.19 (1.08–1.30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.16 (1.05–1.28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.17 (1.07–1.29)</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>814</td>
<td>457</td>
<td>177.94 (157.87–199.84)</td>
<td>1.48 (1.35–1.62)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.31 (1.19–1.44)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30 (1.18–1.44)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.32 (1.20–1.45)</td>
</tr>
</tbody>
</table>

*aAdjusted for clustering by primary GP practice, age, sex and calendar year, *bAdjusted for PS, clustering by GP practice, age and calendar year, *cPS matched (pairwise matching with lithium) adjusted for clustering by GP practice, age and calendar year

Figure 6.iii Time to treatment failure (stopped treatment or addition of mood stabiliser, antipsychotic, antidepressant or benzodiazepine) (unadjusted)
Lithium’s superiority remained after adjustment for clustering by GP practice, age, sex, calendar year, and ethnicity. It also remained after adjusting for PS, age and calendar year, and after matching by PS (Table 6.ii), with olanzapine having the least elevated HR (1.16, 95% CI 1.05 to 1.28). Compared to olanzapine, quetiapine had an increased rate of monotherapy failure (HR 1.12, 95% CI 1.02 to 1.23) in the PS adjusted model. Compared to valproate, olanzapine and quetiapine had similar rates of treatment failure (HR 0.97; 95% CI 0.89 to 1.06 and HR 1.09; 95% CI 0.99 to 1.19 respectively). The proportional hazards assumption held for all analyses. Before pairwise matching, PS scores were most different for lithium (median 0.45, IQR 0.25 to 0.61) and quetiapine (median 0.14, IQR 0.8 to 0.25) (Figure 6.iv). After matching the median PS was 0.21 (IQR 0.13 to 0.30) for lithium and 0.14 for quetiapine (IQR 0.8 to 0.25), this comparison included 626 patients in each group (Figure 6.v).

Individuals prescribed lithium or valproate were more likely to require antipsychotic add-on (19.53% and 18.41% respectively) than those prescribed olanzapine or quetiapine monotherapy (10.25% and 9.02% respectively). Conversely, individuals prescribed olanzapine and quetiapine were more likely to require mood-stabiliser add-on (14.07% and 12.56% respectively) compared to lithium and valproate (6.71% and 5.20% respectively).
Figure 6.iv Propensity score distribution before matching

Figure 6.v Propensity score distribution after pairwise matching with lithium

Continued overleaf
Supplementary analyses produced similar results to the primary analyses. If treatment failure was restricted to stopping medication, swapping to an alternative study drug or add-on of a mood stabiliser or antipsychotic medication then PS adjusted HRs were elevated for all drugs compared to lithium (Table 6.iii). The same was true if patients failing in the first three months of follow-up were excluded from the analysis (Table 6.iii, Figure 6.vi).
Table 6.iii Supplementary analyses using PS adjusted model

<table>
<thead>
<tr>
<th></th>
<th>Excluding benzodiazepine add-on&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Excluding benzodiazepine and antidepressant add-on&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Excluding failures in the first 3 months of treatment&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithium</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Valproate</td>
<td>1.25 (1.14–1.37)</td>
<td>1.18 (1.08–1.29)</td>
<td>1.22 (1.06–1.40)</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>1.10 (1.00–1.22)</td>
<td>1.17 (1.07–1.28)</td>
<td>1.26 (1.09–1.45)</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>1.25 (1.13–1.38)</td>
<td>1.13 (1.03–1.25)</td>
<td>1.20 (1.04–1.40)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Treatment failure represents stopping medication or requiring a mood stabiliser or antipsychotic. <sup>b</sup>Treatment failure represents stopping medication or requiring a mood stabiliser, antipsychotic or antidepressant as add-on. <sup>c</sup>Treatment failure after a 3 month period of stabilisation.

Figure 6.vi Time to treatment failure (excluding failures in the first 3 months of treatment) (unadjusted)
6.5 Discussion

6.5.1 Main findings
As far as I am aware this study represents the only head-to-head comparison of the four most common maintenance treatments for BPD, and has the longest follow-up (up to 14.5 years) and largest cohort (over 5000 patients) of any direct comparison of treatment for BPD. RCTs making these comparisons do not exist and are unlikely to be possible. The overall rate of treatment failure (represented by stopping index medication, swapping to an alternative study medication or requiring add-on of a mood stabiliser, antipsychotic, antidepressant or benzodiazepine) was increased for valproate, olanzapine and quetiapine when compared to lithium. This was also true if failures within the first three months were excluded (i.e., once the patient had been stabilised on the prescribed drug). This analysis may more closely capture the effectiveness of the drug, as tolerability and resolving mood episode issues are most likely to arise in the period directly after drug initiation. These results suggest that monotherapy with lithium may be more successful than the other recommended drugs. Monotherapy treatment failure appear to be common (often early in treatment), but this is consistent with other naturalistic studies (Kessing et al., 2007, Schumann et al., 1999). The rate of treatment failure was also elevated for quetiapine compared to olanzapine, but it was not possible to separate the other drugs from each other.

6.5.2 Comparison with previous literature
One previous study has used EHR to examine the comparative effectiveness of lithium and valproate (Kessing et al., 2011). This study found monotherapy with lithium before treatment failure to be longer than valproate monotherapy, and found greater difference between rates of treatment failure in those treated with lithium and valproate than my
study (HR 1.86, 95% CI 1.59 to 2.16). However, this study was limited by the small number of individuals prescribed valproate, and the potential for unmeasured confounding. I am not aware of other observational studies comparing effectiveness of maintenance mood stabilisers. As described in Chapter 1 and Chapter 5, combined evidence from RCTs does not reliably separate one active drug from the others in terms of superior efficacy (Miura et al., 2014).

6.5.3 Strengths and limitations
The use of contemporaneous, representative medical records mitigated the risk of potential biases relating to selection into the study. Patients could still potentially have been misdiagnosed (this is discussed in Section 9.5.2); however, there is less risk of this occurring in this study compared to the mortality/morbidity study (Study 1) because patients had to be diagnosed with BPD and receive an appropriate mood stabiliser. Information bias should partially have been avoided by the use of prescribing data as exposure; in the UK GPs are responsible for on-going prescribing within the NHS (Health and Social Care Informaiton Centre, 2012), which is detailed and well recorded in THIN. However, exposure to the study drug was approximated through prescriptions issued to patients, and may not reflect how the patient used the medication. Poor adherence to prescribed drug regimens is a problem with all medications, and this is particularly true if side effects are unpleasant, as can be the case with all of the study drugs (Sajatovic et al., 2007, Sajatovic et al., 2006). In this study, stopping the drug will be reflected in the outcome, but erratic adherence cannot be detected. It is possible that erratic adherence is more likely for drugs other than lithium (as this is more closely monitored through regular blood tests) and may have contributed to lithium’s perceived superiority, but other longitudinal cohort studies have not shown differential adherence (Sajatovic et al., 2007).
Treatment failure was defined as stopping the drug or initiation of any mood stabiliser, antipsychotic, antidepressant or benzodiazepine. It is likely that addition of a mood stabiliser or antipsychotic represents more serious treatment failure than addition of an antidepressant (which would only occur during a depressive relapse) or a benzodiazepine (which may be used short-term to avoid a relapse). A supplementary analysis excluding addition of these drugs had similar results. In my initial analysis plan, I had hoped to additionally use psychiatric hospitalisation as an outcome. It was expected that these data would be available via a linkage between THIN and the Hospital Episode Statistics (HES) database, which contains records of secondary care outpatient contacts and inpatient admissions (Sinha et al., 2013). However, by the time of writing up this thesis this linkage remains unavailable, in part because of the failure of the care.data project, which aimed to link numerous UK health data (Carter et al., 2015).

It may be the case that both of these outcomes fail to capture what is important to patients in terms of relapse, reoccurrence, functioning and quality of life. These factors are not measured in the data. However, through examining monotherapy treatment failure I believe I have described a proxy for these important outcomes that captures both tolerability and effectiveness and highlights a very common need for adjunctive drug treatments. This outcome has also been used in a number of RCTs of maintenance drug treatment for BPD and therefore comparison with these results is possible. For example, the largest trial of lithium versus valproate treatment had a primary outcome of “time to new intervention for an emerging mood episode” (Geddes et al., 2010). This trial found similar results to my study (HR 1.41, 95% CI 1.00 to 1.92) but was not powered to directly compare lithium and valproate.
A limitation of interpretation of data from studies such as this is the inability to rule out important unmeasured confounding effects. I attempted to account for confounding by indication by building a PS model that included important clinical predictors of treatment allocation (Rosenbaum and Rubin, 1983). This included physical health variables that may lead a clinician to avoid a certain drug because of its side effect profile, for example renal disease with lithium or CVD with olanzapine. Characteristics such as sex, age, and BMI were also included as valproate is contraindicated in women of childbearing potential (Hayes et al., 2011), and olanzapine has the potential to cause rapid weight-gain (Eder et al., 2014). Adjusting for the GP practice should have accounted for physician preference for a particular drug. Once these covariates were adjusted for, there was a similar propensity for patients to be prescribed valproate, olanzapine or quetiapine, with patients prescribed lithium having slightly higher scores. Despite this, I cannot rule out the possibility that these confounders were imperfectly adjusted for or that other important confounders were not included in the PS model.

Unfortunately, I was unable to separate treatment failure relating to emergent manic (or hypomanic) episodes from depressive episodes, and there is evidence that the study drugs may be differentially effective in preventing a particular polarity of illness (Miura et al., 2014). However, an ideal “mood stabiliser” would protect against both polarities of relapse (Bauer and Mitchner, 2004), and this is what my study captures. I was also unable to examine the physician’s reason for treatment initiation, and it may be that quetiapine’s apparent inferiority is because in some patients it is prescribed not as maintenance treatment, but for shorter-term indications (which I hoped to capture in the supplementary analysis). There were too few patients on monotherapy with other recommended maintenance treatments, such as lamotrigine or aripiprazole, to include these drugs in the analysis.
6.5.4 Conclusions

This study provides necessary supplementary and complementary evidence to RCT findings for maintenance treatments for BPD. In clinical practice, lithium appears to be the most effective treatment to prevent any relapse or reoccurrence of BPD and may prolong the time before adjunctive prescribing is necessary. This finding echoes the results of recent meta-analyses that suggest lithium is superior to these drugs in protecting against both manic and depressive relapse (Miura et al., 2014, Severus et al., 2014). This is important as lithium is often avoided because of its side effect profile (Shine et al., 2015), but monotherapy with valproate, olanzapine or quetiapine is more likely to fail sooner and may result in patients experiencing the additive side effects of multiple psychotropic drugs.

6.5.5 Implications of the findings of this study

As EHRs grow in size and number, it will become possible to run these sorts of comparative effectiveness and tolerability studies for a number of medications and indications. As stated, in terms of BPD treatments, THIN is currently too small to examine newer medications such as aripiprazole or lamotrigine, and combination treatments such as lithium and valproate. These potential studies could powerfully augment RCT results, and there has been a move to elevate the level of evidence provided by such studies by the British Association for Psychopharmacology (Goodwin et al., 2016). Alongside advances in the size of EHRs, analytical techniques will need to be developed further to manage concerns about residual confounding. Potential techniques are discussed in Chapter 9. Further implications for practice, policy and future research of the results of this study and the NMA (Chapter 5) are discussed in Section 9.4. I examine additional issues that need to be addressed in the selection of maintenance treatment in the next two chapters. Firstly, is the potential for adverse physical health outcomes related to the drug (Chapter 7)?
what is the potential for adverse psychiatric outcomes, such as self-harm and suicide (Chapter 8)?
Chapter 7  Comparison of the adverse renal, endocrine, hepatic and metabolic events during treatment with different maintenance mood stabiliser medications for bipolar disorder

7.1  Summary

7.1.1  Objective
To calculate rates of adverse chronic renal, hepatic, endocrine and metabolic effects in individuals prescribed lithium, valproate, olanzapine and quetiapine, accounting for propensity to be prescribed one of these mood stabilisers.

7.1.2  Methods
I conducted a propensity score adjusted cohort study using nationally representative United Kingdom electronic health records from January 1 1995 until 31 December 2013. Included patients had a diagnosis of BPD and were prescribed lithium (N=2148), valproate (N=1670) olanzapine (N=1477) or quetiapine (N=1376) as maintenance mood stabiliser treatment. Adverse outcomes were chronic kidney disease, thyroid disease, hypercalcaemia, weight gain, hypertension, T2DM, CVD and hepatotoxicity. The propensity score included important demographic, physical health and mental health predictors of drug treatment allocation.

7.1.3  Results
Compared to patients prescribed lithium, those taking valproate, olanzapine and quetiapine had reduced rates of chronic kidney disease stage 3 or more severe, following adjustment
for propensity score, age, calendar year and accounting for clustering by primary care practice (valproate HR 0.56; 95% CI 0.45 to 0.69; P<0.001, olanzapine HR 0.57; 95% CI 0.45 to 0.71; P<0.001, quetiapine HR 0.62; 95% CI 0.47 to 0.80; P<0.001). Hypothyroidism was reduced in those taking valproate (HR 0.60; 95% CI 0.40 to 0.89; P=0.012) and olanzapine (HR 0.48; 95% CI 0.29 to 0.77; P=0.003), compared to those taking lithium. Rates of new onset hyperthyroidism (valproate HR 0.24; 95% CI 0.09 to 0.61; P=0.003, olanzapine HR 0.31; 95% CI 0.13 to 0.73; P=0.007) and hypercalcemia (valproate HR 0.25; 95% CI 0.10 to 0.60; P=0.002, olanzapine HR 0.32; 95% CI 0.14 to 0.76; P=0.008, quetiapine HR 0.23; 95% CI 0.07 to 0.73; P=0.013) were also reduced relative to lithium. However, rates of greater than 15% weight gain on valproate, olanzapine and quetiapine were higher (valproate HR 1.55; 95% CI 1.28 to 1.86; P<0.001, olanzapine HR 1.64; 95% CI 1.35 to 2.00; P<0.001, quetiapine HR 1.48; 95% CI 1.16 to 1.87; P<0.001) than in individuals prescribed lithium, as were rates of hypertension in the olanzapine treated group (HR 1.41, 95% CI 1.06 to 1.87; P=0.017). I found no significant difference in rates of chronic kidney disease stage 4 or more severe, T2DM, CVD or hepatotoxicity. Despite estimates being robust following sensitivity analyses, limitations include the potential for residual confounding and ascertainment bias, and an inability to examine dosage effects.

7.1.4 Conclusions

Lithium use is associated with more renal and endocrine adverse events, but less weight gain than commonly used alternative mood stabilisers. Risks need to be offset with the effectiveness and anti-suicidal benefits of lithium, and potential metabolic side effects of alternative treatment options.

A modified version of this chapter was published as Hayes JF, Marston L, Walters K, Geddes JR, King M, Osborn DP. Adverse renal, endocrine, hepatic, and metabolic events
7.2 Introduction

As discussed in Chapter 1, a number of adverse effects of lithium have been identified since its use as a mood stabiliser became established in the 1970s (Bech, 2006), but it is only recently that they have begun to be characterised and quantified (Close et al., 2014, Kessing et al., 2015, McKnight et al., 2012, Murru et al., 2015, Shine et al., 2015). Lithium’s adverse effects include renal, thyroid, and parathyroid dysfunction. Lithium is also recognised to cause weight gain, but the risk of weight gain relative to other potential maintenance therapies has not been widely investigated (McKnight et al., 2012). Alternatives, such as SGAs and valproate, have been found to be obesogenic (Tarricone et al., 2010), especially olanzapine, which is the most commonly prescribed antipsychotic in BPD (Hayes et al., 2011). Weight gain is associated with a number of adverse events such as hypertension, T2DM and CVD (Haupt, 2006). Valproate, olanzapine and quetiapine are metabolised by the liver. Valproate has been found to be associated with a high risk of asymptomatic elevated transaminases and can cause idiosyncratic hepatic failure (Dols et al., 2013, Murru et al., 2015). Olanzapine and quetiapine have also been associated with rare cases of hepatotoxicity (Atasoy et al., 2007, El Hajj et al., 2004, Ozcanli et al., 2006). Therefore, the balance of risks associated with maintenance mood stabiliser selection is not straightforward, and I am aware of no studies that make these comparisons across treatment options.

This study uses EHR to compare rates of major recognised adverse outcomes amongst individuals prescribed lithium, valproate, olanzapine or quetiapine for mood stabilisation in BPD. The adverse events examined are CKD, hypothyroidism, hyperthyroidism, hypercalcemia, weight gain, hypertension, T2DM, CVD and hepatotoxicity (Dols et al., 2013, Murru et al., 2015).
7.3 Methods

7.3.1 Study design & Setting
As with the previous study (Chapter 6), this study used a prospective cohort of patients in THIN from January 1st 1995 and December 31st 2013 (Figure 6.i). See Section 2.5, Section 4.3.1 and Section 6.3.1 for a full description of the data source.

7.3.2 Participants
As previously, patients with a diagnosis of BPD were included if they had 2 or more consecutive prescriptions for treatment lasting 28 days of lithium, valproate, olanzapine or quetiapine after 1 January 1995, or after the date at which the medical records met quality assurance criteria for data entry (based on computer usage and mortality recording rates). Patients were excluded if they were prescribed another study drug at the start of follow-up, or in the month before this. Diagnosis of BPD could occur at any time in the patient record. For each outcome requiring haematological or biochemical confirmation for diagnosis (CKD, thyroid disease, hypercalcemia, hepatotoxicity) patients were excluded from the primary analysis if they did not receive a specific blood test for the outcome, to reduce surveillance bias. For the weight gain outcome, patients were excluded if they did not have a baseline, or pre-treatment weight, and at least one other weight measurement. For the outcome of hyperthyroidism, patients taking thyroxine were excluded, as this can result in thyroid stimulating hormone (TSH) suppression (Beastall et al., 2006). Patients were also excluded if they had the outcome of interest at baseline (as I was interested in new/incident events). Therefore, each outcome has a different number of patients included.

7.3.3 Exposure
Date of first prescription was taken as the start of exposure time. The end of the prescription was calculated from the amount prescribed and dosage instructions coded by
the physician. Patients were considered to have a period of continuous prescribing if another prescription for the drug was issued within three months of the calculated end date. If this did not occur, the date of stopping the study drug was the end date of the final prescription. Three months was added to this end date to account for late development of the adverse event or delayed recording. Each patient could only contribute exposure time to one of the study drugs (the first they received) and did not re-enter the cohort if they restarted the drug after more than 3 months. In contrast to the effectiveness/tolerability study (Study 2: Chapter 6), patients could be prescribed other psychiatric medications but not combinations of the study drugs. If they commenced another study drug (i.e., lithium, valproate, olanzapine, quetiapine) they were censored (to ensure the outcome could be assigned to a particular drug).

7.3.4 Main outcomes
Outcomes of interest were: CKD stage 3 or above (or an estimated glomerular filtration rate [eGFR] of <60ml/min/1.73m²), CKD stage 4 or above (or an eGFR <30ml/min/1.73m²) (Crowe et al., 2008, Vassalotti et al., 2007), (if eGFR was unavailable I calculated it from available creatinine blood tests using the CKD-EPI equation (Levey et al., 2009)), hypothyroidism (or a TSH of >10mU/L), hyperthyroidism (or a TSH<0.1mU/L) (Beastall et al., 2006), hypercalcemia (adjusted calcium>2.65mmol/L) (Smellie et al., 2008), >7% and >15% weight gain from baseline (Manu et al., 2015), hypertension, T2DM (or HBA1c >48mmol/mol) (John, 2012), CVD (defined as any of IHD, MI or CVE) and hepatotoxicity (or alanine transaminase [ALT] >200U/L, or aspartate aminotransferase [AST] >250U/L) (Sabin, 2004). CKD, thyroid disease, T2DM, hypertension, CVD and other chronic health condition diagnoses have been validated in THIN (Blak et al., 2006, Blak et al., 2011).
Patients were followed-up until the earliest of i) the first record of the adverse event of interest, ii) the date of stopping the study drug plus three months, iii) date of switching to another study drug, iv) date of death or date of leaving the primary care practice, v) 31 December 2013.

7.3.5 Propensity score estimation using observed pre-treatment variables

A number of baseline patient characteristics were extracted from THIN. Physical and mental health conditions were considered present if referenced in patient notes and absent if they were not. If a patient had multiple entries of the same (or similar) codes, the start date of the condition was taken as the earliest date of entry.

A PS for each individual was estimated, as described in Section 2.11.1 and Section 6.3.4. Included variables were: sex, age at start of treatment with the study drug, year of entry to the cohort, ethnicity (grouped as White, Black, Asian, mixed, other, with missing values coded as White [44]), IHD diagnosis before baseline, history of MI, history of CVE, hypertension, CKD at baseline, history of hypo or hyperthyroidism, history of liver disease or hepatotoxicity, T2DM, epilepsy, alcohol use (grouped as none/low, moderate, high/dependent), history of illicit drug use, smoking status (grouped as never-smoker, ex-smoker, current smoker), BMI (grouped as healthy weight, overweight (BMI 25 to 30), obese (BMI over 30)), anxiety symptoms or diagnosis before baseline, depressive symptoms or diagnosis, sleep disturbance before baseline, treatment with one of the study drugs at or before baseline and clustering by practice in which the treating physician was working. The PS was checked by comparison of covariate balance across treatments, within strata. The variables in the PS excluded the outcome variable for that particular analysis.
7.3.6 Statistical Analysis

Cox regression analyses were conducted comparing the rates of adverse events in the four treatment groups. As in the effectiveness/tolerability study (Study 3: Chapter 6) the proportional hazards model was tested formally with analysis of Schoenfeld residuals (Schoenfeld, 1982). The PS was calculated using multinomial logistic regression using drug treatment as the dependent variable and the covariates described as independent variables. The PS was then used as a linear term in a Cox regression analysis that also included age, calendar year and clustering by practice (d’Agostino, 1998). In all cases this model was shown to be superior to stratifying on PS using Akaike information criterion and Bayesian information criterion (Lin and Dayton, 1997), and was a more efficient use of data than PS matching (because no patients were excluded). To account for the competing risk of each outcome with death, I plotted graphs of cumulative incidence function, adjusted for PS and age, following competing-risks regression (Fine and Gray, 1999, Noordzij et al., 2013). I conducted sensitivity analyses where individuals who did not receive blood tests or weight measurements were not dropped from the cohort, and where individuals were assigned inverse probability weights (IPTW) based on how likely they were to have blood test or weight records (Seaman and White, 2013). I used multiple demographic and clinical variables to predict missingness for the IPTW model.

7.4 Results

Of the 14,396 individuals with a diagnosis of BPD, 6671 were potentially included in the analysis; 2148 prescribed lithium, 1670 prescribed valproate, 1477 prescribed olanzapine and 1376 prescribed quetiapine (Figure 1.i).
The characteristics of the potentially included cohort are shown in Table 1.ii. Drug exposure ranged from 28 days to 17 years 11 days. People prescribed lithium tended to be older than those taking other study drugs, with more years of follow-up data. These individuals were less likely to have records of depression or anxiety prior to entry into the cohort. Individuals prescribed lithium had no more contacts with primary care services during follow-up than individuals prescribed other drugs. The number of individuals included for each outcome by treatment group is shown in Table 7.ii.
Table 7.i Patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, N</td>
<td>2148</td>
<td>1670</td>
<td>1477</td>
<td>1376</td>
</tr>
<tr>
<td>Female, N (%)</td>
<td>1287 (59.92)</td>
<td>911 (54.55)</td>
<td>791 (53.55)</td>
<td>959 (69.69)</td>
</tr>
<tr>
<td>Age, median (IQR), years</td>
<td>46.28 (35.70–60.67)</td>
<td>42.31 (31.95–54.80)</td>
<td>41.01 (32.03–53.08)</td>
<td>38.08 (29.30–48.71)</td>
</tr>
<tr>
<td>Non-white ethnic background, N (%)</td>
<td>55 (2.56)</td>
<td>85 (5.09)</td>
<td>78 (5.28)</td>
<td>43 (3.13)</td>
</tr>
<tr>
<td>Follow-up, median (IQR), years</td>
<td>2.03 (0.77–4.86)</td>
<td>1.48 (0.65–3.35)</td>
<td>1.28 (0.59–3.29)</td>
<td>1.06 (0.56–2.26)</td>
</tr>
<tr>
<td>Primary care contacts per year, median (IQR)</td>
<td>11.14 (6.54–18.02)</td>
<td>12.51 (7.36–19.96)</td>
<td>11.94 (7.08–19.55)</td>
<td>14.61 (9.21–22.55)</td>
</tr>
<tr>
<td>Health at baseline, N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD history</td>
<td>124 (5.77)</td>
<td>121 (7.25)</td>
<td>68 (4.60)</td>
<td>53 (3.85)</td>
</tr>
<tr>
<td>≥CKD3 (or eGFR&lt;60 ml/min/1.73m²)</td>
<td>52 (2.42)</td>
<td>40 (2.40)</td>
<td>27 (1.83)</td>
<td>32 (2.33)</td>
</tr>
<tr>
<td>Hypothyroidism (or TSH&gt;10mU/L)</td>
<td>183 (8.52)</td>
<td>105 (6.29)</td>
<td>60 (4.06)</td>
<td>61 (4.43)</td>
</tr>
<tr>
<td>Hyperthyroidism (or TSH&lt;0.1mU/L)</td>
<td>16 (0.74)</td>
<td>8 (0.48)</td>
<td>9 (0.61)</td>
<td>9 (0.66)</td>
</tr>
<tr>
<td>T2DM (or HBA1c&gt;48mmol/mol)</td>
<td>108 (5.03)</td>
<td>140 (8.38)</td>
<td>45 (3.05)</td>
<td>86 (6.25)</td>
</tr>
<tr>
<td>Hepatic impairment (or ALT&gt;200U/L or AST&gt;250U/L)</td>
<td>34 (1.58)</td>
<td>41 (2.45)</td>
<td>37 (2.51)</td>
<td>19 (1.38)</td>
</tr>
<tr>
<td>Obesity (BMI&gt;30)</td>
<td>896 (41.71)</td>
<td>716 (42.87)</td>
<td>509 (34.36)</td>
<td>609 (44.26)</td>
</tr>
<tr>
<td>Hypercalcemia (adjusted calcium&gt;2.65mmol/L)</td>
<td>10 (0.47)</td>
<td>4 (0.24)</td>
<td>2 (0.14)</td>
<td>3 (0.22)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>184 (8.57)</td>
<td>173 (10.36)</td>
<td>103 (6.97)</td>
<td>130 (9.45)</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>43 (2.00)</td>
<td>132 (7.90)</td>
<td>50 (3.39)</td>
<td>49 (3.56)</td>
</tr>
<tr>
<td>Previous anxiety problems</td>
<td>144 (6.70)</td>
<td>150 (8.98)</td>
<td>137 (9.28)</td>
<td>201 (14.61)</td>
</tr>
<tr>
<td>Moderate/heavy alcohol use</td>
<td>1189 (57.00)</td>
<td>899 (54.75)</td>
<td>791 (53.55)</td>
<td>708 (52.21)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>711 (33.10)</td>
<td>652 (39.04)</td>
<td>632 (42.79)</td>
<td>567 (41.21)</td>
</tr>
<tr>
<td>Bipolar disorder characteristics at baseline, N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous depressive episode</td>
<td>1238 (57.64)</td>
<td>990 (59.28)</td>
<td>915 (61.95)</td>
<td>1015 (73.76)</td>
</tr>
<tr>
<td>Previous record of taking study drug</td>
<td>1731 (80.59)</td>
<td>1157 (69.28)</td>
<td>886 (59.99)</td>
<td>847 (61.56)</td>
</tr>
</tbody>
</table>

CVD cardiovascular disease; CKD chronic kidney disease; eGFR estimated glomerular filtration rate; TSH thyroid stimulating hormone; T2DM type 2 diabetes mellitus; ALT alanine transaminase; AST aspartate aminotransferase; BMI body mass index

In unadjusted analysis and after adjustment for PS, age, calendar year and clustering by practice in which the primary care physician worked, rates of CKD stage 3 or above in individuals prescribed valproate (HR 0.56; 95% CI 0.45 to 0.69; P<0.001) olanzapine (HR 0.57; 95% CI 0.45 to 0.71; P<0.001) or quetiapine (HR 0.62; 95% CI 0.47 to 0.80; P<0.001) were reduced compared to lithium (Table 7.iii, Figure 7.7.ii).
<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total potentially included, N</strong></td>
<td>2148</td>
<td>1670</td>
<td>1477</td>
<td>1376</td>
</tr>
<tr>
<td>≥CKD stage 3</td>
<td>1541 (71.74)</td>
<td>1116 (66.83)</td>
<td>964 (65.27)</td>
<td>939 (68.24)</td>
</tr>
<tr>
<td>≥CKD stage 4</td>
<td>1642 (76.44)</td>
<td>1176 (70.42)</td>
<td>1016 (68.79)</td>
<td>983 (71.44)</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>1620 (74.95)</td>
<td>916 (54.85)</td>
<td>832 (56.33)</td>
<td>735 (53.42)</td>
</tr>
<tr>
<td>Hypercalcemia</td>
<td>1398 (65.08)</td>
<td>844 (50.54)</td>
<td>775 (52.47)</td>
<td>687 (49.93)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1964 (91.43)</td>
<td>1497 (89.64)</td>
<td>1374 (93.03)</td>
<td>1246 (90.55)</td>
</tr>
<tr>
<td>Hepatotoxicity</td>
<td>1171 (54.52)</td>
<td>852 (51.02)</td>
<td>718 (48.61)</td>
<td>611 (44.40)</td>
</tr>
</tbody>
</table>

CKD, chronic kidney disease; T2DM, type 2 diabetes mellitus; CVD, cardiovascular disease.

### Table 7.iii Adverse effects during maintenance treatment

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥CKD stage 3 (N=4560)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>489</td>
<td>130</td>
<td>121</td>
<td>71</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>51.97</td>
<td>29.85</td>
<td>25.64</td>
<td>16.89</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>9.41 (8.61–10.28)</td>
<td>4.35 (3.67–5.17)</td>
<td>4.72 (3.95–5.64)</td>
<td>4.20 (3.33–5.31)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.46 (0.38–0.55)</td>
<td>0.50 (0.41–0.61)</td>
<td>0.43 (0.33–0.55)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.56 (0.45–0.69)</td>
<td>0.57 (0.45–0.71)</td>
<td>0.62</td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>≥CKD stage 4 (N=4817)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>91</td>
<td>34</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>63.48</td>
<td>32.75</td>
<td>27.77</td>
<td>18.35</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>1.43 (1.17–1.76)</td>
<td>1.04 (0.74–1.45)</td>
<td>0.72 (0.46–1.12)</td>
<td>0.65 (0.37–1.15)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.75 (0.51–1.13)</td>
<td>0.52 (0.32–0.85)</td>
<td>0.47</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.94 (0.59–1.50)</td>
<td>0.65 (0.37–1.12)</td>
<td>0.67</td>
</tr>
<tr>
<td>P-value</td>
<td>0.806</td>
<td>0.127</td>
<td>0.273</td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism (N=4093)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>183</td>
<td>61</td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>59.23</td>
<td>27.78</td>
<td>23.79</td>
<td>15.59</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>3.09 (2.67–3.57)</td>
<td>2.20 (1.71–2.82)</td>
<td>1.72 (1.27–2.34)</td>
<td>2.12 (1.50–2.98)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.69 (0.51–0.94)</td>
<td>0.54 (0.38–0.77)</td>
<td>0.62</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.60 (0.40–0.89)</td>
<td>0.48 (0.29–0.77)</td>
<td>0.63</td>
</tr>
<tr>
<td>P-value</td>
<td>0.012</td>
<td>0.003</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>Hyperthyroidism (N=3704)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>41</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>52.49</td>
<td>25.81</td>
<td>22.62</td>
<td>14.65</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>0.78 (0.58–1.06)</td>
<td>0.19 (0.08–0.47)</td>
<td>0.27 (0.12–0.59)</td>
<td>0.41 (0.18–0.91)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.24 (0.09–0.61)</td>
<td>0.33 (0.14–0.78)</td>
<td>0.48 (0.20–1.17)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.24 (0.09–0.61)</td>
<td>0.31 (0.13–0.73)</td>
<td>0.45 (0.18–1.18)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.003</td>
<td>0.007</td>
<td>0.096</td>
<td></td>
</tr>
</tbody>
</table>

Continued overleaf
<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypercalcemia (N=2094)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>55</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>36.29</td>
<td>17.1</td>
<td>13.44</td>
<td>8.99</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>1.52</td>
<td>0.35</td>
<td>0.45</td>
<td>0.33</td>
</tr>
<tr>
<td>(1.16–1.97)</td>
<td>(0.16–0.78)</td>
<td>(0.20–0.99)</td>
<td>(0.11–1.03)</td>
<td></td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.24</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(0.10–0.56)</td>
<td>(0.14–0.68)</td>
<td>(0.07–0.76)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.25</td>
<td>0.32</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.10–0.60)</td>
<td>(0.14–0.76)</td>
<td>(0.07–0.73)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.002</td>
<td>0.008</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td><strong>T2DM (N=6292)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>150</td>
<td>86</td>
<td>88</td>
<td>51</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>69.11</td>
<td>36.52</td>
<td>34.01</td>
<td>21.28</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>2.17</td>
<td>2.35</td>
<td>2.59</td>
<td>2.40</td>
</tr>
<tr>
<td>(1.85–2.55)</td>
<td>(1.91–2.91)</td>
<td>(2.10–3.19)</td>
<td>(1.82–3.15)</td>
<td></td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.21</td>
<td>1.31</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>(0.94–1.55)</td>
<td>(0.99–1.74)</td>
<td>(0.95–1.82)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.08</td>
<td>1.20</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(0.83–1.42)</td>
<td>(0.89–1.61)</td>
<td>(0.65–1.35)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.586</td>
<td>0.230</td>
<td>0.752</td>
<td></td>
</tr>
<tr>
<td><strong>CVD (N=6305)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>94</td>
<td>32</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>69.49</td>
<td>37.56</td>
<td>33.41</td>
<td>21.17</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>1.35</td>
<td>0.85</td>
<td>0.78</td>
<td>0.95</td>
</tr>
<tr>
<td>(1.11–1.66)</td>
<td>(0.60–1.20)</td>
<td>(0.53–1.14)</td>
<td>(0.62–1.45)</td>
<td></td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.67</td>
<td>0.61</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>(0.44–1.04)</td>
<td>(0.38–0.98)</td>
<td>(0.47–1.35)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.91</td>
<td>0.85</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>(0.59–1.41)</td>
<td>(0.53–1.36)</td>
<td>(0.63–1.96)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.684</td>
<td>0.509</td>
<td>0.732</td>
<td></td>
</tr>
<tr>
<td><strong>&gt;7% weight gain (N=4458)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>467</td>
<td>410</td>
<td>396</td>
<td>299</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>63.20</td>
<td>33.28</td>
<td>30.56</td>
<td>19.05</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>7.39</td>
<td>12.32</td>
<td>12.96</td>
<td>15.70</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.90</td>
<td>1.99</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>(1.64–2.20)</td>
<td>(1.72–2.30)</td>
<td>(2.33–3.16)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.37</td>
<td>1.43</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>(1.17–1.61)</td>
<td>(1.23–1.67)</td>
<td>(1.16–1.62)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td><strong>&gt;15% weight gain (N=4458)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>179</td>
<td>182</td>
<td>189</td>
<td>130</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>63.92</td>
<td>34.27</td>
<td>31.24</td>
<td>19.30</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>2.80</td>
<td>5.31</td>
<td>6.05</td>
<td>6.74</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>2.29</td>
<td>2.57</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>(1.87–2.82)</td>
<td>(2.06–3.22)</td>
<td>(2.61–4.44)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.62</td>
<td>1.84</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>(1.31–2.01)</td>
<td>(1.47–2.30)</td>
<td>(1.24–2.20)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

Continued overleaf
### Hypertension (N=6081)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, N</td>
<td>174</td>
<td>85</td>
<td>89</td>
<td>33</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>64.25</td>
<td>35.66</td>
<td>32.30</td>
<td>20.65</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>2.71</td>
<td>2.38</td>
<td>2.76</td>
<td>1.60</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.98</td>
<td>1.11</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(0.75–1.26)</td>
<td>(0.84–1.46)</td>
<td>(0.50–1.12)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.19</td>
<td>1.41</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>(0.90–1.58)</td>
<td>(1.06–1.87)</td>
<td>(0.59–1.34)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.274</td>
<td>0.017</td>
<td>0.590</td>
<td></td>
</tr>
</tbody>
</table>

### Hepatotoxicity (N=3352)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, N</td>
<td>20</td>
<td>10</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>50.13</td>
<td>25.95</td>
<td>20.39</td>
<td>12.53</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td>0.40</td>
<td>0.39</td>
<td>0.69</td>
<td>1.04</td>
</tr>
<tr>
<td>(0.26–0.62)</td>
<td>(0.21–0.72)</td>
<td>(0.41–1.16)</td>
<td>(0.60–1.79)</td>
<td></td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.96</td>
<td>1.71</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>(0.45–2.07)</td>
<td>(0.86–3.40)</td>
<td>(1.26–5.32)</td>
<td></td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.65</td>
<td>1.23</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>(0.30–1.39)</td>
<td>(0.63–2.42)</td>
<td>(0.54–2.74)</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.274</td>
<td>0.558</td>
<td>0.658</td>
<td></td>
</tr>
</tbody>
</table>

CKD, chronic kidney disease; T2DM, type 2 diabetes mellitus; CVD, cardiovascular disease; PYAR, person-years at risk; HR, hazard ratio; PS, propensity score. Unadjusted hazard ratio accounts for clustering by primary care practice, adjusted hazard ratio is adjusted for propensity score age group and calendar period time varying variables and clustering by primary care practice. P-values for PS adjusted HR.

**Figure 7.7.ii** Cumulative incidence estimates of adverse renal and hepatic event rates.
From PS and age adjusted competing-risks regression. Note differences in scale of y-axis for each plot.
Compared to lithium, rates of hypothyroidism were reduced in those prescribed valproate (HR 0.60; 95% CI 0.40 to 0.89; P=0.012), olanzapine (HR 0.48; 95% CI 0.29 to 0.77; P=0.003), but not quetiapine (HR 0.63; 95% CI 0.38 to 1.05; P=0.074) after adjustment. Rates of hyperthyroidism were lower in those prescribed valproate (HR 0.24; 95% CI 0.09 to 0.61; 0.003) and olanzapine (HR 0.31; 95% CI 0.13 to 0.73; 0.007), but not quetiapine (HR 0.45; 95% CI 0.18 to 1.18; 0.096) compared to lithium. Hypercalcemia was less common in those prescribed valproate (HR 0.25; 95% CI 0.10 to 0.60; P=0.002), olanzapine (HR 0.32; 95% CI 0.14 to 0.76; P=0.008), or quetiapine (HR 0.23; 95% CI 0.07 to 0.73; P=0.013) compared to lithium (Table 7.iii, Figure 7.7.iii).

Figure 7.7.iii Cumulative incidence estimates of adverse endocrine event rates
From PS and age adjusted competing-risks regression. Note differences in scale of y-axis for each plot.
After adjustment, rates of weight gain were higher in valproate, olanzapine and quetiapine than lithium (>15% weight gain: valproate HR 1.62; 95% CI 1.31 to 2.01; P<0.001, olanzapine HR 1.84; 95% CI 1.47 to 2.30; P<0.001, quetiapine HR 1.67; 95% CI 1.24 to 2.20; P<0.001). Rates of hypertension were higher in olanzapine (HR 1.41; 95% CI 1.06 to 1.87; P=0.017) than lithium (Table 7.iii, Figure 7.7.iv).

Figure 7.7.iv Cumulative incidence estimates of adverse metabolic event rates

Continued overleaf
I found no significant difference in rates of CKD stage 4 or above, T2DM, CVD, or hepatotoxicity between groups (Table 7.iii). The median number of eGFR/creatinine and TSH blood tests per year in treatment was higher in those taking lithium than the other
drugs (Table 7.iv). Weight measurement and blood tests for adjusted calcium and ALT/AST were less frequent in patients prescribed lithium, but not significantly so (Table 7.iv).

Table 7.iv Median number (and interquartile range) of tests per year of drug exposure in patients included in analyses

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFR or creatinine</td>
<td>1.84 (1.04–3.03)</td>
<td>1.10 (0.62–1.88)</td>
<td>1.00 (0.56–1.78)</td>
<td>1.20 (0.69–2.01)</td>
</tr>
<tr>
<td>TSH</td>
<td>1.64 (1.05–2.47)</td>
<td>0.85 (0.49–1.46)</td>
<td>0.80 (0.45–1.50)</td>
<td>1.04 (0.62–1.70)</td>
</tr>
<tr>
<td>Adjusted calcium</td>
<td>0.91 (0.38–2.18)</td>
<td>0.90 (0.41–2.12)</td>
<td>1.10 (0.43–2.99)</td>
<td>1.37 (0.62–2.99)</td>
</tr>
<tr>
<td>ALT or AST</td>
<td>0.75 (0.36–1.51)</td>
<td>0.81 (0.40–1.57)</td>
<td>0.73 (0.37–1.40)</td>
<td>0.94 (0.49–1.70)</td>
</tr>
<tr>
<td>Weight</td>
<td>0.98 (0.53–1.73)</td>
<td>1.19 (0.73–2.07)</td>
<td>1.20 (0.63–2.01)</td>
<td>1.44 (0.89–2.44)</td>
</tr>
</tbody>
</table>

eGFR, estimated glomerular filtration rate; TSH, thyroid stimulating hormone; ALT, alanine transaminase; AST, aspartate aminotransferase. Type 2 diabetes mellitus, cardiovascular disease and hypertension were not defined by tests.

For outcomes in which patients had been excluded because of missing tests (CKD, hypothyroidism, hypercalcemia, weight gain and hepatotoxicity) sensitivity analyses including all patients resulted in reduced rate estimates compared to the primary analyses, but had little effect on HRs (Table 7.v). Sensitivity analyses using IPTW suggest results from the primary analyses are robust (Table 7.v). From Schoenfeld residuals, there was no evidence against the assumption of proportional hazards for any outcome.

Table 7.v Sensitivity analyses to account for missing blood tests by 1) including all individuals and 2) performing inverse probability weighting

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥CKD stage 3 (N=6520)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>489</td>
<td>130</td>
<td>121</td>
<td>71</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>60.65</td>
<td>36.62</td>
<td>32.56</td>
<td>21.33</td>
</tr>
<tr>
<td>Rate, per 100 PYAR (95% CI)</td>
<td>8.06 (7.38–8.81)</td>
<td>3.55 (2.99–4.22)</td>
<td>3.72 (3.11–4.44)</td>
<td>3.33 (2.64–4.20)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>0.54 (0.44–0.67)</td>
<td>0.55 (0.44–0.70)</td>
<td>0.55 (0.42–0.72)</td>
<td>0.55 (0.32–0.53)</td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR (95% CI)</td>
<td>0.63 (0.51–0.78)</td>
<td>0.67 (0.53–0.83)</td>
<td>0.70 (0.52–0.93)</td>
<td>0.16 (0.09–0.23)</td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Continued overleaf
# Lithium, Valproate, Olanzapine, Quetiapine

<table>
<thead>
<tr>
<th>Event</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6600</td>
<td>6600</td>
<td>6600</td>
<td>6600</td>
</tr>
<tr>
<td>Rate (per 100 PYAR)</td>
<td>1.25 (1.03–1.56)</td>
<td>0.86 (0.61–1.20)</td>
<td>0.58 (0.37–0.89)</td>
<td>0.53 (0.30–0.93)</td>
</tr>
<tr>
<td>Unadjusted HR</td>
<td>0.73 (0.49–1.08)</td>
<td>0.49 (0.30–0.79)</td>
<td>0.45 (0.24–0.83)</td>
<td>0.45 (0.24–0.83)</td>
</tr>
<tr>
<td>PS Adjusted HR</td>
<td>0.90 (0.57–1.43)</td>
<td>0.61 (0.36–1.05)</td>
<td>0.58 (0.28–1.18)</td>
<td>0.12 (0.01–0.23)</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>1.04 (0.64–1.68)</td>
<td>0.70 (0.40–1.21)</td>
<td>0.77 (0.36–1.64)</td>
<td>0.59 (0.30–0.99)</td>
</tr>
</tbody>
</table>

# Hypothyroidism (N=6262)

<table>
<thead>
<tr>
<th>Event</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6262</td>
<td>6262</td>
<td>6262</td>
<td>6262</td>
</tr>
<tr>
<td>Rate (per 100 PYAR)</td>
<td>2.87 (2.49–3.32)</td>
<td>1.64 (1.28–2.10)</td>
<td>1.22 (0.90–1.66)</td>
<td>1.50 (1.06–2.11)</td>
</tr>
<tr>
<td>Unadjusted HR</td>
<td>0.56 (0.41–0.76)</td>
<td>0.41 (0.29–0.59)</td>
<td>0.48 (0.33–0.70)</td>
<td>0.48 (0.33–0.70)</td>
</tr>
<tr>
<td>PS Adjusted HR</td>
<td>0.59 (0.42–0.84)</td>
<td>0.43 (0.29–0.63)</td>
<td>0.47 (0.31–0.73)</td>
<td>0.47 (0.31–0.73)</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>0.54 (0.38–0.78)</td>
<td>0.42 (0.28–0.63)</td>
<td>0.49 (0.32–0.78)</td>
<td>0.49 (0.32–0.78)</td>
</tr>
</tbody>
</table>

# Hyperthyroidism (N=6220)

<table>
<thead>
<tr>
<th>Event</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6220</td>
<td>6220</td>
<td>6220</td>
<td>6220</td>
</tr>
<tr>
<td>Rate (per 100 PYAR)</td>
<td>0.72 (0.53–0.98)</td>
<td>0.14 (0.06–0.34)</td>
<td>0.19 (0.08–0.41)</td>
<td>0.28 (0.13–0.62)</td>
</tr>
<tr>
<td>Unadjusted HR</td>
<td>0.19 (0.07–0.49)</td>
<td>0.25 (0.11–0.59)</td>
<td>0.37 (0.15–0.91)</td>
<td>0.37 (0.15–0.91)</td>
</tr>
<tr>
<td>PS Adjusted HR</td>
<td>0.19 (0.07–0.51)</td>
<td>0.25 (0.11–0.59)</td>
<td>0.34 (0.13–0.90)</td>
<td>0.34 (0.13–0.90)</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>0.20 (0.08–0.53)</td>
<td>0.28 (0.12–0.67)</td>
<td>0.34 (0.14–0.99)</td>
<td>0.34 (0.14–0.99)</td>
</tr>
</tbody>
</table>

# Hypercalcemia (N=6652)

<table>
<thead>
<tr>
<th>Event</th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6652</td>
<td>6652</td>
<td>6652</td>
<td>6652</td>
</tr>
<tr>
<td>Rate (per 100 PYAR)</td>
<td>0.76 (0.58–0.99)</td>
<td>0.15 (0.07–0.33)</td>
<td>0.17 (0.08–0.38)</td>
<td>0.13 (0.04–0.40)</td>
</tr>
<tr>
<td>Unadjusted HR</td>
<td>0.22 (0.09–0.51)</td>
<td>0.24 (0.11–0.54)</td>
<td>0.20 (0.06–0.64)</td>
<td>0.20 (0.06–0.64)</td>
</tr>
<tr>
<td>PS Adjusted HR</td>
<td>0.23 (0.09–0.56)</td>
<td>0.28 (0.12–0.66)</td>
<td>0.21 (0.06–0.68)</td>
<td>0.21 (0.06–0.68)</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>0.22 (0.09–0.57)</td>
<td>0.24 (0.10–0.55)</td>
<td>0.22 (0.06–0.76)</td>
<td>0.22 (0.06–0.76)</td>
</tr>
<tr>
<td></td>
<td>Lithium</td>
<td>Valproate</td>
<td>Olanzapine</td>
<td>Quetiapine</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>&gt;7% weight gain (N=6671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>467</td>
<td>410</td>
<td>396</td>
<td>299</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>73.36</td>
<td>39.54</td>
<td>35.96</td>
<td>23.09</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>6.35</td>
<td>(5.80–6.96)</td>
<td>11.01</td>
<td>(11.52–14.46)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.91</td>
<td>(1.65–2.21)</td>
<td>2.01</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>2.77</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.37</td>
<td>(1.17–1.61)</td>
<td>1.46</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1.28</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>1 [reference]</td>
<td>1.23</td>
<td>(1.08–1.41)</td>
<td>1.30</td>
</tr>
<tr>
<td>(95% CI)</td>
<td></td>
<td>0.002</td>
<td>&lt;0.001</td>
<td>1.52</td>
</tr>
<tr>
<td><strong>&gt;15% weight gain (N=6671)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>179</td>
<td>182</td>
<td>189</td>
<td>130</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>73.95</td>
<td>40.53</td>
<td>36.65</td>
<td>23.35</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>2.43</td>
<td>(2.10–2.82)</td>
<td>5.16</td>
<td>(4.69–6.61)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>2.28</td>
<td>(1.85–2.81)</td>
<td>2.57</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>3.44</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.63</td>
<td>(1.31–2.02)</td>
<td>1.86</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1.54</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>1 [reference]</td>
<td>1.35</td>
<td>(1.15–1.58)</td>
<td>1.44</td>
</tr>
<tr>
<td>(95% CI)</td>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1.62</td>
</tr>
<tr>
<td><strong>Hepatotoxicity (N=6540)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N</td>
<td>20</td>
<td>10</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>PYAR (100s)</td>
<td>72.08</td>
<td>39.57</td>
<td>34.13</td>
<td>22.86</td>
</tr>
<tr>
<td>Rate, per 100 PYAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(95% CI)</td>
<td>0.28</td>
<td>(0.18–0.43)</td>
<td>0.41</td>
<td>(0.30–0.69)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.94</td>
<td>(0.44–2.02)</td>
<td>1.53</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>0.883</td>
<td>0.229</td>
<td>2.22</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>0.63</td>
<td>(0.30–1.34)</td>
<td>1.05</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>0.228</td>
<td>0.897</td>
<td>0.98</td>
</tr>
<tr>
<td>IPTW PS Adjusted HR</td>
<td>1 [reference]</td>
<td>0.58</td>
<td>(0.27–1.26)</td>
<td>0.97</td>
</tr>
<tr>
<td>(95% CI)</td>
<td></td>
<td>0.166</td>
<td>0.938</td>
<td>1.12</td>
</tr>
</tbody>
</table>

CKD, chronic kidney disease; PYAR, person-years at risk; HR, hazard ratio; PS, propensity score; IPTW, inverse probability weighted; outcomes for cardiovascular disease, type 2 diabetes mellitus and hypertension had no missing data. N for each outcome varies because of potential diagnoses pre-baseline. Unadjusted HR accounts for clustering by primary care practice. PS adjusted HR is adjusted for propensity score, age group and calendar period time varying variables and clustering by primary care practice, IPTW PS adjusted HR accounts for probability of being a complete record, adjusted for propensity score, age group and calendar period time varying variables and clustering by primary care practice.
7.5 Discussion

7.5.1 Main findings
In a large dataset of nearly 7000 individuals treated for BPD with lithium, valproate, olanzapine or quetiapine, with follow-up times of up to 17 years, I found differential rates of a number of adverse events. Those prescribed lithium were significantly more likely to have a decline in renal function and develop hypo- or hyperthyroidism, and hypercalcemia. However, they were less likely to gain significant weight. Individuals prescribed olanzapine had the highest rate of weight gain and new onset hypertension. I did not find any differences in the rate of new T2DM, CVD or hepatotoxicity across drug treatment groups.

7.5.2 Comparison with previous literature
Severe CKD (stage 4 or above) was uncommon in the cohort (approximately 1 in 100 PYAR) and I did not find differences by drug treatment, but less severe CKD (stage 3 or above) occurred most frequently in patients prescribed lithium. Whilst many of these patients (i.e., those with CKD stage 3) would not progress to a clinically relevant decline in renal function, a number of them would be at increased risk of doing so. It remains unclear if this result is due to 1) lack of power to determine a true difference in rates of severe CKD, 2) surveillance bias due to increased monitoring of renal function in those taking lithium, which would lead to apparent increased rates of asymptomatic CKD stage 3, or 3) lithium treatment truly increasing the risk of reduced renal function without increasing severe CKD risk. Previous studies have found similar results and have not been able to account for this potential bias (Aiff et al., 2014a, Close et al., 2014, Kessing et al., 2015, Shine et al., 2015)[12-14,56]. Clos et al. found no decline in eGFR in individuals taking lithium, using a similar active comparator design, but were also limited by potential ascertainment bias (Clos et al., 2015).
Rates of both hypothyroidism and hyperthyroidism were increased in individuals prescribed lithium compared to valproate and olanzapine (but not quetiapine). Increased hypothyroidism has been shown previously (Kibirige et al., 2013, McKnight et al., 2012) but literature on the association between lithium and hyperthyroidism is inconsistent (Shine et al., 2015) and lithium induced hyperthyroidism is considered rare (Lazarus, 2009). Monitoring thyroid dysfunction in BPD is vital because of evidence that abnormalities are associated with longer time to remission and more symptoms during the maintenance period (Fagiolini et al., 2006). Thyroid function potentially normalises on cessation of lithium, but only one study has investigated this (Souza et al., 1991). Hypercalcemia is also recognised to be associated with lithium prescribing (Khandwala and Uum, 2006, Lehmann and Lee, 2013, McKnight et al., 2012, Shine et al., 2015). Calcium monitoring in patients prescribed lithium was rare in my representative sample of primary care (37% had one or more calcium blood test result), despite it being recommended in the 2006 NICE guidance (National Institute for Health and Care Excellence, 2006).

The rate of individuals gaining more than 7%, and more than 15% of their baseline weight was greater in those prescribed olanzapine, quetiapine or valproate than those prescribed lithium. This degree of weight gain represents a significant risk factor for a number of adverse physical health outcomes including CVD and T2DM (Manu et al., 2015). I may not have captured increased rates of CVD or T2DM because of the relatively brief median follow-up time, in relation to the time taken to develop these diseases. Olanzapine had the highest adjusted rate of greater than 15% weight gain compared to lithium, and the highest rate of new onset hypertension. This has been shown previously in comparisons of antipsychotic drugs (Newcomer, 2005), and in trials of olanzapine versus lithium or valproate (Nashed et al., 2011).
Hepatotoxicity was rare in the cohort, and before PS adjustment rates appeared to be elevated in the quetiapine group, compared to lithium. This association has been identified previously (Atasoy et al., 2007). After PS adjustment, there was no evidence of between group differences.

7.5.3 Strengths and limitations
As with the study of comparative effectiveness and tolerability (Study 2) the major strength of this study, beyond size and length of follow-up, is the direct comparison between BPD maintenance mood stabiliser treatment options for a number of adverse effects. The use of electronic health records also means it is possible to adjust for a number of demographic and physical health characteristics that may have influenced the clinician’s decision to treat with a particular medication or potentially confound the relationship between treatment and adverse outcome. Despite including numerous variables in the PS, it is possible that residual confounding remained, especially as those prescribed lithium were older and were more likely to have taken the drug previously, perhaps reflecting a more chronic illness course. It may be that important patient or clinician features were not captured by the score, and despite the balance of observed covariates I cannot confirm balance of unobserved covariates (Austin et al., 2005, Stukel et al., 2007). I was also unable to consider dosage differences across the different treatment groups in this analysis. Periods of lithium toxicity may be particularly important with regards to developing renal failure and I was unable to capture this information from the available data. Missing data can be a problem in studies utilising EHRs, especially as there may be a clinical reason why information is missing. Because of the way outcomes were defined, T2DM, CVD and diagnoses of hypertension had no missing data, and no covariates in the PS had missing values.
Patients prescribed lithium had no more physician contacts than those taking other mood stabiliser medication. In individuals that ever received tests during treatment exposure, testing frequency was similar in all study drugs for adjusted calcium, liver function and weight (*Table 7.iv*). Frequency of testing renal and thyroid function was higher in those taking lithium, which reflects the guidance for monitoring (National Institute for Health and Care Excellence, 2006). Patients prescribed lithium were also more likely to have at least one renal function, thyroid function, calcium or liver function test compared to patients taking other drugs. This is likely to be due to both drug related indications for monitoring, and the longer drug exposure seen in those taking lithium. IPTW sensitivity analysis to account for this difference did not alter my conclusions (*Table 7.v*). In the primary analysis the likely effect of this differential missingness would be to reduce the HRs for lithium compared to the other drugs, relative to their true values, as blood tests in the non-lithium group are more likely to be related to clinical symptoms than monitoring guidance (for instance, this is likely to represent an underestimation of the true hypercalcemia HR for lithium versus other drugs). The median number of weight measurements was similar in each group suggesting detection of weight gain was not related to differential monitoring. The sensitivity analyses including individual’s irrespective of blood tests produced similar adjusted HRs to the primary analyses for each outcome, but often with reduced incidence of the outcome in each treatment group (*Table 7.v*). These analyses may more accurately reflect testing occurring because of clinical indication. Further limitations that are general across all studies in the thesis are discussed in *Chapter 9*.

### 7.5.4 Conclusions

Lithium remains an important treatment option for individuals with BPD. However there is clear evidence that that its use is associated with a number of adverse events. These risks need to be offset with the potentially superior effectiveness and anti-suicidal benefits of
the drug compared to other treatment options (Cipriani et al., 2013a, Miura et al., 2014, Severus et al., 2014). It is also true that other recommended maintenance treatments can have serious side effects, often related to weight gain, and are not suitable for use in certain patient groups (such as the contraindication of valproate in women of childbearing potential (National Institute for Health and Care Excellence, 2014)).

7.5.5 Implications of the findings of this study
Assiduous monitoring of patients prescribed lithium should ameliorate some risk associated with effects on renal physiology and endocrine systems. Given the need to balance an array of risks and benefits, an individualised and collaborative approach to treatment choice is likely to be most appropriate. To achieve this, further research identifying patient characteristics that are risk factors for specific side effects and an understanding of the risks and benefits of stopping treatment in those who experience adverse effects is necessary. Implications for policy and practice are discussed comprehensively in Chapter 9. As a next step (in Study 4; Chapter 8), I go on to explore whether there are indeed anti-suicidal effects of lithium, which may offset some of the physical problems quantified in this chapter.
Chapter 8  Comparison of self-harm, accidental injury and suicide in individuals with bipolar disorder during treatment with different maintenance mood stabiliser medications for bipolar disorder

8.1 Summary

8.1.1 Objective

To calculate and compare rates of self-harm, unintentional injury and suicide in individuals prescribed lithium, valproate, olanzapine and quetiapine, while accounting for propensity to be prescribed one of these mood stabilisers.

8.1.2 Methods

I conducted a propensity score adjusted and matched longitudinal cohort study in a nationally representative United Kingdom sample of electronic health record data collected January 1 1995 to December 31 2013. All patients diagnosed with BPD prescribed lithium (N=2148), valproate (N=1670), olanzapine (N=1477), or quetiapine (N=1376) as maintenance mood stabiliser treatment were included. The primary outcome was any record of self-harm. Secondary outcomes were accidental injury and suicide.

8.1.3 Results

Self-harm rates were lower in patients prescribed lithium (205 per 10,000 PYAR; 95% CI 175 to 241) compared with those prescribed valproate (392 per 10,000 PYAR; 95% CI 334 to 460), olanzapine (409 per 10,000 PYAR; 95% CI 345 to 483) or quetiapine (582 per 10,000...
PYAR; 95% CI 489 to 692). This relationship was maintained after propensity score adjustment (valproate, olanzapine or quetiapine versus lithium; HR 1.40; 95% CI 1.12 to 1.74) and matching (HR 1.51; 95% CI 1.21 to 1.88). After propensity score adjustment, accidental injury rates were lower in lithium compared to valproate (HR 1.32; 95% CI 1.10 to 1.58) and quetiapine (HR 1.34; 95% CI 1.07 to 1.69), but not olanzapine. The suicide rate in the cohort was 14 per 10,000 PYAR (95% CI 9 to 21). Although this was lower in the lithium group than for other treatments, there were too few events to allow statistical comparison.

8.1.4 Conclusions
Patients taking lithium had reduced self-harm and accidental injury rates. This finding augments limited trial and smaller observational study results. It supports the hypothesis that lithium reduces impulsive aggression as well as stabilising mood.

8.2 Introduction

Self-harm is a major cause of morbidity in BPD (Singhal et al., 2014), and drug treatments that reduce suicidal and non-suicidal self-harm could improve quality of life for individuals with BPD and their families (Berghöfer, 2013). My mortality/morbidity cohort study (Study 1: Chapter 4) found rates of self-harm in BPD to be 25 times higher than the general population. Furthermore, individuals who self-harm have a substantially increased suicide risk (Owens et al., 2002). My cohort study found BPD is associated with 13 times the rate of suicide (Chapter 4) and my systematic review demonstrated lifetime risk of suicide almost 15 times greater (Chapter 3). RCTs of maintenance medication show that drugs such as lithium, valproate, olanzapine and quetiapine can stabilise mood compared to placebo (Chapter 5). However, balancing the relative benefits and potential risks (Chapter 7) of these medications is not straightforward; potential drug effects on self-harm have been under-examined in this regard.

As trials often exclude those with a history of suicidal behaviour, drug effects on self-harm have been difficult to quantify due to low event rates (Perlis, 2011). A meta-analysis of 48 trials suggested that suicide was less likely in people prescribed lithium than placebo or active comparator groups, but found no difference in self-harm rates (Cipriani et al., 2013a). Observational studies have suggested that lithium may reduce fatal and non-fatal self-harm compared to maintenance treatment alternatives, most commonly anticonvulsant medication (Baldessarini et al., 2006b, Goodwin et al., 2003, Schou, 1998, Smith et al., 2009, Søndergård et al., 2008), but the findings are not always consistent (Ahearn et al., 2013, Bowden et al., 2000, Marangell et al., 2008). Following a warning from the United States Food and Drug Administration that anticonvulsant medications carry an increased risk of suicidal self-harm (US Food and Drug Administration, 2009), a number of
studies investigated this in BPD. A meta-analysis including only BPD patients (Redden et al., 2011), and several observational studies (Arana et al., 2010, Leon et al., 2014, Reid, 2011) did not replicate this finding. There are relatively sparse data on the association between antipsychotic medication and self-harm. Small retrospective cohorts have shown no difference in suicidal self-harm in patients taking olanzapine or quetiapine (Koek et al., 2012) and higher rates of suicide attempts in those prescribed SGAs compared to lithium or valproate (Ahearn et al., 2013, Yerevanian et al., 2007).

Risk of accidental injury has also been understudied in BPD, despite deaths from accidents being around 6 times higher in BPD than in the general population (Hoang et al., 2011). Though accidents are often recorded in drug trials, they are rarely reported as important outcomes (Matson et al., 2006). Observational studies of drug treatments are even more limited (Elvik, 2013). It has been suggested that accidents are associated with (hypo)manic rather than depressive morbidity (Khalsa et al., 2008), in which case drugs with the strongest anti-accident properties may not be those with the strongest anti-suicidal effects.

Three mechanisms for lithium’s potentially superior anti-suicidal effects have been proposed. Firstly, that lithium reduces risk through reducing depressive relapse, in which case drugs that also protect against depressive relapse should show comparable effects (for example quetiapine) (Cipriani et al., 2013a). Secondly, that there are specific serotonin-mediated effects of lithium that result in reduced aggressive behaviour, risk-taking, and impulsivity (Fawcett, 2001, Kovacsics et al., 2009, Müller-Oerlinghausen and Lewitzka, 2010), in which case one would also expect to see reductions in accidental injury in this group. Thirdly, that the close monitoring of patients taking lithium may provide psychosocial support lacking with other drug treatments, mitigating suicide risk (Tondo and
Baldessarini, 2009), in which case one would expect to see variability in service use across treatment groups.

8.3 Method

8.3.1 Study design, setting and participants
This study used the same cohort as the adverse effects study (Chapter 7).

8.3.2 Exposure
Exposure to each of the study drugs, namely lithium, valproate, olanzapine or quetiapine was defined in the way described in Chapter 7.

8.3.3 Outcomes
The primary outcome of interest was emergency department or primary care attendance for self-harm during the period of drug exposure and the three months afterwards (both outcomes will be included in the EHR). This outcome included Read codes for intentional poisoning, intentional self-injurious behaviour, and self-harm acts of uncertain intent. The positive predictive value of this outcome in THIN has been shown to be 97% (Arana et al., 2010). It was not possible to separate non-suicidal self-harm from self-harm with suicidal intent, or grade the event’s severity. However my unitary categorisation of non-suicidal and suicidal self-harm is consistent with UK research norms (Haw et al., 2015). Secondary outcomes were accidental injury (such as falls or road traffic accidents) presenting to primary or secondary care, and a record of the patient’s suicide during this period, defined in line with previous research (Arana et al., 2010).  

8.3.4 Propensity score estimation
I developed a new PS model based on factors, decided a priori and based on existing research and clinical experience as described in Chapter 6 and Chapter 7. Variables in the
PS differed slightly from those in my previous studies, because of the recommendation that variables in a PS should be associated with the outcome, in this case self-harm (Rubin and Thomas, 1996). Included variables were: sex, age at start of treatment with the study drug, year of entry to the cohort, ethnicity (grouped as White, Black, Asian, mixed, other, with missing values coded as White) (Chisholm, 1990), CVD diagnosis before baseline, hypertension, chronic kidney disease at baseline, history of hypo- or hyperthyroidism, history of liver disease, T2DM, epilepsy, alcohol use (grouped as none or low, moderate or heavy alcohol use, or dependence), history of illicit drug use, smoking status (grouped as never-smoker, ex-smoker, current smoker), BMI (grouped as healthy weight, overweight (BMI 25 to 30), obese (BMI over 30)), anxiety symptoms or diagnosis before baseline, depressive symptoms or diagnosis before baseline, sleep disturbance before baseline, treatment with study drug at or before baseline, history of previous self-harm.

8.3.5 Statistical analysis

A similar analysis to that described in Chapters 6 and 7 was performed. Cox regression analyses were conducted comparing the rates of self-harm, accidental injury and suicide in the four treatment groups. Time to adverse outcome was summarised by Kaplan-Meier curves. Analysis of Schoenfeld residuals was completed to test the assumption of proportional hazards (Schoenfeld, 1982). The PS was calculated by multinomial logistic regression using the covariates described as independent variables and drug treatment as the dependent variable. The PS was used as a linear term in a Cox regression analysis. A one-to-one PS matched analysis was also completed, with each patient in the valproate, olanzapine, or quetiapine group matched to a lithium patient with a 0.01 caliper, dropping all other patients from the analysis. As mentioned in Chapter 2 and Chapter 6, these two approaches to PS analysis have different strengths: the adjusted analysis may be more generalisable and is a more efficient use of the data (as no patients are dropped); the
matched analysis may provide a more valid estimate of treatment effect as only patients with similar observed characteristics are included (d’Agostino, 1998, Rosenbaum and Rubin, 1985). Both adjusted and matched PS models were also adjusted for time-updated variables (age and calendar year) and clustering of patients by primary care practice.

### 8.4 Results

The key features of the cohort are described Section 7.4 and the flow diagram of included patients is shown in Figure 8.i. The characteristics of these patients important to this study are shown in Table 8.i. Of note, individuals prescribed lithium were less likely to have a history of self-harm prior to entry into the cohort. Individuals prescribed lithium had no more contacts with primary care services during follow-up than individuals prescribed other drugs.

---

**Figure 8.i Flow diagram of included patients**
Table 8.1 Patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>lithium</th>
<th>valproate</th>
<th>olanzapine</th>
<th>quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, N</td>
<td>2148</td>
<td>1670</td>
<td>1477</td>
<td>1376</td>
</tr>
<tr>
<td>Female, N (%)</td>
<td>1287 (59.92)</td>
<td>911 (54.55)</td>
<td>791 (53.55)</td>
<td>959 (69.69)</td>
</tr>
<tr>
<td>Age, median (IQR), years</td>
<td>46.28 (35.70-60.67)</td>
<td>42.31 (31.95-54.80)</td>
<td>41.01 (32.03-53.08)</td>
<td>38.08 (29.30-48.71)</td>
</tr>
<tr>
<td>Duration of drug exposure, median (IQR), years</td>
<td>2.03 (0.77-4.86)</td>
<td>1.48 (0.65-3.35)</td>
<td>1.28 (0.59-3.29)</td>
<td>1.06 (0.56-2.26)</td>
</tr>
<tr>
<td>Non-white ethnic background, N (%)</td>
<td>55 (2.56)</td>
<td>85 (5.09)</td>
<td>78 (5.28)</td>
<td>43 (3.13)</td>
</tr>
<tr>
<td>Primary care contacts per year, median (IQR)</td>
<td>11.14 (7.36-19.92)</td>
<td>12.51 (7.36-19.95)</td>
<td>11.94 (7.08-19.55)</td>
<td>14.61 (9.21-22.55)</td>
</tr>
<tr>
<td>Physical health characteristics at baseline, N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD history</td>
<td>124 (5.77)</td>
<td>121 (7.25)</td>
<td>68 (4.60)</td>
<td>53 (3.85)</td>
</tr>
<tr>
<td>Thyroid disease</td>
<td>234 (10.89)</td>
<td>130 (7.78)</td>
<td>92 (6.23)</td>
<td>87 (6.32)</td>
</tr>
<tr>
<td>Liver disease</td>
<td>33 (1.54)</td>
<td>40 (2.40)</td>
<td>36 (2.44)</td>
<td>19 (1.38)</td>
</tr>
<tr>
<td>T2DM</td>
<td>108 (5.03)</td>
<td>140 (8.38)</td>
<td>45 (3.05)</td>
<td>86 (6.25)</td>
</tr>
<tr>
<td>Obesity (BMI≥30)</td>
<td>896 (41.71)</td>
<td>716 (42.87)</td>
<td>509 (34.36)</td>
<td>609 (44.26)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>184 (8.57)</td>
<td>173 (10.36)</td>
<td>103 (6.97)</td>
<td>130 (9.45)</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>43 (2.00)</td>
<td>132 (7.90)</td>
<td>50 (3.39)</td>
<td>49 (3.56)</td>
</tr>
<tr>
<td>Moderate/heavy alcohol use</td>
<td>1189 (57.00)</td>
<td>899 (54.75)</td>
<td>791 (54.82)</td>
<td>708 (52.21)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>711 (33.10)</td>
<td>652 (39.04)</td>
<td>632 (42.79)</td>
<td>567 (41.21)</td>
</tr>
<tr>
<td>Illicit drug use history</td>
<td>93 (4.33)</td>
<td>148 (8.86)</td>
<td>179 (12.12)</td>
<td>160 (11.63)</td>
</tr>
<tr>
<td>Mental health characteristics at baseline, N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous suicidal or non-suicidal self-harm</td>
<td>468 (22.44)</td>
<td>424 (25.82)</td>
<td>349 (24.19)</td>
<td>473 (34.88)</td>
</tr>
<tr>
<td>Sleep problems</td>
<td>200 (9.59)</td>
<td>197 (12.00)</td>
<td>191 (13.24)</td>
<td>230 (16.96)</td>
</tr>
<tr>
<td>Depression symptoms or diagnosis</td>
<td>1238 (57.64)</td>
<td>990 (59.28)</td>
<td>915 (61.95)</td>
<td>1015 (73.76)</td>
</tr>
<tr>
<td>Anxiety symptoms or diagnosis</td>
<td>144 (6.70)</td>
<td>150 (8.98)</td>
<td>137 (9.28)</td>
<td>201 (14.61)</td>
</tr>
<tr>
<td>Previous exposure to drug</td>
<td>1731 (80.59)</td>
<td>1157 (69.28)</td>
<td>886 (59.99)</td>
<td>847 (61.56)</td>
</tr>
</tbody>
</table>

Table 8.1 Patient characteristics

8.4.1 Self-harm

The rate of self-harm reported to primary care physicians in individuals prescribed maintenance mood stabiliser medication for BPD was 340 per 10,000 PYAR (95% CI 313 to 370 per 10,000 PYAR). In unadjusted analysis, self-harm rates were reduced in people taking lithium, compared to those taking valproate, olanzapine or quetiapine (Table 8.ii, Figure 8.ii). This was also the case after adjustment for PS, age, calendar year, and primary care practice (all other study drugs compared to lithium: HR 1.40; 95% CI 1.12 to 1.74).

After one-to-one PS matching with lithium, rates of self-harm remained higher in individuals prescribed valproate (N=1,186; HR 1.31; 95% CI 1.01 to 1.70), olanzapine
(N=1,100; HR 1.33; 95% CI 1.01 to 1.75) and quetiapine (N=790; HR 1.51; 95% CI 1.21 to 1.88). One-to-one matching of individuals taking lithium with those taking any other study drug showed higher self-harm rates in the non-lithium group (N=1501; HR 1.51; 95% CI 1.21 to 1.88).

Table 8.ii Rates of self-harm, accidental injury, and suicide by mood stabiliser

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
<th>Valproate, olanzapine or quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-harm</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N PYAR</td>
<td>146</td>
<td>152</td>
<td>137</td>
<td>128</td>
<td>417</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>205 (175-241)</td>
<td>392 (334-460)</td>
<td>409 (345-483)</td>
<td>582 (489-692)</td>
<td>442 (402-487)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.68 (1.34-2.12)</td>
<td>1.76 (1.39-2.23)</td>
<td>2.21 (1.74-2.82)</td>
<td>1.84 (1.52-2.23)</td>
</tr>
<tr>
<td>Model 1st HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.39 (1.08-1.78)</td>
<td>1.39 (1.07-1.79)</td>
<td>1.52 (1.15-2.01)</td>
<td>1.40 (1.12-1.74)</td>
</tr>
<tr>
<td>Model 2nd HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.31 (1.01-1.70)</td>
<td>1.33 (1.01-1.75)</td>
<td>1.36 (1.00-1.87)</td>
<td>1.51 (1.21-1.88)</td>
</tr>
<tr>
<td><strong>Accidental Injury</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N PYAR</td>
<td>388</td>
<td>255</td>
<td>190</td>
<td>154</td>
<td>599</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>583 (528-644)</td>
<td>669 (592-757)</td>
<td>569 (494-655)</td>
<td>705 (602-825)</td>
<td>641 (592-694)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.18 (1.01-1.39)</td>
<td>1.00 (0.84-1.19)</td>
<td>1.29 (1.06-1.56)</td>
<td>1.13 (1.00-1.29)</td>
</tr>
<tr>
<td>Model 1st HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.32 (1.10-1.58)</td>
<td>1.14 (0.95-1.37)</td>
<td>1.34 (1.07-1.69)</td>
<td>1.26 (1.07-1.47)</td>
</tr>
<tr>
<td>Model 2nd HR (95% CI)</td>
<td>1 [reference]</td>
<td>1.34 (1.09-1.65)</td>
<td>1.17 (0.94-1.47)</td>
<td>1.44 (1.09-1.91)</td>
<td>1.19 (1.01-1.41)</td>
</tr>
<tr>
<td><strong>Suicide</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, N PYAR</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Rate, per 10,000 PYAR (95% CI)</td>
<td>7 (3-16)</td>
<td>17 (8-36)</td>
<td>20 (9-42)</td>
<td>22 (9-52)</td>
<td>19 (12-32)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 [reference]</td>
<td>2.35 (0.74-7.46)</td>
<td>2.73 (0.86-8.64)</td>
<td>2.85 (0.81-10.06)</td>
<td>2.60 (0.96-7.03)</td>
</tr>
<tr>
<td>Model 1st HR (95% CI)</td>
<td>1 [reference]</td>
<td>2.71 (0.75-9.80)</td>
<td>3.18 (0.86-11.73)</td>
<td>3.01 (0.68-13.38)</td>
<td>2.86 (0.88-9.26)</td>
</tr>
</tbody>
</table>

PYAR, person-years at risk; HR, hazard ratio; CI, confidence interval. aAdjusted for propensity score, clustering by primary care practice, age and calendar year. bPropensity score matched (pairwise matching with lithium) adjusted for clustering by primary care practice, age and calendar year. ctoo few events for propensity score matched analysis.
8.4.2 Accidental injury

The rate of accidental injury was 616 per 10,000 PYAR (95% CI 579 to 656 per 10,000 PYAR). Rates of accidental injury were lower in people taking lithium compared to those taking valproate and quetiapine, but not olanzapine in unadjusted, PS adjusted, and PS matched analyses (Table 8.ii). Individuals prescribed lithium had lower accidental injury rates compared to those taking other study mood stabilisers, whether following adjustment for PS, calendar year, age and primary care practice (HR 1.26; 1.07 to 1.47), or following one-to-one PS matching with people taking valproate, olanzapine or quetiapine (HR 1.19; 95% CI 1.01 to 1.41).

8.4.3 Suicide

The rate of suicide deaths in the cohort was 14 per 10,000 PYAR; 95% CI 9 to 21 per 10,000 PYAR. The number of suicides was too low to show differences by individual drugs. The HR
point estimate for suicide was elevated for all other study drugs compared to lithium, but 95% CIs overlapped unity, indicating no effect (unadjusted HR 2.60; 95% CI 0.96 to 7.03, and PS adjusted HR 2.86; 95% CI 0.88 to 9.26) (Table 8.ii).

8.5 Discussion

8.5.1 Main findings

As far as I am aware this is the largest naturalistic longitudinal study of fatal and non-fatal self-harm rates in individuals with BPD treated with lithium, valproate, olanzapine or quetiapine. I found increased rates of self-harm in individuals prescribed valproate, olanzapine or quetiapine, compared to those prescribed lithium. I did not find differences in rates among valproate, olanzapine and quetiapine. This association remained after PS adjustment and matching. I also found reduced rates of accidental injury in those prescribed lithium, an important association that has not been widely investigated or found previously.

8.5.2 Comparison with previous literature

I did not find differences in rates of suicide because of the small number of suicides in the cohort. However, the point estimates for rates of suicide on lithium and valproate matched those found in the US retrospective cohort study by Goodwin et al. (7 per 10,000 PYAR and 17 per 10,000 PYAR respectively) (Goodwin et al., 2003) and are similar to other studies (Smith et al., 2009).

The lower rates of self-harm in those prescribed lithium may be due either to improved mood stabilisation compared to other treatments or specific effects on impulsive aggression and risk-taking. The similarity of the negative association between lithium and accidental injury and that between lithium and self-harm supports the latter hypothesis, as
there is little reason to expect that lower rates of depressive symptoms would reduce accidental injury (Khalsa et al., 2008). Also there is little evidence that lithium is superior to quetiapine in preventing depressive episode relapse (Miura et al., 2014).

8.5.3 Strengths and limitations

This study uses a large nationally representative sample to examine rates of fatal and non-fatal self-harm and accidental injury. This is something that RCTs have struggled to do, because they tend to exclude suicidal individuals and event rates are often too low for statistical comparison. The use of EHR to capture those episodes of self-harm managed entirely in primary care, as well as those admitted to secondary care, captures the true burden of self-harm morbidity—both in the community and hospital-presenting. Consequently, rates of recorded self-harm in my study were slightly higher than in previous cohort studies (Gibbons et al., 2009, Goodwin et al., 2003). However, it was not possible to test if all secondary care self-harm presentations are appropriately coded in THIN. As in any analysis of health records, this study would have missed episodes of untreated and unreported self-harm: only population survey methods could capture these episodes, and estimates generated in this manner are prone to response biases. The number of suicides was low, so I was not able to examine differences in rates between drugs. It is possible that misclassification or non-recording of suicides occurred. However, similarities with other cohorts suggest this is not a major problem (Goodwin et al., 2003, Smith et al., 2009) – or it is a problem with all cohorts. There is no evidence to suggest misclassification of cause of death would differ by drug. Exposure time was foreshortened because of both left truncation (for example, practices joining THIN later than 1995) and right censoring (for example, switch to or addition of another drug, patients leaving the primary care practice or dying of non-suicide causes). This censoring was equally distributed by exposure drug.
Potential biases relating to selection into the study should have been avoided by the use of contemporaneous, representative medical records, and information bias minimised by the use of THIN’s detailed and well recorded prescribing data as the exposure. However, exposure to the study drug is approximated through prescriptions issued, and may not reflect how the patient used the medication. It is possible that erratic adherence is more likely for drugs other than lithium (as lithium is more closely monitored through regular blood tests) and may have contributed to lithium’s perceived superiority. However these patients had no more physician contacts than those taking other medication, all individuals had to collect more than one prescription during their follow-up period (suggesting drug adherence), and other longitudinal cohort studies have not shown differential adherence (Sajatovic et al., 2007). As people taking lithium tended to be older, suicides could have occurred in this group prior to the start of follow-up, thus reducing the observed rate relative to other treatment groups. However, this should not be the case in the matched analysis.

Through PS adjustment and matching, I attempted to account for potential confounding, including confounding by indication, and it is reassuring that both analyses produce similar results. Despite the numerous variables included in the PS, it is possible that residual confounding remained. It may be that important socio-demographic or clinical factors were not captured by the score, and I cannot confirm balance of unobserved covariates (Austin et al., 2005, Stukel et al., 2007). Notably, detailed information on educational level, individual socioeconomic status and social support is lacking from the database. However, although these covariates are likely to be associated with self-harm, accidental injury, and suicide, they should not be associated with treatment allocation. For these (or any) unmeasured covariates to have an important impact on the results they would have to be
strongly associated with exposure and outcome and be independent of covariates included in the PS (Psaty et al., 1999, Schneeweiss, 2006).

Previously it has been shown that a combination of lithium and valproate was associated with the lowest rate of suicide attempt (Ahearn et al., 2013). This group (and other combinations) were excluded from my study, as I wanted to examine the association with monotherapy, and in fact this combination was rarely prescribed in THIN, despite recommendations for its use (Geddes et al., 2010).

8.5.4 Conclusions
In this representative UK study, individuals with BPD who were prescribed lithium had lower rates of self-harm and accidental injury than those with BPD taking other commonly prescribed maintenance treatments. Contrary to the FDA warning, I did not find higher self-harm rates in those prescribed valproate than those taking other (non-lithium) maintenance drug treatments. These findings are important as they support and augment the existing evidence from RCT and smaller cohort studies. Self-harm, accidental injury and suicide are important outcomes in BPD that appear to be amenable to modification through appropriate drug treatment.

8.5.5 Implications of the findings of this study
Implications for policy and practice are discussed in Chapter 9. Lithium for Suicidal Behaviours in Mood Disorders (The Li+ Study) is an RCT that is currently recruiting from within the US Department of Veterans Affairs. Participants will be patients with BPD or depression who have survived a recent episode of suicidal self-directed violence or were hospitalised specifically to prevent suicide. In theory, this impressive lithium plus treatment as usual versus treatment as usual trial will put pay to any doubt about the anti-suicidal properties of lithium (Smith and Attenburrow, 2016).
If lithium has anti-suicidal properties that are independent of mood stabilisation, then it may be appropriate to research its use in a wider range of diagnoses that are associated with self-harm and suicide. In particular, there is a paucity of evidence about the use of lithium in emotionally unstable personality disorder (Lieb et al., 2010, Rombold et al., 2014) and schizophrenia (Leucht et al., 2007), both of which have high rates of suicide and self-harming behaviour (as demonstrated in Chapter 3 for schizophrenia).

Apart from lithium, there are no pharmacological treatments that appear to have specific effects on self-harm or suicide (Saunders and Smith, 2016). A more comprehensive understanding of lithium’s mechanisms of action may lead to potential new drug development; this is further discussed in Section 9.7.4.
Chapter 9  Discussion, implications and conclusions

9.1  Summary

This chapter summarises the main findings of this thesis, and places them in the current research and clinical context. It more fully explores the meaning, implications, strengths, and limitations of the completed studies and the routine data approaches used. Finally, future research and plans for dissemination are discussed.
9.2 Key findings

The main objectives of this thesis were related to long-term outcomes in BPD and the effects of maintenance treatments:

i) To summarise previous observational studies of long-term prognosis in individuals with a diagnosis of BPD by examining all-cause and cause specific mortality via systematic review and meta-analysis (Chapter 3)

ii) To calculate recent time trends in all-cause mortality in the UK in individuals with BPD compared to individuals with schizophrenia and the general population (Study 1: Chapter 4)

iii) To determine relative rates of i) CVD deaths, ii) suicide, iii) CVD diagnoses, iv) self-harm in individuals diagnosed with BPD or schizophrenia compared to the general population, while accounting for sociodemographic factors (Study 1: Chapter 4)

iv) To determine relative efficacy of the four most commonly used maintenance mood stabiliser medications (lithium, valproate, olanzapine and quetiapine) via a network meta-analysis of all head-to-head and placebo controlled RCTs (Chapter 5)

v) To assess comparative effectiveness and tolerability of the four most common mood stabilisers by calculating rates of time to cessation of treatment, or add-on of another psychotropic medication in individuals prescribed lithium, valproate, olanzapine and quetiapine, while accounting for propensity to be prescribed one of these mood stabilisers (Study 2: Chapter 6)

vi) To calculate rates of adverse events on these four mood stabilisers, specifically chronic renal, hepatic, endocrine and metabolic effects, accounting for propensity to be prescribed one of these mood stabilisers (Study 3: Chapter 7)
vii) To determine rates of self-harm, unintentional injury and suicide on these four mood stabilisers, while accounting for propensity to be prescribed one of these mood stabilisers (Study 4: Chapter 8)

9.2.1 Mortality and morbidity in bipolar disorder

9.2.1.1 All-cause mortality in bipolar disorder

In my systematic review and meta-analysis of studies examining SMR (Chapter 3), all-cause mortality in BPD was 2.05 times that of the general population (95% CI 1.89 to 2.23) but heterogeneity was high ($I^2=96.2\%$) and could not be accounted for by study level factors, such as country or years of data collection.

In my cohort study of 17,341 with BPD followed up between 2000 and 2014 (Chapter 4), all-cause mortality was elevated by 1.79 times (95% CI 1.67 to 1.88) compared to the general population (after adjustment for age, sex, calendar year, area level deprivation and ethnicity). For comparison, all-cause mortality in schizophrenia was elevated by 2.08 times (95% CI 1.98 to 2.19). However, the overall summary estimate for BPD masks a widening mortality gap between individuals with this diagnosis and the general population. Beginning in 2006 the HR increased by 0.14 per year (95% CI 0.10 to 0.19) until 2014. Similarly the mortality gap between those with schizophrenia diagnoses and the general population increased by 0.11 (95% CI 0.04 to 0.17) per year between 2004 and 2010, and from 2010 to 2014 increased by 0.34 (95% CI 0.18 to 0.49).

9.2.1.2 Cardiovascular disease mortality

In my systematic review and meta-analysis of mortality in BPD, of all previous studies of mortality in BPD, 14 included estimates of mortality from circulatory disease (Chapter 3). The pooled SMR was 1.73 (95% CI 1.54 to 1.94). There was no evidence that this differed by sex (men: 1.81; 95% CI 1.61 to 2.05, women: 1.72; 95% CI 1.46 to 2.03).
In my mortality/morbidity cohort study (Chapter 4), using a more restricted definition of CVD (MI, CHD and CVE), there was no evidence of an increase in mortality after accounting for age, sex calendar year, area level deprivation, ethnicity and number of GP contacts (HR 1.10; 95% CI 0.84-1.46). Again, there was no difference by sex or age group. However, stratification by 5 year calendar period found that rates were elevated from 2010-2014 (HR 1.92; 95% CI 1.24-2.98). By way of comparison, CVD mortality in individuals with schizophrenia was elevated (HR 1.39; 95% CI 1.12 to 1.73), and dramatically so in those aged 50 or younger (HR 3.20; 95% CI 1.62 to 6.31). Additional adjustment for traditional CVD risk factors (smoking, high cholesterol, high blood pressure, BMI, and T2DM) explained this association.

9.2.1.3 Cardiovascular disease diagnoses

Despite having no increase in CVD death rates compared to the general population, rates of CVD diagnosis were elevated in individuals with BPD diagnoses, after adjustment for age, sex, calendar period, area level deprivation, ethnicity and primary care contacts (HR 1.41; 95% CI 1.26 to 1.58). The rate of CVD diagnoses was similarly elevated in individuals with schizophrenia. Unlike CVD deaths, the relationship between SMI and CVD diagnosis was not fully explained by additional adjustment for CVD risk factors. In both diagnostic groups, it was men who had increased rates of CVD diagnoses, compared to men in the general population.

9.2.1.4 Suicide

Of the 31 studies included in the systematic review and meta-analysis, 15 reported SMRs for suicide in BPD, with a summary estimate of 14.44 (95% CI 12.43 to 16.78) (Chapter 3). There was no evidence that this was differentially elevated by sex. In my cohort study the HR for suicide in BPD compared to the general population, after adjusting for age, sex
calendar year, area level deprivation, ethnicity and number of GP contacts was 12.66 (95% CI 7.79 to 20.58) (Chapter 4). There was no evidence of an interaction by sex, age group or calendar period. Suicide rates in people with schizophrenia were similarly elevated, relative to the general population.

9.2.1.5 Self-harm

Elevated rates of self-harm compared to the general population were identified in those with BPD (adjusted HR 25.24; 95% CI 22.37 to 28.49) and rates were similarly elevated in those with schizophrenia (Chapter 4). Individuals aged 50 or younger in both diagnostic groups had markedly elevated self-harm rates (adjusted HR 55.74; 95% CI 45.35 to 68.53 in those with BPD and adjusted HR 52.07; 95% CI 42.43 to 63.92 in those with schizophrenia).

9.2.2 Maintenance treatment for bipolar disorder

9.2.2.1 Effectiveness and tolerability of the four most common maintenance mood stabiliser medications

A NMA of RCTs examining any combination of lithium versus valproate versus olanzapine versus quetiapine versus placebo found no statistically significant difference between active drugs for effectiveness or tolerability (Chapter 5). All active drugs were superior to placebo for both outcomes. Ranking of active drugs suggested that olanzapine is potentially most effective, and quetiapine most well tolerated. Despite this, lithium is viewed favourably because of the results of a number of recent RCTs, which have been optimised for newer study drugs (such as quetiapine) and have used lithium as an active control. These RCTs have provided support for lithium’s mood stabilising properties beyond alternatives.

In my cohort study of 5,089 patients, lithium monotherapy had a better effectiveness and tolerability profile (defined as stopping, swapping to an alternative mood stabiliser, or add-
on of another psychotropic), than those prescribed valproate, olanzapine or quetiapine (Chapter 6). However, for all medications, time to monotherapy treatment failure was short (median 102 days for lithium, for example). Differences remained after accounting for clustering by GP practice, age, sex, calendar year and ethnicity. This was also true when I explored whether rate differences could be explained by variation between those prescribed each of the different drugs after PS adjustment and matching. In the matched analysis HRs were 1.20 (95% CI 1.10 to 1.32) for valproate, 1.17 (95% CI 1.07 to 1.29) for olanzapine and 1.32 (95% CI 1.20 to 1.45) for quetiapine, relative to lithium. Less strict definitions of treatment failure, and examining only those individuals stable for at least three months on medication did not alter these conclusions.

9.2.2.2 Adverse effects of maintenance mood stabiliser medication

In a cohort of 6,671 patients, those prescribed valproate, olanzapine and quetiapine had lower rates of ≥CKD stage 3 than those prescribed lithium, after accounting for PS, age, calendar year and clustering by GP practice (Chapter 7). HRs were 0.56 (95% CI 0.45 to 0.69), 0.57 (95% CI 0.45 to 0.71) and 0.62 (95% CI 0.47 to 0.80) for valproate, olanzapine and quetiapine respectively. However, there was no difference in rates of more severe renal failure (≥CKD stage 4) between groups.

In the fully adjusted analysis, rates of hypothyroidism were reduced in individuals taking valproate (HR 0.60; 95% CI 0.40 to 0.89) and olanzapine (HR 0.48; 95% CI 0.29 to 0.77), relative to those in the lithium treated group. Similarly, rates of hyperthyroidism and hypercalcemia were reduced in those taking alternatives to lithium.

Conversely, rates of >7% and >15% weight gain were significantly elevated in groups treated with valproate, olanzapine or quetiapine, compared to those taking lithium. For
example in the fully adjusted PS analysis patients taking valproate (HR 1.62; 95% CI 1.31 to 2.01), olanzapine (HR 1.84; 95% CI 1.47 to 2.30), and quetiapine (HR 1.67; 95% CI 1.24 to 2.20) had elevated rates of >15% weight gain. I could find no significant difference in rates of CVD, T2DM or hepatotoxicity, but these outcomes were rare. A number of sensitivity analyses to address missing data did not alter these conclusions.

9.2.2.3 Anti-suicidal effects of maintenance mood stabiliser medications

In the same cohort as the previous study, rates of self-harm were lower in those individuals prescribed lithium, than those prescribed valproate, olanzapine or quetiapine, after adjusting for PS, age, calendar year and clustering by GP practice (Chapter 8). This was also true after one-to-one PS matching with lithium (valproate HR 1.31; 95% CI 1.01 to 1.70, olanzapine HR 1.33; 95% CI 1.01 to 1.75), and quetiapine HR 1.36; 95% CI 1.00 to 1.87). The number of suicides was too low to show differences by individual drugs (comparing lithium with other study drugs: PS-adjusted HR 2.86; 95% CI 0.88 to 9.26). Similarities between lower self-harm rates and lower accidental injury rates in those taking lithium suggested that lithium might have specific effects on impulsive aggression and risk taking that the other maintenance mood stabilisers do not.

9.3 Implications of findings from my mortality and morbidity studies

9.3.1 Key points for clinicians and patients

i) Mortality is elevated in BPD relative to the general population

ii) Mortality rates are decreasing in people with BPD and schizophrenia

iii) Mortality rates are not decreasing in line with the general population – the mortality gap is widening
iv) From 2000, CVD mortality is not elevated in the group with BPD (except 2010-2014), but CVD diagnosis is more frequent than in the general population

v) Suicide and self-harm rates are dramatically elevated in people with BPD

vi) Non-fatal self-harm presenting to emergency department or primary care may provide an opportunity for intervention and reduce the risk of suicide

vii) Non-CVD causes of death contribute to the elevated natural mortality rate in BPD and also need to be monitored

9.3.2 Key points for policy makers

i) Mortality rates reflect the quality of psychiatric and physical healthcare provided to individuals

ii) Current interventions and policies are not successfully addressing the mortality gap

iii) With the expansion of EHRs tracking mortality in real time will become feasible

iv) Policies to reduce CVD deaths in SMI are admirable, but measuring impact on mortality rates is challenging in short time periods

v) The apparent increase in CVD mortality in BPD relative to the general population after 2010 needs to be monitored to see if this is a continuing trend; it may be a true elevation or it may reflect better/earlier CVD diagnosis in this group

vi) There is a need for a clear and consistent message about primary care and psychiatrists’ responsibilities for their patients physical health monitoring

vii) Increased rates of CVD diagnosis suggest CVD risk factor monitoring is occurring

9.3.3 Key points for researchers

i) The SMR is the most common measure of mortality, but does not account for important confounding factors. Life expectancy or years of life lost may be easier for patients, clinicians and policy makers to understand
ii) SMRs are highly heterogeneous by time, place, treatment setting and other unidentified covariates

iii) Further research is necessary to better identify predictors of specific cause mortality, particularly suicide

iv) Further research is necessary to understand the effects of behaviour, lifestyle and medication on mortality and morbidity in BPD, and not simply assume they are the same as in schizophrenia

v) Disease processes beyond CVD require further investigation in BPD

vi) Reasons for non-detection of CVD before a fatal event in schizophrenia should be investigated. For example: do people with schizophrenia potentially get assessed differently when they present with chest pain? Do they present with atypical symptoms (for example “black out”) more commonly? Why does this appear to be different in people with BPD?

vii) There is a need to understand if, once diagnosed with non-fatal CVD, individuals with SMI receive equitable treatment and if risk factor management differs between SMI and non-SMI groups. My findings suggest this may be different for BPD and schizophrenia

9.3.4 All-cause mortality in people with bipolar disorder

The mortality rate is elevated in people with BPD relative to the general population, and this gap appears to be increasing (Chapter 4). Despite focus of UK policy on improving the health of individuals with schizophrenia (and SMI more generally), mortality for this group is also increasing relative to the general population (Schizophrenia Commission, 2012). This finding masks the fact that for both groups there has been a gradual decline in mortality rates since 2000. Clearly, this reduction in all-cause mortality is important and should not be overlooked. However, it has been argued that reductions in health inequalities should
be a clinical priority, especially as in this case the inequality is likely to be largely avoidable and would provide cost effective benefits if addressed (Woodward and Kawachi, 2000).

An over-all decline in mortality rate may suggest that some interventions in addressing SMI morbidity and disability have had an impact. It may also suggest that increased prescription of SGAs (Hayes et al., 2011, Verdoux et al., 2010) has not resulted in the increased mortality because of weight gain, metabolic abnormalities, T2DM and CVD that was anticipated (Kahn et al., 2008, Lieberman et al., 2005), though it may be too soon to observe this potential impact on mortality. From premature all-cause mortality we can extrapolate to more years lived with disability in individuals with BPD than the general population. More continuous monitoring of mortality in people with BPD and schizophrenia might guide us in evaluating the impact of interventions to manage physical comorbidity, reduce inequalities in medical care provision and prevent inequalities in their background risk factors. This sort of research becomes increasingly straightforward with the wider coverage of EHR and increasing availability of record linkages. Potentially this could become a semi-automated process that provided close-to-real-time information on mortality trends in different sociodemographic groups.

9.3.5 Cause-specific mortality in people with bipolar disorder

9.3.5.1 Cardiovascular disease mortality

As with the general population, CVD is the commonest cause of death in individuals with BPD and schizophrenia but this is not necessarily reflected in elevated SMRs or HRs (National Institute for Health and Care Excellence, 2014). The pooled SMR for circulatory disorders in BPD (Chapter 3) was elevated. However, overall in the period 2000-2014, I found no evidence that CVD death was more common in individuals with BPD than in the general population (although it was higher in the period 2010-2014) (Chapter 4). This may
suggest that since 2000 the elevated risk of death in the population with BPD is from other causes and that there is no CVD mortality gap to close. Certainly, this problem is not as marked in BPD as in individuals with schizophrenia, where CVD deaths in the 16-50 age group were more than three times as common as in the general population. Despite these differences in CVD mortality, diagnoses of CVD are similarly elevated in both BPD and schizophrenia relative to the general population. This suggests people with BPD are at increased risk of MI, CHD and CVE, but may either have better prognosis CVD events (e.g., non-ST elevation MIs) or receive better care that means these events are not fatal. I did not explore potential reasons for this, but it does not seem to come from increased opportunity to receive a diagnosis because frequency of contact with primary care was similar in both groups. Risk factors for CVD were also comparable in both groups (apart from smoking, which was more common in individuals with schizophrenia). Adjustment for these risk factors did not fully explain the elevation in CVD above population baseline.

9.3.5.2 Suicide

Suicide is the cause of death that is consistently most elevated relative to the general population (Chapter 3 and Chapter 4). Rates appear to be similar in both BPD and schizophrenia, with point estimates tending to be higher in BPD (Chapter 4). Whilst it is hard to distinguish differences in suicide rate by age, self-harm is dramatically elevated in the younger (16-50) population. Suicide rates probably reflect the quality of care provided (both from primary and secondary care), the level of social support received by the patient, and the clustering of factors that increase risk (such as comorbid substance misuse or physical illness). Predicting who will die by suicide remains a major challenge in psychiatry (Glenn and Nock, 2014). In BPD, risk factors tend to be similar to those in the general population. Beyond these, depressive or mixed episodes and rapid cycling appear to have
detrimental effects (Gonda et al., 2012). However, clinically applicable prediction models do not exist.

9.3.5.3 Other causes of mortality in bipolar disorder

All cause-specific SMRs in BPD are elevated (Chapter 3), and are comparable to those for schizophrenia (Saha et al., 2007). Beyond CVD, in terms of natural causes, particular attention needs to be paid to potentially synergistic effects of metabolic abnormalities, inflammatory abnormalities, adverse health behaviours, early life experiences and lifestyle factors. It is likely that these factors cluster in a manner that increases all morbidity and mortality. Specifically with relation to unnatural deaths, little research has focused on the increased risk of accidental death (Chapter 8). Accidental injury, unlike suicide, is likely to be more common in the (hypo)manic phase of illness (Khalsa et al., 2008) and therefore may be more amenable to treatment.

9.3.6 Why is there a still “mortality gap”?

A number of factors are likely to contribute to the increased mortality and morbidity in people with BPD. These include poor access to healthcare, lifestyle factors, social deprivation, the effects of psychiatric treatment and factors intrinsic to the disorder itself.

9.3.6.1 Reduced access to healthcare compared to the general population

Poor access to healthcare can occur through both healthcare system failures and failure of the patient to engage “appropriately” with the healthcare system. Psychiatrists and other doctors may regard reporting of a physical symptom as a sign of mental illness (diagnostic overshadowing) (Jones et al., 2008). Clinicians may focus on the mental health issue at the expense of physical healthcare and may not possess the appropriate skills to diagnose the physical complaint. Resources for diagnosis and treatment may be lacking in a psychiatric setting. Physicians, surgeons and emergency department staff may be reticent to treat
people with severe mental illness because engagement is complicated by the mental illness, because of stigmatising attitudes or because of issues with capacity, which they may feel less well versed in (Viron and Stern, 2010).

Patients may actively avoid contact with general healthcare services (although this is not suggested from the baseline characteristics in my mortality and morbidity study (Study 1; Chapter 4). Poor general treatment adherence may also play a part. Some may have difficulty interpreting and communicating their physical health needs and problems in general, or may be unaware of physical health problems because of cognitive deficits or symptoms. Some patients may have difficulty undertaking tasks such as making appointments, comprehending healthcare or carrying out recommended lifestyle changes, without additional support.

9.3.6.2 Negative health behaviours in individuals with bipolar disorder

Individuals with BPD have been shown to engage in more unhealthy and high risk behaviours than the general population (Parks et al., 2006). They have been found to have poorer diets, exercise less frequently and smoke more than the general population (Kupfer, 2005). They are more likely to use illicit substances. They are less likely to practice safe sex. They are more at risk of coercion, exploitation and violence (Baxter et al., 2016). Many of these covariates are not well captured in primary care EHR and are a limitation of the data source (see Section 9.5.9).

9.3.6.3 Medication effects on physical health

Maintenance mood stabilisers (especially antipsychotics) cause weight gain (see Chapter 7), as do most antidepressants. This in turn can result in CVD, hyperlipidemia and T2DM. Medications also have side effects such as sedation or Parkinsonism that can contribute to reduced physical activity (Connolly and Kelly, 2005).
9.3.6.4 The direct effects of mental illness on physical health

It has been shown that prior to the introduction of antipsychotic medication in the 1950s, individuals with schizophrenia were at increased risk of having high BMIs and T2DM (Green et al., 2000). Additionally there is evidence that drug naïve patients with schizophrenia have increased intra-abdominal fat, impaired fasting glucose tolerance, more insulin resistance than the general population (Ryan et al., 2003) and increased risk of metabolic syndrome (Vancampfort et al., 2013b). There is no similar evidence for BPD, because of a lack of research in this area. However, as with schizophrenia (Andreassen et al., 2013), there is now evidence of shared genetic risk for BPD, T2DM and elevated BMI (Winham et al., 2014). This suggests that there may be something intrinsic to the illness itself that reduces physical health, or that there is a common susceptibility to SMI and cardiometabolic disorders.

9.3.7 What should be done to address the “mortality gap”?

9.3.7.1 Physical health monitoring

Worldwide there are multiple guidelines for the monitoring of physical health in SMI. In the UK, two key guidelines are in place: the NICE BPD guidelines (National Institute for Health and Care Excellence, 2014) and the QOF for primary care (Employers and Committee, 2012). The guidelines focus on the monitoring of BMI, blood pressure, HbA1c or glucose and the ratio of total cholesterol to high-density lipoprotein. These guidelines do not state what comprises adequate frequency of testing, but NICE suggests monitoring physical health at least once a year. In addition, they state that it is primary care, rather than psychiatrists, who have responsibility for this monitoring, but that health professionals in secondary care should ensure this monitoring is happening. This has potentially led to some confusion about who takes responsibility for physical health issues and may have contributed to the increasing difference between mortality in BPD and the general
population. Some have therefore argued that mental health services should take responsibility for clearly understanding and monitoring the physical health problems of their patients, especially in relation to potential adverse effects of the medication they are prescribing (Hasselt et al., 2015). There may be additional benefits to psychiatrists and other mental health staff being involved in the physical healthcare of people with BPD. It is clear that a whole range of chronic health problems have a significant effect on mental state and complicate the diagnosis and treatment of the disorders, so in identifying and treating patients physical complaints we may go some way to improving their mental state. There is also the opportunity to strengthen the therapeutic alliance through addressing physical problems that the patient may feel have been neglected by other healthcare providers. There is evidence that routine monitoring via blood tests is not meeting standards set by guidelines (Chapter 7). In addition, there is some evidence that the increase in monitoring that has occurred has not had an effect on reducing the mortality gap (Chapter 4).

9.3.7.2 Smoking cessation

Considering the contribution of smoking to CVD, cancer and respiratory mortality, there is evidence that smokers lose at least 10 years of lifespan, and that stopping smoking before the age of 40 can prevent more than 90% of the excess mortality caused by continuing (Pirie et al., 2013). There is also evidence that generic smoking cessation programmes have been preferentially adopted by the non-SMI population (Lawrence and Kisely, 2010). This could partially explain the increasing HR seen in recent years (Chapter 4). However, systematic reviews of RCTs confirm that treatment interventions based on behavioural support and pharmacotherapy that work in the general population are also (and approximately equally) effective in smokers with mental illness and do not
appear to worsen psychiatric symptoms (Banham and Gilbody, 2010, Ratschen et al., 2011).

9.3.7.3 Exercise and healthy eating

Strategies to help individuals with SMI manage their weight include restriction of caloric intake, pharmacological interventions, and behavioural interventions. Behavioural healthy eating programmes offer the best evidence for sustained change in the SMI population (Daumit et al., 2013). A review of healthy eating programmes in SMI patients suggests that basic approaches to caloric reduction can be as effective as more comprehensive efforts (Cabassa et al., 2010).

Exercise interventions have been shown to be effective in increasing levels of physical activity in patients with SMI (Ussher et al., 2007). A simple intervention such as walking, either in the form of supervised walking groups or unsupervised, is one of the easiest, safest and most inexpensive types of exercise to promote. It is also one of the most popular forms of exercise for those with and without chronic illness (Siegel et al., 1995).

9.3.7.4 Prescribing to reduce morbidity and mortality

Controversy and uncertainty remain as to whether psychotropic prescribing increases overall mortality (Murray et al., 2016). Most of the literature is focused on high dose, long-term prescribing of antipsychotics in schizophrenia, but these issues also apply to BPD. Recently it has been suggested that psychiatrists should 1) treat with the minimum necessary dose, 2) use a weight-sparing medications, 3) reduce to the lowest possible dose following recovery and 4) ensure access to non-pharmacological treatments (Murray et al., 2016). However, this is further complicated in BPD because of the episodic nature of illness,
inherent risk of different polarities of illness, potential for kindling effects and the limited evidence for effective psychological approaches (Goodwin et al., 2016, Oud et al., 2016).

It may be that psychotropic medications reduce suicide (Torniainen et al., 2015) but increase CVD deaths (Saha et al., 2007), but again this is not clear-cut. We continue to know very little about the very long-term effects of maintenance treatment, and I have tried to address this in Chapter 7 and Chapter 8. However, these studies do not examine medicated vs. unmedicated groups because taking medication is likely to be a surrogate marker for both illness severity and overall healthier behaviour (Mace et al., 2015).

9.4 Implications of findings from my maintenance treatment studies

9.4.1 Key points for clinicians and patients

i) Individuals prescribed lithium are likely to remain on monotherapy for longer than those prescribed valproate, olanzapine or quetiapine, suggesting it is more effective and tolerable

ii) Lithium should be considered the first-line treatment for BPD

iii) Lithium is associated with increased rates of mild or moderate CKD, but not necessarily end-stage renal failure (although I cannot rule this out from my study results)

iv) Lithium is associated with increased detection of hypothyroidism, hyperthyroidism and hypercalcemia

v) Alternatives to lithium are associated with higher rates of weight gain

vi) Patients taking olanzapine additionally have increased rates of hypertension
vii) Although CVD and T2DM diagnoses rates did not differ by drug type it is likely that increased weight gain, CKD and hypertension are major risk factors for these adverse events.

viii) Monitoring of physical health and blood test parameters is necessary.

ix) In particular calcium is not monitored well in primary care, despite being part of NICE guidance.

x) If adverse events occur, a discussion about the risks and benefits of continued treatment should be initiated with the patient.

xi) Both self-harm and accidental injury rates are lower in patients prescribed lithium, this should be discussed as an additional potential benefit with patients.

9.4.2 Key points for policy makers

i) Supporting clinicians to prescribe lithium may be necessary. Much like specialist clinics mitigate some of the risks and complications of prescribing clozapine, lithium clinics could allow better and safer prescribing.

ii) Improved documentation of physical health monitoring relating to specific prescribing choices may improve outcomes.

iii) Calcium monitoring should be added to QOF.

9.4.3 Key points for researchers

i) It is unclear if individuals who have the best chance of responding to lithium can be identified in advance or early in treatment.

ii) Research into recovery from mood stabiliser adverse events is very limited, it is unclear whether (and at what point) clinicians should stop or alter medication.

iii) If lithium, beyond its mood stabiliser properties acts directly on impulsivity and aggression, then it may have therapeutic benefit in other psychiatric disorders.
iv) A clearer understanding of lithium’s mechanism of action may advance our understanding of the pathophysiology of BPD and help identify drugs suitable for repurposing or targets for new drug development.

9.4.4 Effectiveness and tolerability of maintenance mood stabiliser medications

A number of reviews and guidelines have aligned to suggest lithium should be the first-line prescribing choice for maintenance treatment in BPD (Goodwin et al., 2016, Miura et al., 2014, National Institute for Health and Care Excellence, 2014, Severus et al., 2014). None of these were in existence at the time the studies in this thesis were conceived. It remains the case that RCTs and meta-analyses of these RCTs do not suggest that lithium is overwhelmingly superior to active comparators. However, lithium is the only drug that has been found to be efficacious and tolerable compared to active comparators under non-enriched conditions (i.e., RCTs that are set up to favour the non-Lithium drug) (Licht, 2012, Miura et al., 2014, Chapter 5). Because the outcome measure in recent RCTs has been time to first relapse or recurrence, any treatment continued beyond this point is not evidence-based. This also does not allow for the fact that a partial response may be clinically useful, and that mood stabilisers often require time to work (including dose optimisation) (Licht, 2012). My study (Chapter 6) robustly showed that individuals prescribed lithium took it for longer as monotherapy, before stopping medication, swapping to an alternative or having medication added to their treatment regimen. This outcome represents a combination of effectiveness and tolerability that has been widely used in RCTs (Miura et al., 2014). Stopping, swapping and add-on events were common with all drugs, but this is consistent with other naturalistic studies (Kessing et al., 2007, Schumann et al., 1999). It is reassuring that this study corroborates this treatment approach. However, in clinical practice, the perception that lithium is complicated to prescribe and associated with acute and long-
term adverse effects constitutes a barrier to its effective prescription and management (these issues are discussed below and in Chapter 7). Additionally, it was not possible in the cohort from THIN to examine different effects of combination therapies, which are becoming increasingly common (Licht, 2012).

### 9.4.5 Adverse events during maintenance mood stabiliser treatment

The four commonly used mood stabiliser medications investigated are associated with a number of adverse effects (Chapter 7). Specifically, with regards to lithium the potential for irreversible long-term renal damage is likely to be the clinician’s greatest concern. With regards to all of the studied drugs, but particularly olanzapine and quetiapine, significant weight gain and the related cardiometabolic sequelae are a major concern. Neither of these adverse events had previously been quantified relative to other medications. I was able to examine these adverse events, and a number of others related to each drug.

#### 9.4.5.1 Renal impairment during maintenance mood stabiliser treatment

That lithium is associated with mild or moderate renal impairment cannot be disputed. It is seen clinically, and has been shown in a number of longitudinal studies (Chapter 7). However, the implications of this are unclear. Previous studies have been hampered by lack of active comparator treatments and inability to appropriately manage potential surveillance bias. The study in this thesis addresses these previous limitations by comparing lithium with valproate, olanzapine and quetiapine and by completing an analysis accounting for the probability of having a blood test.

≥CKD stage 3 was twice as common in those taking lithium. However, the majority of these cases will be at the mild end. Guidance for management of CKD stage 3 suggests active monitoring and lifestyle advice (unless there are signs of progression of renal disease such as proteinuria or haematuria). I was not able to examine records of these events in THIN,
and often a finding of eGFR<60 ml/min/1.73m\(^2\) but >30 ml/min/1.73m\(^2\) will result in no change in management beyond that which would be received by people with SMI in general (i.e., lifestyle advice and additional monitoring) (Frankel et al., 2005, Levey and Coresh, 2012). At this stage, however, there is already an approximately 40% increased risk of cardiovascular events (Go et al., 2004) but continuing lithium is not contraindicated (Werneke et al., 2012). Risk of progression from mild to end-stage CKD does not seem to be modified by continuing or stopping lithium (Bocchetta et al., 2015). Overall ≥CKD stage 4 was rare in my cohort, such that it was not possible to discern differences in rates between those taking lithium and patients taking other mood stabilisers (Chapter 7). End-stage CKD has also been found to be rare in other recent cohort studies (Aiff et al., 2014a, b, Bendz et al., 2010). As such, the risks surrounding CKD in lithium treated patients who are maintained on a non-toxic, clinically therapeutic dose are likely to have been overestimated. I am not aware of methods for identifying patients at greatest risk of progression to end-stage renal failure from CKD stage 3 whilst taking lithium. This is an important future research focus.

9.4.5.2 Thyroid disease during maintenance mood stabiliser treatment

Hypothyroidism rates were elevated in lithium treated patients, compared to valproate or olanzapine treated patients. It was not possible to be certain that rates were higher than in those taking quetiapine, as CIs were wide and included unity. Hypothyroidism has long been recognised as a complication of lithium treatment. However, compared to alternative mood stabilisers the risk has not been quantified. In Study 2 in this thesis, rates of hypothyroidism during lithium treatment were less than double rates during treatment with alternatives. This is a lower relative risk than comparisons with the general population (McKnight et al., 2012) and potentially reflects the higher baseline risk of thyroid disease
associated with BPD itself (Kupka et al., 2002). The rate (approximately 2 per 100 PYAR) is ten times what is seen in general population studies (Garmendia Madariaga et al., 2014).

Although hyperthyroidism is very rare, it was more common in those taking lithium, than those taking valproate or olanzapine (but not quetiapine). As far as I am aware, this is the first study to be large enough to show definitive differences in hyperthyroidism rates between lithium and other drugs. Four previous case-control studies exist, with a pooled effect estimate overlapping no effect (McKnight et al., 2012) and one cohort study, with multiple limitations, also found no elevated risk in those taking lithium (Shine et al., 2015).

Early identification and appropriate treatment of thyroid disease is vital in BPD as it can complicate the clinical picture and reduce the chance of recovery (Fagiolini et al., 2006). Even low but normal range thyroid function is associated with slower treatment response (Cole et al., 2002) Thyroid function monitoring is more common in individuals prescribed lithium than other drugs (75% vs 55%). However, given the increased risk to all patients with BPD, and the overlap in symptomatology, all of these patients should have been screened for thyroid abnormalities (National Institute for Health and Care Excellence, 2014).

9.4.5.3 Hypercalcemia during maintenance mood stabiliser treatment

Clinically elevated levels of calcium were more than 4 times as common in individuals taking lithium than other mood stabilisers. However, the majority of individuals never received a calcium blood test during their exposure period. This suggests that this problem often goes unrecognised in clinical practice. Other studies have found more elevated rates compared to the general population, but there are no active comparator studies (Grünfeld and Rossier, 2009, McKnight et al., 2012). As stated in Chapter 7 the HR reported is likely to
be an underestimation, due to missing data. In treatment with lithium, most hypercalcemia will be due to hyperparathyroidism.

Detecting hypercalcemia is particularly important as some of the symptoms may be confused with side effects of lithium or signs of lithium toxicity, such as polyuria, polydipsia, dyspepsia, fatigue, nausea, cognitive impairment and muscle weakness (Inzucchi, 2004). Hypercalcemia can be a causal factor in renal failure, and is also associated with arrhythmias and osteoporosis (Inzucchi, 2004).

**9.4.5.4 Weight gain during maintenance mood stabiliser treatment**

Weight gain was lowest in patients prescribed lithium. Point estimates for >7% and >15% weight gain were highest in the group taking olanzapine. The rate of >15% weight gain was around 6 per 100 PYAR in patients taking alternatives to lithium. As far as I am aware, this is the first comparative study of weight gain during BPD treatment. Irrespective of medication effects, patients with BPD are at increased risk of being overweight or obese, relative to the general population (Keck and McElroy, 2003). This can be seen in the baseline characteristics of patients in my mortality/morbidity study (Study 1: Chapter 4). Weight gain is associated with increased mortality risk caused by CVD, T2DM, cancer and respiratory problems, but it also has psychological consequences such as low self-esteem and self-image, reduced socialisation and reduced activity (Torrent et al., 2008). Weight gain is also likely to have an impact on medication adherence and consequently on illness course. It is therefore important that risks of weight gain be minimised, if possible, in these patients. Additionally, although weight gain has been traditionally seen as an early event with regards to antipsychotic treatment (Sussman, 2001), in my study rates of >15% weight gain were stable over a 5 year follow-up period (Chapter 7). Despite lithium having the
lowest propensity, of the drugs studied, to cause weight gain it is still associated with weight gain compared to placebo (McKnight et al., 2012). In contrast to risks of renal and endocrine adverse events, there is no tension between the preferential prescribing of lithium for preventing relapse or recurrence and the risk of weight gain.

9.4.5.5 Cardiovascular disease and diabetes during maintenance mood stabiliser treatment

Both weight gain and CKD increase the risk of CVD, therefore although we may hypothesise that olanzapine will be associated with more CVD, this may not necessarily be true. In Study 3, I did not find differentially elevated rates of CVD by drug, although 95% CI were wide. I am not aware of any study that has been able to separate these medications in terms of CVD risk. Compared with the full BPD population (Chapter 4), rates of CVD in those treated with lithium, valproate, olanzapine or quetiapine were up to twice as high. However, this cannot be not be taken as proof that these medications double the risk of CVD, as there are likely to be many other differences between individuals treated with these drugs and those who simply receive a BPD diagnosis. I did not examine how these patients were treated following diagnosis of their CVD and this would have important implications for their mortality risk.

T2DM is also a multifactorial disease that is associated with considerable illness burden. As with CVD, I was unable to find differences in rates by drug. I am aware of no literature that suggests that there is a direct link between lithium prescription and T2DM, but increased rates in this group are possible via weight gain, CKD and CVD risk increases. Conversely, olanzapine has been found to increase the risk of T2DM by more than 4 times that seen in untreated patients (Gianfrancesco et al., 2003) and appears to carry a direct risk that is independent of weight gain (Lean and Pajonk, 2003). Compared to the population who receive a diagnosis of BPD, individuals treated with these drugs clearly need increased
monitoring of cardiometabolic parameters to reduce the risk of morbidity and mortality from these diseases.

9.4.5.6 Hepatotoxicity during maintenance mood stabiliser treatment

I found no differences in rates of hepatotoxicity, but this outcome was extremely rare. Contrary to expectations, impaired liver function was not elevated in individuals taking valproate (Betrosian and Frantzeskaki, 2006). In the unadjusted analysis, it was elevated in those taking quetiapine. The potential risk of quetiapine induced liver damage remains limited to case-reports, the same is true of olanzapine (Sedky et al., 2012). Of all of the adverse events covered in this thesis, this is likely to be the rarest and should have least impact on prescribing choice, unless there is evidence of pre-existing liver impairment (Sedky et al., 2012). Co-prescribing of antipsychotic medications, anticonvulsant moods stabilisers and/or antidepressants may increase the risk of hepatotoxicity. Additionally alcohol and other drug use may have an impact on liver function, but this has not been widely investigated (Sedky et al., 2012).

9.4.5.7 Other adverse effects

I was unable to examine a number of acute side effects of these drugs, such as tremor and thirst with lithium. Whilst acute adverse effects such as these may be disturbing to patients and may reduce medication adherence, they are not life threatening and are unlikely to lead to a change in prescribing if occurring in isolation. I therefore focused on serious and potentially life threatening (direct or indirect) adverse effects.

9.4.6 Self-harm, accidental injury and suicide during treatment with maintenance mood stabiliser medications

Self-harm was less common in individuals taking lithium (Chapter 8) and this finding is supported by existing literature (Cipriani et al., 2013a). Accidental injury was also less
common in patients taking lithium. I found no evidence to support the FDA warning that valproate is associated with increased suicide risk (Busco, 2008). Whilst it is important to understand the risk of these adverse outcomes in and of themselves, Study 4 also attempted to further understand the potential mechanism by which lithium may display anti-suicidal properties.

Three mechanisms have been proposed:

1. Lithium requires closer monitoring, increasing psychosocial support and therefore reducing self-harm
2. Lithium reduces depressive episodes, therefore reducing self-harm
3. Lithium specifically reduces impulsive aggression, therefore reducing self-harm

Mechanism 1 is unlikely to be correct as Study 4 showed that patients taking lithium had no extra primary care contacts than those taking other study drugs, additionally I would not necessarily expect accidental injury rates to be effected by closer monitoring. Mechanism 2 is unlikely to be correct because lithium is not significantly better at preventing or shortening depressive episodes than other mood stabilisers such as quetiapine. In addition, I would not expect depressive episodes to be associated with risk of accidental injury (in fact accidental injury is associated with mood elevation (Khalsa et al., 2008)). If mechanism 3 were correct then I would expect to see reduced rates of self-harm and accidental injury in patients prescribed lithium: and this is what was observed. The theory that lithium has specific serotonin-mediated effects, which reduce aggressive, risk-taking and impulsive behaviour is also supported by previous literature (Fawcett, 2001, Kovacsics et al., 2009, Müller-Oerlinghausen and Lewitzka, 2010).
The point estimate for suicide rate in those taking lithium was lower than in patients taking other medications, but the study was underpowered to be certain of this difference. That rates were similar to previous studies is reassuring (Goodwin et al., 2003, Smith et al., 2009) and a meta-analysis of longitudinal studies may be an appropriate next step to investigate drug-specific suicide rates.

Taking the evidence as a whole, there is a considerable argument for psychiatrists preferentially recommending lithium where self-harm is part of the clinical presentation. Potentially most benefit will come from starting lithium early in the illness course, where rates of self-harm and completed suicide are highest relative to the general population (Chapter 4).

9.5 Strengths and limitations of studies making up this thesis

Study specific strengths and limitations are discussed in each chapter. The more fundamental limitations are discussed here:

9.5.1 Use of electronic health records

Some of the strengths and limitations of using EHR for the studies in this thesis were discussed in Chapter 1 (section 1.6 and 1.7) and Chapter 2 (section 2.8 and 2.9). In the main, the questions addressed in this thesis were not ones that can, or will be adequately addressed by experimental studies or specially designed prospective cohort studies. However, EHR have a number of limitations that are particularly important and are linked to their use primarily as a clinician record, rather than a research tool. Some of these limitations can be overcome by study design and analysis techniques, others cannot. Overcoming the limitations of EHR can mean relying on a number of assumptions, but often
these assumptions are based on the logical thinking that to be useful to a treating clinician, data in EHR must be complete, correct, plausible, and display concordance (across records or other sources), and that entries must also be appropriately timed (Weiskopf and Weng, 2013). Additionally, we must assume that there is little risk of differential recording because of a particular state or trait, for example: GPs behave consistently and are just as likely to make a diagnosis of T2DM in the general population as they are in someone with BPD. In many instances in the studies in this thesis, this assumption will hold true. In others, such as self-harm and suicide there may be more detection and coding in people with SMI compared to the general population, with evidence that EHRs tend to under-record suicide in the general population (Thomas et al., 2013a). However, this should not be the case in the study comparing different maintenance medications (Chapter 8) as all individuals are considered high risk of suicide and rates appear similar to other cohort studies (Goodwin et al., 2003). In general, the HRs should be considered accurate, and the raw rates interpreted more cautiously.

9.5.2 Diagnostic validity of bipolar disorder in THIN

Diagnoses of BPD in THIN have not been directly validated. It has generally been considered that BPD is not a diagnosis that most GPs would make alone, and therefore the majority will be diagnosed in secondary care by a psychiatrist and will meet ICD criteria for the disorder (National Institute for Health and Care Excellence, 2014), however this heuristic approach is still vulnerable to error. I have previously shown that BPD incidence in THIN is similar to other European estimates, and shows expected age, sex and SES distributions (Hardoon et al., 2013). There may be potential for misdiagnosis, but attempts were made to minimise this by using the most recent diagnostic code in the patient record, at which point the treating clinician would have the most information about the patient’s illness course, and therefore be able to make the most accurate diagnosis. Clearly, misdiagnosis cannot be
ruled out, but this would be the case in any situation where detailed research diagnostic criteria were not rigorously applied, with the added complication of the phasic nature of the illness in the case of BPD. The opportunity for misdiagnosis in BPD is potentially greater than in other mental illness. The psychotic episodes in mania and depression could be misdiagnosed as schizophrenia, schizoaffective disorder or other psychotic illness. BPD II symptoms could be misdiagnosed as unipolar depression or emotionally unstable personality disorder. It has been reported that around 70% of BPD is initially misdiagnosed, with over one third waiting more than 10 years for a correct diagnosis (Hirschfeld and Vornik, 1899). It is not possible to separate BPD I and II in THIN, and this may mean a heterogeneous group of individuals have been examined. Additionally, there may be diagnostic biases at play that lead to clinicians making a BPD, as opposed to alternative, diagnosis. These biases may exist because of baseline characteristics of the patients. For example, there has been a longstanding clinical (and historical research) belief that non-White ethnicity and lower SES is more likely to be associated with schizophrenia than BPD (Eid et al., 2013, Garb, 1997), having these prejudices in mind, along with a heuristic that interprets specific symptom clusters as definitely delineating one diagnosis from another, will lead to misdiagnosis.

Within the mortality and morbidity studies, it is clear that, whether truly reflecting the illness state recognised as BPD, individuals who receive a code consistent with the diagnosis have a greatly elevated rate of all-cause mortality, CVD diagnosis, suicide and self-harm compared to the general population. These rates are comparable with other studies (Chapter 3) and comparable to people who receive a code consistent with a schizophrenia diagnosis (which is likely to be those at the severe end of psychotic illness, as there has been a move away from diagnosing schizophrenia specifically)(Haroon et al., 2013). Within the maintenance medication studies it is likely that receiving a BPD diagnosis and
being treated with a maintenance mood stabiliser increase the chance that the BPD diagnosis is valid.

Beyond THIN, there is some evidence supporting the validity of BPD diagnosis in UK primary care records. A small CPRD study validated renal failure in patients with BPD treated with lithium and found that all those expected to have BPD fulfilled diagnostic criteria (Close et al., 2014). In general, mental and behavioural disorders in CPRD are as well validated as other diagnoses, based on diagnostic algorithms, record checks and questionnaires sent to GPs (Herrett et al., 2010). Previously psychotic illness has been validated in primary care records (Nazareth et al., 1993). Other EHR sources also display good case validity for BPD (Kessing, 1998, Sellgren et al., 2011).

9.5.3 Comparing characteristics of my cohort of people with bipolar disorder with other cohort studies

It is difficult to compare baseline sociodemographic and clinical characteristics (Table 4.i, Table 7.i) for people with BPD in this thesis with previous cohorts from the UK, because studies do not exist or do not report specific BPD characteristics. Internal comparisons with schizophrenia suggest individuals diagnosed with BPD are more likely to be female, White ethnicity and live in higher SES areas, as reported elsewhere (Laursen et al., 2007). Smoking prevalence is similar to that in patients recruited to the Stanley Research Program in the US (Dickerson et al., 2013). Compared to Danish population data, the sex and age distributions are similar, as is the all-cause mortality rate (Medici et al., 2015). In Study 2, 3 and 4 baseline covariates were collected over specific time periods, with the most recent being used if there was more than one record. This approach minimises missing data, but may introduce bias because of failure to update records when a status changes. In these studies I would hope that this bias would be non-differential. However, there are cases where it may not be. For example, individuals commencing olanzapine may be more likely to have a
weight measure at the point of commencing the drug (because clinicians are concerned about the risk of weight gain). Whereas, if a drug with less recognised propensity to cause weight gain is commenced, weight may not be measured and therefore a more historical (i.e., inaccurate) weight may be used.

9.5.4 Defining drug exposure using EHR

Although most long-term prescribing for stable patients is likely to occur in primary care, prescriptions for mood stabilisers issued in secondary care, for example, in specialist psychiatric clinics or during inpatient care may not be recorded in THIN. Thus, drug exposure may be underestimated in these patients and may be at risk of ‘immeasurable time bias’. One previous CPRD study has suggested that lithium prescribing may be underestimated by primary care records (Close et al., 2014). However, there is no reason to believe that this would be differential by study drug, and therefore the study design used in this thesis minimises the potential impact of this problem. I also considered the issuing of a new prescription within 3 months of the previous predicted end date to signify a period of continuous prescribing to reduce the potential impact of this issue. I am aware of no EHR that has better recording of medications than those that contain records from UK primary care. Additionally, in the case of the effectiveness/tolerability study (Study 2: Chapter 6) this move to psychiatrist prescribing or hospital care would reflect treatment failure and is therefore an outcome of interest.

A larger issue is the patient’s potential non-adherence with medication. I aimed to minimise this threat to validity by only including patients with more than one prescription issued, and including a window of 3 months from the previous predicted end date in which a new prescription could be issued. Despite this, collecting a prescription can only be a proxy for taking a medication. Potentially, adherence is better in patients where medication levels
are measured, such as lithium and valproate. However, studies of cohorts with BPD have found that full and partial adherence rates are similar across groups prescribed lithium, valproate and SGAs (Sajatovic et al., 2007, Sajatovic et al., 2006). Adherence with any medication cannot be guaranteed, but non-adherence rates are similar in BPD populations and populations with other long-term health conditions (Horne et al., 2005).

9.5.5 Using drug monotherapy to investigate effects

To be able to understand the specific effects of drugs, studies 2, 3 and 4 included people prescribed monotherapy of one of lithium, valproate, olanzapine or quetiapine. Specific combinations of maintenance mood stabilisers could not be compared. This was due to relatively small numbers, the increasingly complex study design necessary, and the likely differences in patient characteristics related to adherence with a polypharmacy regimen. This study design therefore allows us to consider potential early/single drug treatment options for BPD, but may not be generalisable to situations where prescribing becomes more complicated. To put this limitation in context, in the UK (where prescribing is relatively conservative), there is evidence that for patients taking lithium, 20% take lithium alone, 45-50% take a second drug, about 30% a third and 5% a forth (Goodwin et al., 2016, Hayes et al., 2011). However, in each of these patients it is likely that they began with just one mood stabiliser prescribed; therefore it is important to understand which monotherapy has the best chance of keeping people well and the lowest risk of adverse events. In the case of adverse effects related to these drugs it remains unclear how combinations of mood stabilisers may interact.

9.5.5.1 Considering the effects of drug dosage

In addition to studying monotherapy only, I did not include drug dosage in any of the studies. This was too complex to add to this thesis, but is likely to be important for both
comparative effectiveness and adverse effect profiles. In effectiveness studies, an increase in drug dose may reflect treatment failure and in adverse effect studies, higher doses increase the risk of adverse effects.

9.5.6 Follow-up time for patients with bipolar disorder in THIN

Despite the potential for long follow-up time in THIN, the median follow-up time in the studies conducted in this thesis was relatively short. For example, in the mortality/morbidity study (Study 1: Chapter 4) patients with BPD had a median follow-up of 2.3 years and the general population comparator group 2.0 years. It is likely that this is due to more practices joining the database and coming up to agreed levels of data quality later during the study period, rather than high rates of de-registration from practices. There is some evidence that people with SMI are more likely to move accommodation than their healthy counterparts (McCarthy et al., 2007). However, this is not reflected in THIN data. Importantly, there does not appear to be differential follow-up by diagnosis (Chapter 4) or treatment group (Chapter 7, Chapter 8). In the study in Chapter 6, differential follow-up is the outcome of interest.

9.5.7 Validity of outcomes used in this thesis

The majority of outcomes examined in this thesis have been validated previously. Specific to studies examining THIN, all-cause mortality is well recorded, with >99% positive predictive value and sensitivity (Hall, 2009). Applying an algorithm for suicide gives a positive predictive value of 97% (Arana et al., 2010). CVD diagnoses have positive predictive values greater than 90% (Hammad et al., 2008, Ruigómez et al., 2010). Rates of CKD, hypertension, hypothyroidism, CVD and T2DM in THIN reflect UK national QOF prevalence (Blak et al., 2011). With regards to these outcomes in CPRD (a comparable system) the proportion of confirmed cases is: endocrine and metabolic, 88%; circulatory, 85%; and
injury and poisoning, 90% across multiple studies (Herrett et al., 2010). Despite these reassuring validation results, there remains the risk of potential misclassification when relying on algorithms rather than gold standard research criteria or death certificate information.

In a number of studies, a more fine-grained definition of outcome would have been interesting to investigate. It would have been potentially informative to separate CVD into MI (ST and non-ST elevated MI), CHD and CVE (ischemic and hemorrhagic), but even in the full cohort of individuals with BPD this would have resulted in low power, because events were rare. Similarly, it would have been interesting to separate self-harm events into overdoses and other forms of self-harm, because of the different relationship these two types of acts might have with impulsivity and aggression. Again, this was not possible because of the rarity of the events. In Chapter 8, it would have been useful to be able to look at grades of blood test abnormalities. These limitations highlight the fact that very large datasets are required to answer a number of epidemiological and particularly pharmacoepidemiological questions, for example by combining data from several routinely collected data sources.

9.5.8 Comparing bipolar disorder mortality with general population rates

As well as GPs recording death, THIN relies on information being returned from the ONS about patient mortality. The accuracy of these data is checked regularly and, as discussed in Chapter 2, only practices with AMR contributed to this thesis. Although I would not necessarily expect the mortality in the healthy population comparison group in Study 1 to reflect exactly the UK population (because they were selected to match with BPD or schizophrenia patients rather than be representative of the general population), it should be similar. As such, SMRs for THIN general population vs. ONS general population should be
approximately 1.00. To check the generalisability of mortality recording for my cohort I generated annual SMRs (accounting for age group and sex) comparing my healthy population comparison group with ONS data at the start (2000) midpoint (2007) and end (2014) of the study. In 2000, the THIN SMR was 0.97 (95% CI 0.81 to 1.16). In 2007, the SMR was 1.13 (95% CI 1.03 to 1.22). In 2014, the SMR was 0.69 (95% CI 0.63 to 0.75). This suggests that there is general agreement between these data sources. By 2014, it appears that the THIN population had lower mortality than the general UK population, but this is potentially an error based on the large number of individuals in the THIN comparison group by this time (i.e., the 95% CI is likely to be too precise).

9.5.9 Approach to handling missing data in THIN

In general, I took a simple approach to missing data throughout this thesis, in line with previous EHR studies. In Study 1, I performed a complete case analysis, where the only missing data were considered to be quintiles of Townsend score (3% missing). Less than 5% missing data is considered unlikely to produce biased estimates (Bennett, 2001). As discussed in Chapter 2 all individuals with missing ethnicity data were considered White (Hippisley-Cox et al., 2008) and all individuals with missing smoking data were coded as non-smokers (Marston et al., 2014) in line with previous research. Other confounders in this study were coded as the worst ever recorded over follow-up (BMI, hypercholesterolemia, hypertension, T2DM). In Studies 2, 3 and 4 variables were dealt with similarly, but a number were only considered present if they were recorded before baseline (drug initiation). The adverse events study (Study 3: Chapter 7), in particular, suffered from missing blood test recording, which was used to define the outcome. I attempted to overcome this limitation by conducting a number of sensitivity analyses (including IPTW). As discussed in Chapter 2, multiple imputation approaches for missing data are currently incompatible with PS analyses. However, in the case of the adverse effects study (Study 3),
using two-fold fully conditional multiple imputation (Welch et al., 2014) to impute missing blood tests and then attempting to identify the time points at which blood tests first became abnormal would have been a potentially interesting addition.

Beyond the necessary simplification of variables that are recorded in THIN, a number of other variables that would enhance the quality of the completed work are not included in the EHR. Illness and symptom severity is particularly difficult to quantify. For example, it could be argued that the observed lower rates of self-harm in individuals prescribed lithium are simply due to that group already having lower impulsivity (and therefore a clinician being happier to prescribe a potentially toxic drug). To be certain that this is not the case we would need everyone to receive a standardised measure of impulsivity, which clearly is not going to happen in clinical practice. All we can therefore draw on are covariates that are potential markers of impulsivity (for example: age, prior self-harm, smoking, alcohol and substance misuse) which suggest there are no dramatic differences by drug group.

A major limitation of THIN in general for research purposes is the lack of information about individual level SES and other social factors (such as employment, education and relationships). In all UK primary care EHRs, SES is defined by an area level measure, either the Townsend score (see Section 2.3.3 and Section 2.9.5), or Index of Multiple Deprivation (IMD). These area level measures are considered suitable proxies for individual SES, but are clearly vulnerable to ecological fallacy (i.e that the SES of the group does not accurately reflect all individual SES in the group) (Galobardes et al., 2006). Townsend score, used in this thesis, is based on 2001 census data, and so may not adequately reflect the area by the end of cohort follow-up in 2014. IMD is a more recent measure of area SES and is now available in THIN. The problem is that there are different IMDs for each country of the UK, and these cannot be combined (Payne and Abel, 2012). Use of IMD would therefore have
reduced the cohort to individuals living in England (approximately 70%). Because of this, I decided to use Townsend score despite its limitations. Other data that would enhance the included studies would be measures of health behaviours, such as exercise and healthy eating, but again these would require some form of standardised longitudinal measurement as they are not routinely coded in primary care EHRs.

9.5.9.1 How missing blood test results may have affected the results of maintenance treatment studies

In the sensitivity analyses in the adverse events study (Study 3; Chapter 7), attempts were made to minimise the effects of potential surveillance bias in blood tests using IPTW. These results were consistent with the primary analysis. However, there remains the problem that GPs will complete blood tests when indicated, rather than at random, and it is unlikely that the effect of this indication can be totally overcome by statistical approaches. Interpretation of the potential effects of this on my results is complex. For example, GP will/should routinely complete renal function tests for people taking lithium. If people taking lithium have more tests, an abnormal result (especially mildly abnormal) is more likely to be detected. However, being on olanzapine is not necessarily an indication for testing renal function, and therefore when a test is completed there is likely to be a clinical (i.e., symptom related) reason for the test being completed. If there are clinical signs, a blood test is potentially more likely to be abnormal. In the case of calcium testing the effects are even more difficult to predict, because, in spite of guidelines recommending calcium monitoring for patients taking lithium it is only the minority of this group that receive tests.

9.5.10 Unmeasured and residual confounding

It remains impossible to rule out unmeasured and residual confounding in each of the completed studies, as it does in any observational research. In the mortality/morbidity
study (Study 1: Chapter 4), this is potentially less important as the aim was to understand differences with relation to the general population accounting for a small number of potential confounders. In Studies 2, 3 and 4 the use of PS techniques goes beyond traditional multivariable regression approaches in terms of reducing confounding. For unmeasured confounders to have a dramatic impact on the results of any of these studies they would have to be strongly associated with drug-exposure and outcome and be independent of other confounders: it is unlikely that such covariates could be identified (Lin et al., 1998, Psaty et al., 1999). However to be more confident in this conclusion it may have been beneficial to have conducted an analysis to identify the strength of unmeasured or residual confounding that would be necessary to explain the observed association between drug-exposure and outcome, as suggested by Schneeweiss (Schneeweiss, 2006).

9.5.10.1 Confounding by indication

As discussed in Chapter 2 and the Potential limitations section of Chapters 6, 7, and 8, Studies 2, 3 and 4 may be at particular risk of confounding by indication. Although attempts were made to limit this via study and analysis design, it is impossible to rule out that there were unmeasured differences in patient groups that lead to treatment with one particular drug, rather than an alternative. In fact this type of bias may be an intractable threat to validity in all observational studies (Bosco et al., 2010). It then becomes vital to show that results are consistent across “different persons, places, circumstances and time” (Hill, 1965). Certainly, this appears to be the case when comparing the results of the effectiveness/tolerability study (Study 2) with RCT results and the adverse events and suicide/self-harm studies (Studies 3 and 4) with other cohort studies.
9.5.11 Are propensity score approaches superior to traditional techniques?

It has been argued that PSs are no better at managing confounding than traditional regression techniques (Biondi-Zoccai et al., 2011, Shah et al., 2005, Stukel et al., 2007). Even if this were the case, there are still fundamental benefits to PS analysis, beyond multivariable regression. The first benefit is the ability to synthesise succinctly several contributors of confounding together and to visualise the distribution of PSs across exposure statuses – this allows the researcher to understand how comparable two (or more) groups may be in terms of baseline characteristics. The second benefit is the ability to add more parameters to a model than multivariable regression and therefore adjust for more confounders without model instability. The third is their intuitive appeal as a quasi-randomised adjustment method (Biondi-Zoccai et al., 2011). Other authors still maintain that PS models outperform multivariable models. Simulation studies show that quintiles of PS give results closer to the true marginal treatment effect than logistic regression (Martens et al., 2008), and that PSs are less biased, more robust and more precise than logistic regression when the number of outcome events per confounder is low (Cepeda et al., 2003).

9.5.12 Limitations related to study design

The mortality/morbidity study (Study 1) used a classical regression approach to examine the morbidity and mortality in BPD. The other studies compared individuals treated with lithium, valproate, olanzapine or quetiapine monotherapy. Over the majority of the study period these medications were recommended as first-line treatments (National Institute for Health and Care Excellence, 2006). The main reason for designing the studies in this way was to minimise the potential for confounding by indication. Therefore I did not include a group receiving non-treatment or a group co-prescribed lithium and valproate, which would have been more likely to have underlying differences based on compliance and
illness severity (amongst other things). Additionally, at the point of designing these studies (before the NICE 2014 update (National Institute for Health and Care Excellence, 2014)), this was a real clinical dilemma facing psychiatrists and their patients: which drug to choose given that NICE considers them equal. A potential limitation of this approach, however, is that a large number of people with BPD are not prescribed monotherapy of one of these drugs, and thus their data were not included in any analysis. Additionally, to have large enough groups of individuals treated with each drug I had to start follow-up in 1995, which was before the first use of olanzapine and quetiapine for BPD (1997 and 1998 respectively) (Hayes et al., 2011). This means that clinical choice, and therefore clinical equipoise was not the same at the start and end of the study. It is unclear if clinician drug preference based on clinical presentation changed over this time; however, all of these factors (or their proxies) are included in the PS.

9.6 Bipolar disorder now and in the future

9.6.1 Is bipolar disorder becoming more common?

It has been stated that in clinical practice the diagnosis of BPD has become more commonplace, that this is due to the introduction of BPD II (a new diagnosis in DSM-IV), and perhaps in part because of the marketing of SGAs and anticonvulsants as specific treatments for BPD (Ostacher et al., 2016). This is supported by studies showing rapid increases in rates of diagnosis in young people in the US (Moreno et al., 2007). However, Global Burden of Disease data do not endorse this argument and suggest that despite evidence of an approximately 50% increase in prevalent cases worldwide between 1990 and 2013, the age-standardised prevalence remains stable at 0.7% (and all differences are explained by demographic shifts in population size and age composition) (Ferrari et al., 2016). We previously found relatively stable incident recording of BPD in THIN (Hardoon et
Despite this, globally by 2013, years lived with disability in BPD was comparable with those of more prevalent conditions such as asthma or Alzheimer’s disease (over 100 per 100,000) (Ferrari et al., 2016).

What does seem apparent is that there is a growing interest in BPD from the general public and individuals experiencing mental health problems, with a lower level of stigma associated with BPD than other SMI (Chan and Sireling, 2010), perhaps related to the idea that BPD is associated with creativity (Kyaga et al., 2013, MacCabe et al., 2010).

9.6.2 Underfunded and under researched

In both clinical and research terms, BPD has been overlooked relative to other physical and mental health problems. In general, it is recognised that mental health problems receive a tiny proportion of global annual research budgets (Young, 2006).

In relation to schizophrenia, Clement et al. found that for every one BPD publication on Medline there are 4.4 for schizophrenia. When only trials are taken into account this ratio increases to 1:7.6 (Clement et al., 2003). More recently it has been shown that investment in BPD research has a tenth of that in schizophrenia (Goodwin and Geddes, 2007). It has been argued that this disparity may be due to clinicians’ and researchers’ perception of the seriousness of BPD. Historically there has been a feeling that schizophrenia represents a greater overall disease burden on individuals, highlighted by higher hospitalisation rates, poorer global functioning and greater economic costs (Craig et al., 2000, Das Gupta and Guest, 2002, Grossman et al., 1991). However, research that is more recent has suggested that this is not the case and reduced quality of life, healthcare costs and premature mortality are similar in both disorders (Dean et al., 2004, Hoang et al., 2013).
9.7 Future questions for bipolar disorder research

9.7.1 Can early diagnosis improve functioning and reduce adverse long-term outcomes?

Systematic, timely diagnosis of BPD remains a challenge. Potentially, an accurate diagnosis early in the illness course (ideally before the first episode of mania or hypomania) could help to prevent the long-term detrimental effects of the illness and inappropriate treatment. Changes to DSM-V and ICD-11 may help this (Phillips and Kupfer, 2013), as may the increased use of clinician and patient rating scales, and more precise assessment of patients with depression for any symptoms or signs of (hypo)mania. The potential impact of each of these approaches needs to be comprehensively assessed. Alternatives to categorical diagnostic approaches, such as Research Domain Criteria (Insel et al., 2010) and dimensional approaches (such as an affective disorders continuum) (Angst, 2007) could feasibly redefine BPD in terms of pathophysiological processes (Vieta and Phillips, 2007). These approaches may enhance our understanding of the illness, suitable treatments and potential outcomes (Phillips and Kupfer, 2013). In particular, early and appropriate treatment could reduce the adverse outcomes studied in this thesis.

9.7.2 Can staging models improve outcomes?

Staging models, such as those developed for schizophrenia (incorporating prodrome, first episode and chronic phase) have only recently been proposed and developed for BPD (Grande et al., 2014, Vieta et al., 2011). Research has shown that psychosis staging is mirrored by anatomical brain and cognitive function changes and that treatment is improved by using stage specific strategies (McGorry et al., 2010). For such models to be useful in BPD, they need to include occurrence of psychiatric comorbidities, medical disorders and risk factors. However, unlike schizophrenia, onset and longitudinal course of these elements remains poorly understood (Leboyer et al., 2012). Development and use of...
staging models, alongside stage specific treatment, has the potential to improve long-term outcomes.

9.7.3 How do we better use the drugs that we have?

As discussed in this thesis, pharmacological interventions have the potential to reduce the severity, and therefore disability in BPD. However, full lifetime remission is rare (American Psychiatric Association, 2013), and it has been estimated that even in a hypothetical situation where every individual with BPD received optimised evidence-based interventions only 40% of the burden would be averted (Andrews et al., 2004). This highlights the fact that although treatment strategies exist, these are far from ideal.

There is now a body of evidence that suggests that lithium is the most effective mood stabiliser medication currently available. It has been argued that, with the exception of electroconvulsive therapy, lithium is the single most effective treatment in psychiatry (Shorter, 2009). However, naturalistic studies have shown that approximately 40% of those treated with lithium show no response, 30% will be partial responders and 30% excellent responders (Baldessarini and Tondo, 2000, Garnham et al., 2007). Currently there are no reliable ways of identifying responders early after drug initiation (Geoffroy et al., 2014). Complicating this is the reduction in use of a drug that requires clinician skill to be safely managed. Increasingly, trainees in psychiatry have become untutored in lithium use and are uncomfortable prescribing it (Healy, 2008). Whilst I hope the studies in this thesis will reassure clinicians and patients of the benefits of lithium, and that the risks are lower than previously thought, increasing the safe use of lithium is likely to require a cultural shift. This shift could be supported by increased education, systems support for monitoring of blood levels and potentially provision of tools for identifying likely responders and individuals at risk of adverse effects before they develop.
9.7.4 Could a better understanding of mechanism of lithium help develop new treatments?

Lithium is also of interest to those considering the development of new drug treatments in BPD, because it remains uniquely effective in this condition (whereas other psychiatric drug treatments are effective across a range of disorders). As such, understanding its mechanism of action may be crucially important to the identification of pharmacological targets. Detail of lithium’s biochemical effects could be the focus of a thesis in itself, but it is interesting to consider the pharmacological properties of lithium that are not present in antipsychotic or anticonvulsant medications. Lithium’s physiological effects are multiple, and it remains unclear which are vital for therapeutic efficacy. Additionally, studies of the action of lithium are also linked to studies of the pathophysiology of BPD.

As it currently stands, lithium’s cellular level effects can be grouped onto: 1) Regulation of cell membrane transport, 2) regulation of ion distribution, 3) Regulation of cell membrane properties, 4) intracellular signalling regulation and, 5) neurotransmitter regulation. These properties interact as multilevel cascades (Alda, 2015). Perhaps, more usefully, we can consider the mechanisms that appear to mediate the relationship between molecular level and clinical level effects. These have been defined as neuroprotection, chronobiology and neuronal activity stabilisation (Alda, 2015). There is growing evidence that lithium is neurotropic, and that this suppresses stress effects and restores plasticity lost through illness (Gray and McEwen, 2013, Hajek et al., 2012, Moore et al., 2000). This has also led to research into the potential positive effects of lithium in individuals with dementia (Sutherland and Duthie, 2015) and neurodegenerative diseases such as multiple sclerosis, progressive supranuclear palsy and multiple system atrophy (Saccà et al., 2013). Sleep deprivation, jet lag and time zone shifts, and daytime light levels are associated with BPD relapse. Lithium modifies these biological rhythms, although mechanisms are not well
understood (McCarthy and Welsh, 2012, Seggie et al., 1983). Finally, lithium is able to inhibit excessive neuronal activity whilst augmenting reduced activity, mostly via signal transduction and transcription factors (Jope, 1999).

Additionally, work has begun on the task of identifying more immediate proxies for good lithium response, which could be employed in future drug development. The most promising target appears to be mood instability (Geddes and Miklowitz, 2013), but studies are also collecting measures of impulsivity, physical activity, sleep, gene expression and neural dynamics during magnetoencephalography and functional magnetic resonance imaging (Saunders et al., 2016).

9.8 Other research directions

9.8.1 Study design

PS studies are accessible and easy to communicate, they are likely to become increasingly important with the further expansion and linkage of EHR, where many covariates will be available. High-dimensional PS studies, using hundreds or thousands of covariates will become possible (Schneeweiss et al., 2009) and PS models may be improved by machine learning techniques (Lee et al., 2010). However, these techniques are not without their problems. It is clear that even simpler PS approaches are not well defined (Chapter 2) and there is evidence that a wide range of estimates can result from apparently minor changes in model build and implementation strategy (Hill et al., 2011). Further statistical work that clarifies the possible impact of covariate inclusion/exclusion is necessary. Additionally, it is unclear how approaches that include as many potential confounders as possible in a PS sit with casual models which warn of the possibility of increase confounding through over adjustment (backdoor colliders) (Greenland et al., 1999). Currently there appears to be no literature on this likely contradiction or how it should be managed. In addition, there is no
agreed approach to combining multiple imputation and PS estimation. Missing data is a major problem in EHR and for full utility in research robust statistical techniques to combine these methods will be necessary. Some advances have been made in this field since I conducted the studies in this thesis (Leyrat et al., 2016, Mitra and Reiter, 2016).

Other techniques for minimising confounding in observational studies exist and these may be useful in EHR studies. Nevertheless, these techniques also have methodological limitations and are potentially more complicated to apply and communicate. Instrumental variables have been used in a number of EHR studies, with GP prescribing practice used most frequently as an instrument (associated with exposure, only associated with outcome through its effect on the exposure and not associated with any confounding factors) (Davies et al., 2013, Thomas et al., 2013b). However, there is evidence that in some situations instrumental variable analysis will be more biased than traditional multivariable techniques (Davies, 2015, Garabedian et al., 2014) and in particular clinician prescribing preference can fail to perform effectively as an instrument, as there may be unidentified instrument-outcome confounders (Garabedian et al., 2014, Kollhorst et al., 2016). Other approaches such as bootstrapping, exact methods, classification and regression tree analyses, mixed effect and Bayesian methods have been far less widely used (Biondi-Zoccai et al., 2011).

9.8.2 Electronic health records

Current EHR systems are primarily clinical in focus. This underestimates their true potential for research and advances in public health practice and policy, comparative effectiveness studies and trial recruitment and passive follow-up (Haneuse and Daniels, 2016, Kukafka et al., 2007). There may be simple advances, which make them suitable for multiple research purposes, without compromising (and potentially improving) functionality. Widespread
linkage of data sources would improve validity and provide further detail about potential confounders. Linkage to genetic and imaging databanks would be particularly powerful.

9.9 Dissemination

At the time of completion of this thesis, modified versions of the following chapters had been published:

i) A version of the meta-analysis in Chapter 3 had been published in Acta Psychiatrica Scandinavica (Appendix 4.1)

ii) A version of the mortality/morbidity study (Study 1: Chapter 4) had been published in British Journal of Psychiatry

iii) A version of the effectiveness/tolerability study (Study 2: Chapter 6) had been published in World Psychiatry (Appendix 4.2)

iv) A version of the adverse effects study (Study 3: Chapter 7) had been published in PLoS Medicine (Appendix 4.3)

v) A version of the suicide/self-harm study (Study 4: Chapter 8) had been published in JAMA Psychiatry (Appendix 4.4)

Additionally I presented elements of this research at conferences internationally:

i) A poster of the results of the meta-analysis (Chapter 3) was presented at the Royal College of Psychiatrists’ International Congress in July 2015, where I won Junior Researcher of the Year

ii) The results of Study 2 (Chapter 6) and Study 3 (Chapter 7) were presented at Eleventh International Conference of the European Network for Mental Health Service Evaluation in October 2015
iii) A poster of the results of Study 3 (Chapter 7) was presented at the 18th Annual Conference of the International Society for Bipolar Disorder in July 2016

iv) An oral presentation of the results of Study 4 (Chapter 8) was presented at the Royal College of Psychiatrists’ International Congress in July 2016, where I won the Research prize

v) An oral presentation of the results of Study 2, Study 3 and Study 4 was presented at the Royal College of Psychiatrists’ International Congress in June 2017

Dissemination to a wider audience, including the members of the public included:

i) Study 3 (Chapter 7) was the subject of an article in The Guardian (Appendix 5.1) and a news piece on BBC Radio Wales

ii) The meta-analysis in Chapter 3 was the subject of a Mental Elf blog (Appendix 5.2)

iii) Study 4 (Chapter 8) was the subject of a Mental Elf blog (Appendix 5.3)

My aim is to continue to disseminate the findings of this thesis. I have been invited to join the International Group for the Study of Lithium Treated Patients and I will use this network to build future research collaborations and communicate these results.

9.10 Conclusions

This thesis set out to determine the impact of receiving a BPD diagnosis on mortality and morbidity, and the potential for outcomes to be modified by mood stabiliser medication. Included studies provide new evidence for a widening mortality gap between individuals with BPD and the general population, and highlight the persistent elevated risk of suicide and self-harm in this population. Additionally, I found that patients with BPD, unlike those with schizophrenia have no increase in CVD deaths, but increased rates of CVD diagnoses
relative to the general population. All other causes of death are elevated in BPD, relative to the general population. These findings should be used to further the argument that parity of esteem and parity of care is not yet a reality in the UK.

Maintenance mood stabiliser medication has the potential to influence the risk of these adverse events, but before the studies in this thesis were completed the impact had not been well quantified. RCTs do not show resounding superiority of any one mood stabiliser. This thesis includes the first study to show that lithium monotherapy is potentially more effective and tolerable than valproate, olanzapine or quetiapine. I have quantified, for the first time: 1) the relative rates of renal impairment, thyroid disease and hypercalcemia; which are all elevated in individuals taking lithium relative to the other drugs examined, 2) the relative rates of severe CKD, T2DM, CVD and hepatotoxicity; which all occur at similar rates in people taking lithium, valproate, quetiapine or olanzapine, 3) the rate of hypertension; which is greater in those prescribed olanzapine, and 4) the relative rate of weight gain; which is greater in alternatives to lithium. Additionally, I examined the potential anti-suicidal effects of lithium and found that individuals prescribed this drug have lower rates of self-harm and accidental injury than patients taking other mood stabilisers. I hope that this information can be used by clinicians, in collaboration with patients, to balance the potential risks and benefits of currently available and recommended treatment options.
References


Harvey, S., Howard, J., Patel, M., Paxton, J., Ranford, W., Rogerson, S., Slade, W., Tomney, B., Van Riel, H. & Wade, J. (2012). Why are men reticent to visit their GP: what can be done to address this situation. Exeter University.


Personal correspondence with IMS Health (2016).


Reid, S. (2011). Current use of antiepileptic drugs is associated with an increased risk of suicidality in people with depression but not in people with epilepsy or bipolar disorder. *Evidence Based Mental Health* 14, 3-3.


US Food and Drug Administration (2009). Suicidal behavior and ideation and antiepileptic drugs.


Appendix

1. Scientific approval

2. Code lists for bipolar disorder and schizophrenia

3. Meta-analysis search terms
   3.1 A systematic review and meta-analysis of mortality in individuals with bipolar disorder
   3.2 Systematic review and network meta-analysis comparing the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder

4. Published papers
   4.1 A systematic review and meta-analysis of premature mortality in bipolar affective disorder
   4.2 Lithium vs. valproate vs. olanzapine vs. quetiapine as maintenance monotherapy for bipolar disorder: a population-based UK cohort study using electronic health records
   4.3 Adverse Renal, Endocrine, Hepatic, and Metabolic Events during Maintenance Mood Stabilizer Treatment for Bipolar Disorder: A Population-Based Cohort Study
   4.4 Self-harm, Unintentional Injury, and Suicide in Bipolar Disorder During Maintenance Mood Stabilizer Treatment: A UK Population-Based Electronic Health Records Study

5. Media
   5.1 The Guardian (online 14.8.2016, print 15.8.2016) – Lithium should be more widely used for bipolar disorder, researchers say
   5.2 The Mental Elf (online 2.6.15) – Premature mortality in bipolar disorder
   5.3 The Mental Elf (online 18.8.16) – Lithium for bipolar disorder: The best maintenance mood stabiliser protection against self-harm and suicide?
1. Scientific approval

SRC Feedback

Researcher Name: Joseph Hayes
Organisation: UCL
SRC Reference Number: 14-087
Date: 9th Feb 2015
Study title: Long term outcomes in bipolar affective disorder in UK Primary Care
Committee opinion: Approved

The following feedback has been supplied by the SRC.

Notes from the Chair:

Approved

We are pleased to inform that you can proceed with the study as this is now approved. CSD Medical Research will let the relevant Ethics committee know this study has been approved by the SRC.

Once the study has been completed and published, it is important for you to inform CSD Medical Research in order for us to advise the SRC and your reference number to be closed.

References to all published studies are added to our website enabling other researchers to become aware of your work. In order to identify your study as using the THIN database, we recommend that you include the words “The Health Improvement Network (THIN)” within your title. Copies of publication(s), where available, will be appreciated.

I wish you and your team all the best with the study progression.

Mustafa Dunganwalla
Research Associate
### 2. Code lists for bipolar disorder and schizophrenia

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Description</th>
<th>Read code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar disorder</td>
<td>Unspecified bipolar affective disorder</td>
<td>E117.00</td>
</tr>
<tr>
<td></td>
<td>Unspecified bipolar affective disorder, severe with psychosis</td>
<td>E117400</td>
</tr>
<tr>
<td></td>
<td>Manic psychoses</td>
<td>E11...00</td>
</tr>
<tr>
<td></td>
<td>Recurrent manic episodes, severe without mention psychosis</td>
<td>E111300</td>
</tr>
<tr>
<td></td>
<td>Mixed bipolar affective disorder</td>
<td>E116.00</td>
</tr>
<tr>
<td></td>
<td>[X] Recurrent manic episodes</td>
<td>Eu31y12</td>
</tr>
<tr>
<td></td>
<td>[X] Bipolar affective disorder, current episode mixed</td>
<td>Eu31600</td>
</tr>
<tr>
<td></td>
<td>Bipolar affective disorder, currently manic, mild</td>
<td>E114100</td>
</tr>
<tr>
<td></td>
<td>Recurrent manic episodes, partial or unspecified remission</td>
<td>E111500</td>
</tr>
<tr>
<td></td>
<td>Other and unspecified manic-depressive psychoses</td>
<td>E11y.00</td>
</tr>
<tr>
<td></td>
<td>[X] Manic episode</td>
<td>Eu30.00</td>
</tr>
<tr>
<td></td>
<td>[X] Bipolar II disorder</td>
<td>Eu31y11</td>
</tr>
<tr>
<td></td>
<td>[X] Bipolar affect disorder cur epi mild or moderate depressn</td>
<td>Eu31300</td>
</tr>
<tr>
<td></td>
<td>Unspecified bipolar affective disorder, unspecified</td>
<td>E117000</td>
</tr>
<tr>
<td></td>
<td>[X] Bipolar affect disorder cur epi manic wout psychotic symp</td>
<td>Eu31100</td>
</tr>
<tr>
<td></td>
<td>Bipolar affective disorder, currently depressed, unspecified</td>
<td>E115000</td>
</tr>
<tr>
<td></td>
<td>Mixed bipolar affective disorder, severe, without psychosis</td>
<td>E116300</td>
</tr>
<tr>
<td></td>
<td>Mixed bipolar affective disorder, moderate</td>
<td>E116200</td>
</tr>
<tr>
<td></td>
<td>Hypomanic psychoses</td>
<td>E110.11</td>
</tr>
<tr>
<td></td>
<td>[X] Hypomania</td>
<td>Eu30000</td>
</tr>
<tr>
<td></td>
<td>[X] Bipol aff disord, curr epis sev depress, no psychot symp</td>
<td>Eu31400</td>
</tr>
<tr>
<td></td>
<td>Recurrent manic episode NOS</td>
<td>E111z00</td>
</tr>
<tr>
<td></td>
<td>Manic-depressive - now depressed</td>
<td>E115.11</td>
</tr>
<tr>
<td></td>
<td>Recurrent manic episodes, unspecified</td>
<td>E111000</td>
</tr>
<tr>
<td></td>
<td>Unspecified bipolar affective disorder, NOS</td>
<td>E117z00</td>
</tr>
<tr>
<td></td>
<td>Bipolar affective disorder, currently manic, full remission</td>
<td>E114600</td>
</tr>
<tr>
<td></td>
<td>Single manic episode, moderate</td>
<td>E110200</td>
</tr>
<tr>
<td></td>
<td>[X] Other bipolar affective disorders</td>
<td>Eu31y00</td>
</tr>
<tr>
<td></td>
<td>Bipolar affective disorder, currently manic, moderate</td>
<td>E114200</td>
</tr>
<tr>
<td></td>
<td>Other mixed manic-depressive psychoses</td>
<td>E11y300</td>
</tr>
<tr>
<td></td>
<td>[X] Manic-depress psychosis, depressed type + psychotic symptoms</td>
<td>Eu33312</td>
</tr>
<tr>
<td></td>
<td>Atypical manic disorder</td>
<td>E11y100</td>
</tr>
<tr>
<td></td>
<td>[X] Bipolar affective disorder, unspecified</td>
<td>Eu31z00</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>EU Code</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar affective disorder, current episode</td>
<td>Eu31000</td>
<td></td>
</tr>
<tr>
<td>hypomaniac</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[X] Manic episode, unspecified</td>
<td>Eu30z00</td>
<td></td>
</tr>
<tr>
<td>[X] Manic stupor</td>
<td>Eu30213</td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar disorder, single manic episode</td>
<td>Eu30.11</td>
<td></td>
</tr>
<tr>
<td>Manic disorder, single episode NOS</td>
<td>E110z00</td>
<td></td>
</tr>
<tr>
<td>Single manic episode in full remission</td>
<td>E110600</td>
<td></td>
</tr>
<tr>
<td>Single manic episode in partial or unspecified</td>
<td>E110500</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, unspecified</td>
<td>E116000</td>
<td></td>
</tr>
<tr>
<td>Unspecified manic-depressive psychoses</td>
<td>E11y000</td>
<td></td>
</tr>
<tr>
<td>Unspecified bipolar affect disord, partial/unspec</td>
<td>E117500</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrent manic episodes</td>
<td>E111.00</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, now depressed, in full</td>
<td>E115600</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manic-depressive - now manic</td>
<td>E114.11</td>
<td></td>
</tr>
<tr>
<td>Single manic episode, severe without mention of</td>
<td>E110300</td>
<td></td>
</tr>
<tr>
<td>psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile of mood states, bipolar</td>
<td>ZRby100</td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disord,currently manic, part/unspec</td>
<td>E114500</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrent manic episodes, moderate</td>
<td>E111200</td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disorder, now depressed, severe with</td>
<td>E115400</td>
<td></td>
</tr>
<tr>
<td>psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[V] Personal history of manic-depressive psy</td>
<td>ZV11112</td>
<td></td>
</tr>
<tr>
<td>[X] Mania with psychotic symptoms</td>
<td>Eu30200</td>
<td></td>
</tr>
<tr>
<td>[X] Mania without psychotic symptoms</td>
<td>Eu30100</td>
<td></td>
</tr>
<tr>
<td>[X] Mania with mood-congruent psychotic symptoms</td>
<td>Eu30211</td>
<td></td>
</tr>
<tr>
<td>Recurrent manic episodes, mild</td>
<td>E111100</td>
<td></td>
</tr>
<tr>
<td>Unspecified bipolar affective disorder, in full</td>
<td>E117600</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[X] Manic-depress psychosis,depressd,no psychotic</td>
<td>Eu33213</td>
<td></td>
</tr>
<tr>
<td>symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unspecified bipolar affective disorder, severe, no</td>
<td>E117300</td>
<td></td>
</tr>
<tr>
<td>psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H/O: manic depressive disorder</td>
<td>146D.00</td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disorder, currently manic, severe with</td>
<td>E114400</td>
<td></td>
</tr>
<tr>
<td>psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disorder, now depressed, part/unspec</td>
<td>E115500</td>
<td></td>
</tr>
<tr>
<td>remission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar affect disorder cur epi manic with</td>
<td>Eu31200</td>
<td></td>
</tr>
<tr>
<td>psychotic symp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently depressed</td>
<td>E115.00</td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, NOS</td>
<td>E116z00</td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, severe, with</td>
<td>E116400</td>
<td></td>
</tr>
<tr>
<td>psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrent manic episodes, in full remission</td>
<td>E111600</td>
<td></td>
</tr>
<tr>
<td>[X] Mania with mood-incongruent psychotic symptoms</td>
<td>Eu30212</td>
<td></td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>Code</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently manic, unspecified</td>
<td>E114000</td>
<td></td>
</tr>
<tr>
<td>[X] Other manic episodes</td>
<td>Eu30y00</td>
<td></td>
</tr>
<tr>
<td>Other and unspecified manic-depressive psychoses NOS</td>
<td>E11yz00</td>
<td></td>
</tr>
<tr>
<td>Recurrent manic episodes, severe, with psychosis</td>
<td>E111400</td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disorder, currently manic, severe, no psychosis</td>
<td>E114300</td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar affect disorder, currently manic, severe, with psychosis</td>
<td>Eu31500</td>
<td></td>
</tr>
<tr>
<td>Unspecified bipolar affective disorder, mild</td>
<td>E117100</td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, partial/unspecified remission</td>
<td>E116500</td>
<td></td>
</tr>
<tr>
<td>Single manic episode, unspecified</td>
<td>E110000</td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, in full remission</td>
<td>E116600</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently manic</td>
<td>E114.00</td>
<td></td>
</tr>
<tr>
<td>Manic disorder, single episode</td>
<td>E110.00</td>
<td></td>
</tr>
<tr>
<td>[X] Manic-depressive illness</td>
<td>Eu31.11</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently depressed, NOS</td>
<td>E115z00</td>
<td></td>
</tr>
<tr>
<td>Mixed bipolar affective disorder, mild</td>
<td>E116100</td>
<td></td>
</tr>
<tr>
<td>[V] Personal history of manic-depressive psychosis</td>
<td>ZV11111</td>
<td></td>
</tr>
<tr>
<td>Unspecified bipolar affective disorder, moderate</td>
<td>E117200</td>
<td></td>
</tr>
<tr>
<td>[X] Manic-depressive reaction</td>
<td>Eu31.13</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently manic, NOS</td>
<td>E114z00</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently depressed, mild</td>
<td>E115100</td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar affective disorder, currently in remission</td>
<td>Eu31700</td>
<td></td>
</tr>
<tr>
<td>Bipolar affective disorder, currently depressed, moderate</td>
<td>E115200</td>
<td></td>
</tr>
<tr>
<td>Bipolar affect disorder, now depressed, severe, no psychosis</td>
<td>E115300</td>
<td></td>
</tr>
<tr>
<td>Bipolar psychoses</td>
<td>E11..11</td>
<td></td>
</tr>
<tr>
<td>Single manic episode, mild</td>
<td>E110100</td>
<td></td>
</tr>
<tr>
<td>Single manic episode, severe, with psychosis</td>
<td>E110400</td>
<td></td>
</tr>
<tr>
<td>[X] Bipolar affective disorder</td>
<td>Eu31.00</td>
<td></td>
</tr>
<tr>
<td>[X] Mania NOS</td>
<td>Eu30z11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schizophrenia</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute exacerbation subchronic schizophrenia</td>
<td>E107300</td>
</tr>
<tr>
<td>Acute exacerbation of chronic paranoid schizophrenia</td>
<td>E103400</td>
</tr>
<tr>
<td>Schizophrenic language</td>
<td>ZS7C611</td>
</tr>
<tr>
<td>Cyclic schizophrenia</td>
<td>E107.11</td>
</tr>
<tr>
<td>[X] Hebephrenic schizophrenia</td>
<td>Eu20100</td>
</tr>
<tr>
<td>[X] Schizophreniform psychosis NOS</td>
<td>Eu20y13</td>
</tr>
<tr>
<td>[X] Schizophreniform disorder NOS</td>
<td>Eu20y12</td>
</tr>
<tr>
<td>Latent schizophrenia</td>
<td>E105.00</td>
</tr>
<tr>
<td>Acute exacerbation of subchronic hebephrenic schizophrenia</td>
<td>E101300</td>
</tr>
<tr>
<td>[X] Pseudoneurotic schizophrenia</td>
<td>Eu21.16</td>
</tr>
<tr>
<td>[X] Schizophrenic reaction</td>
<td>Eu23214</td>
</tr>
<tr>
<td>[X] Schizotypal personality disorder</td>
<td>Eu21.18</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Code</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Hebephrenic schizophrenia in remission</td>
<td>E101500</td>
</tr>
<tr>
<td>Other schizophrenia</td>
<td>E10y.00</td>
</tr>
<tr>
<td>[X]Paraphrenia - late</td>
<td>Eu22013</td>
</tr>
<tr>
<td>[V]Personal history of schizophrenia</td>
<td>ZV11000</td>
</tr>
<tr>
<td>Latent schizophrenia in remission</td>
<td>E105500</td>
</tr>
<tr>
<td>Other schizophrenia NOS</td>
<td>E10yz00</td>
</tr>
<tr>
<td>Schizophrenic disorders</td>
<td>E10..00</td>
</tr>
<tr>
<td>Simple schizophrenia NOS</td>
<td>E100z00</td>
</tr>
<tr>
<td>[X]Simple schizophrenia</td>
<td>Eu20600</td>
</tr>
<tr>
<td>Simple schizophrenia</td>
<td>E100.00</td>
</tr>
<tr>
<td>[X]Undifferentiated schizophrenia</td>
<td>Eu20300</td>
</tr>
<tr>
<td>Paranoid schizophrenia NOS</td>
<td>E103z00</td>
</tr>
<tr>
<td>[X]Prepsychotic schizophrenia</td>
<td>Eu21.14</td>
</tr>
<tr>
<td>Chronic latent schizophrenia</td>
<td>E105200</td>
</tr>
<tr>
<td>[X]Mixed schizophrenic and affective psychosis</td>
<td>Eu25212</td>
</tr>
<tr>
<td>[X]Disorganised schizophrenia</td>
<td>Eu20111</td>
</tr>
<tr>
<td>Chronic paranoid schizophrenia</td>
<td>E103200</td>
</tr>
<tr>
<td>Paranoid schizophrenia</td>
<td>E103.00</td>
</tr>
<tr>
<td>[X]Pseudopsychopathic schizophrenia</td>
<td>Eu21.17</td>
</tr>
<tr>
<td>Latent schizophrenia NOS</td>
<td>E105z00</td>
</tr>
<tr>
<td>Unspecified latent schizophrenia</td>
<td>E105000</td>
</tr>
<tr>
<td>[X]Restzustand schizophrenic</td>
<td>Eu20512</td>
</tr>
<tr>
<td>[X]Latent schizophrenia</td>
<td>Eu21.13</td>
</tr>
<tr>
<td>Schizophrenia NOS</td>
<td>E10z.00</td>
</tr>
<tr>
<td>H/O: schizophrenia</td>
<td>1464</td>
</tr>
<tr>
<td>[X]Paranoid schizophrenia</td>
<td>Eu20000</td>
</tr>
<tr>
<td>[X]Residual schizophrenia</td>
<td>Eu20500</td>
</tr>
<tr>
<td>[X]Schizophrenia, schizotypal and delusional disorders</td>
<td>Eu2...00</td>
</tr>
<tr>
<td>[X]Schizophrenic catalepsy</td>
<td>Eu20212</td>
</tr>
<tr>
<td>Atypical schizophrenia</td>
<td>E10y000</td>
</tr>
<tr>
<td>[X]Schizophrenic catatonia</td>
<td>Eu20213</td>
</tr>
<tr>
<td>Acute exacerbation of subchronic latent schizophrenia</td>
<td>E105300</td>
</tr>
<tr>
<td>Subchronic paranoid schizophrenia</td>
<td>E103100</td>
</tr>
<tr>
<td>Acute exacerbation of chronic catatonic schizophrenia</td>
<td>E102400</td>
</tr>
<tr>
<td>[X]Cycloid psychosis with symptoms of schizophrenia</td>
<td>Eu23112</td>
</tr>
<tr>
<td>[X]Other schizophrenia</td>
<td>Eu20y00</td>
</tr>
<tr>
<td>[X]Chronic undifferentiated schizophrenia</td>
<td>Eu20511</td>
</tr>
<tr>
<td>[X]Borderline schizophrenia</td>
<td>Eu21.12</td>
</tr>
<tr>
<td>Cenesthopathic schizophrenia</td>
<td>E10y.11</td>
</tr>
<tr>
<td>[X]Catatonic schizophrenia</td>
<td>Eu20200</td>
</tr>
<tr>
<td>Hebephrenic schizophrenia</td>
<td>E101.00</td>
</tr>
<tr>
<td>Subchronic schizophrenia</td>
<td>E100100</td>
</tr>
<tr>
<td>[X]Atypical schizophrenia</td>
<td>Eu20311</td>
</tr>
<tr>
<td>[X]Schizophreniform psychosis, manic type</td>
<td>Eu25012</td>
</tr>
<tr>
<td>Acute exacerbation of chronic latent schizophrenia</td>
<td>E105400</td>
</tr>
<tr>
<td>Acute exacerbation of chronic hebephrenic schizophrenia</td>
<td>E101400</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Code</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Prodromal schizophrenia</td>
<td>Eu21.15</td>
</tr>
<tr>
<td>Catatonic schizophrenia in remission</td>
<td>E102500</td>
</tr>
<tr>
<td>Latent schizophrenic reaction</td>
<td>Eu21.11</td>
</tr>
<tr>
<td>Schizophrenia, unspecified</td>
<td>Eu20z00</td>
</tr>
<tr>
<td>Paraphrenic schizophrenia</td>
<td>Eu20011</td>
</tr>
<tr>
<td>Restzustand - schizophrenia</td>
<td>E106.11</td>
</tr>
<tr>
<td>Acute exacerbation of subchronic paranoid schizophrenia</td>
<td>E103300</td>
</tr>
<tr>
<td>Acute exacerbation of subchronic catatonic schizophrenia</td>
<td>E102300</td>
</tr>
<tr>
<td>Acute schizophrenic episode</td>
<td>E104.00</td>
</tr>
<tr>
<td>Coenesthopathic schizophrenia</td>
<td>E10y100</td>
</tr>
<tr>
<td>Schizo-affective schizophrenia NOS</td>
<td>E107z00</td>
</tr>
<tr>
<td>Catatonic schizophrenia NOS</td>
<td>E102z00</td>
</tr>
<tr>
<td>Acute exacerbation of subchronic schizophrenia</td>
<td>E100300</td>
</tr>
<tr>
<td>Cyclic schizophrenia</td>
<td>Eu25211</td>
</tr>
<tr>
<td>Chronic schizophrenic</td>
<td>E100200</td>
</tr>
<tr>
<td>Cenesthopathic schizophrenia</td>
<td>Eu20y11</td>
</tr>
<tr>
<td>Schizophrenic flexibilatis cerea</td>
<td>Eu20214</td>
</tr>
<tr>
<td>Residual schizophrenia</td>
<td>E106.00</td>
</tr>
<tr>
<td>Unspecified hebephrenic schizophrenia</td>
<td>E101000</td>
</tr>
<tr>
<td>Hebephrenic schizophrenia NOS</td>
<td>E101z00</td>
</tr>
<tr>
<td>Schizophreniform psychosis, depressive type</td>
<td>Eu25112</td>
</tr>
<tr>
<td>Catatonic schizophrenia</td>
<td>E102.00</td>
</tr>
<tr>
<td>Paranoid schizophrenia in remission</td>
<td>E103500</td>
</tr>
<tr>
<td>Subchronic catatonic schizophrenia</td>
<td>E102100</td>
</tr>
<tr>
<td>Unspecified paranoid schizophrenia</td>
<td>E103000</td>
</tr>
<tr>
<td>Unspecified schizophrenia</td>
<td>E100000</td>
</tr>
<tr>
<td>Unspecified catatonic schizophrenia</td>
<td>E102000</td>
</tr>
<tr>
<td>Schizophrenia in remission</td>
<td>E100500</td>
</tr>
<tr>
<td>Chronic catatonic schizophrenia</td>
<td>E102200</td>
</tr>
<tr>
<td>Post-schizophrenic depression</td>
<td>Eu20400</td>
</tr>
<tr>
<td>Chronic hebephrenic schizophrenia</td>
<td>E101200</td>
</tr>
<tr>
<td>Subchronic hebephrenic schizophrenia</td>
<td>E101100</td>
</tr>
<tr>
<td>Schizophrenia simplex</td>
<td>Eu20.00</td>
</tr>
<tr>
<td>Acute exacerbation of chronic schizophrenia</td>
<td>E100400</td>
</tr>
<tr>
<td>Subchronic latent schizophrenia</td>
<td>E105100</td>
</tr>
</tbody>
</table>
3. Meta-analysis search terms

3.1 A systematic review and meta-analysis of mortality in individuals with bipolar disorder

Search in PsychINFO, Medline and EMBASE on 30.7.14

**All MeSH terms auto exploded**

**Medline (OVID version)**
Exposure: BIPOLAR DISORDER (MeSH) or “bipolar illness” “manic depression” “bipolar disorder” “bipolar affective disorder” “manic” “mania” “hypomania” (keyword)
AND
Outcome: MORTALITY, LIFE EXPECTANCY, “CAUSE OF DEATH”, DEATH (MeSH) or “life expectancy” “mortality” “death” “mortality rate” “standardised mortality ratio” “standardized mortality ratio” (keyword)
N=699

**Embase (including Embase classic)**
Exposure: BIPOLAR DISORDER, BIPOLAR MANIA, MANIA, “MIXED MANIA AND DEPRESSION” (MeSH) or “bipolar illness” “manic depression” “bipolar disorder” “bipolar affective disorder” “mania” “hypomania” “manic” (keyword)
AND
Outcome: MORTALITY, LIFE EXPECTANCY, DEATH, STANDARDIZED MORTALITY RATIO, (MeSH) or “life expectancy” “mortality” “death” “mortality rate” “standardised mortality ratio” “standardized mortality ratio” (keyword)
N=611

**Psychinfo**
Exposure: BIPOLAR DISORDER, MANIA, HYPOMANIA (MeSH) or “bipolar illness” “manic depression” “bipolar disorder” “bipolar affective disorder” “manic” “mania” “hypomania” (keyword)
AND
Outcome: MORTALITY, LIFE EXPECTANCY, “DEATH AND DYING”, MORTALITY RATE (MeSH) or “life expectancy” “mortality” “death” “standardised mortality ratio” “standardized mortality ratio” (keyword)
N=3196

After combining databases and removal of duplicates N=3522

3.2 Systematic review and network meta-analysis comparing the effectiveness and tolerability of lithium, valproate, olanzapine and quetiapine as maintenance medication in bipolar disorder

Search in the Cochrane Central Register of Controlled Trials on 9.4.2015:

#1 MeSH descriptor: [Bipolar Disorder] explode all trees (N=1624)

#2 (((bipolar or bi?polar or bi polar) near/5 (disorder* or depress*)) or ((cyclothymi* or rapid or ultradian) near/5 cycl*) or hypomani* or mania* or manic* or "mixed episode*" or rcbd):ti or (((bipolar or bi?polar or bi polar) near/5 (disorder* or depress*)) or ((cyclothymi* or rapid or ultradian) near/5 cycl*) or hypomani* or mania* or manic* or "mixed episode*" or rcbd):ab (N=3499)

#3 #1 or #2 (N=3865)

#4 (continuation or "long term" or maintenance or prevent* or prophylactic or prophylaxis or recurrence or relapse or relapses):ti or ("long term" or maintenance or prevent* or prophylaxis or recurrence or relapse or relapses):ab (N=153959)

#5 ("valproic ac* or "2 propylpentanoate" or "2 propylpentanoic acid" or "2 propylpentanoic acid" or "2 propylvalerate sodium" or "2 propylvaleric acid" or "2 propylvaleric acid sodium" or "alpha propylvaler*" or apilepsin* or convulex or convulsofin* or depon or depakone or depakene or depakine* or depakote or deprakin* or "di n propylacetat*" or "di n propylacetat* sodium" or "di n propylacetic acid" or "dipropyl acetate" or "dipropyl acetic acid" or dipropylacetat* or dipropylacetatic or diprosin* or divalproate or divalproex or epilim or epival or ergenyl or everiden* or goilim or labazen*
or leptilan* or leptilanil* or mylproin* or "myproic acid" or "n dipropylacetic acid" or orfiril
or orlept or propymal* or "sodium 2 propylpentanoat*" or "sodium 2 propylvalerat*" or
"sodium di n propyl acetate" or "sodium di n propylacetat*" or "sodium dipropyl acetate"
or "sodium dipropylacetate" or "sodium n dipropylacetate" or valerin* or valparin* or
valpro or valproate or vupral) (N=1948)

#6 (olanzapin* or lanzac or midax or olansek or olzapin or rexapin or zalasta or zolafren or
zydis or zypadhera or zyprex*) (N=2560)

#7 (lithium* or camcolit or candamid* or carbolith or carbolitium or cibalith or contemnol*
or dilithium or eskalith or hypnorex or "li salt" or limas or linthane or liskonium or liskonum
or litarex or lithane or lithiofor or lithionit or lithiophor or lithobid or lithocarb or lithonate
or lithotabs or maniprex or mesin or micalith or neurolepsin or neurolithium or plenur or
priadel or quilinormretard or quilonorm or quilonum or teralithe or theralite or
teralithe):ti,ab,kw (N=2088)

#8 (quetiapine* or ketipinor or quelin or seroquel or tienapin*).ti,ab. (N=1310)

#9 #3 and #4 (N=835)

#10 #5 or #6 or #7 or #8 (N=6600)

#11 #9 and #10 (N=458)

After removal of non-RCTs, N=396, After removal of duplicates N= 382
4.3 Adverse Renal, Endocrine, Hepatic, and Metabolic Events during Maintenance Mood Stabilizer Treatment for Bipolar Disorder: A Population-Based Cohort Study

Abstract

Background

There is limited, poorly characterized information about adverse events occurring during maintenance treatment of bipolar disorder. We aimed to determine adverse event rates during treatment with lithium, valproate, olanzapine, and quetiapine.

Methods and Findings

We conducted a propensity score adjusted cohort study using nationally representative United Kingdom electronic health records from January 1, 1995, until December 31, 2013. We included patients who had a diagnosis of bipolar disorder and were prescribed lithium (n = 2148), valproate (n = 1670), olanzapine (n = 1477), or quetiapine (n = 1376) as maintenance mood stabilizer treatment. Adverse outcomes were chronic kidney disease, thyroid disease, hypercholesterolemia, weight gain, hypertension, type 2 diabetes mellitus, cardiovascular disease, and hepatotoxicity. The propensity score included important demographic, physical health, and mental health predictors of drug treatment allocation. The median duration of drug treatment was 1.48 y (interquartile range 0.64–3.43). Compared to patients prescribed lithium, those taking valproate, olanzapine, and quetiapine had reduced rates of chronic kidney disease stage 3 or more severe, following adjustment for propensity score, age, and calendar year, and accounting for clustering by primary care practices (valproate hazard ratio [HR] 0.56; 95% confidence interval [CI] 0.45–0.69, p < 0.001, olanzapine HR 0.57; 95% CI 0.45–0.71; p < 0.001, quetiapine HR 0.62; 95% CI 0.47–0.80; p = 0.001).

Hypothyroidism was reduced in those taking valproate (HR 0.60; 95% CI 0.40–0.89; p = 0.012) and olanzapine (HR 0.48; 95% CI 0.29–0.77; p = 0.003), compared to those taking lithium. Rates of new onset hypercholesterolemia (valproate HR 0.24; 95% CI 0.09–0.61; p = 0.003, olanzapine HR 0.31; 95% CI 0.13–0.73; p = 0.007) and hypercalcemia (valproate...
Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: AIT, alanine transaminase; AST, aspartate aminotransferase; BMI, body mass index; BPA, bipolar disorder; CVD, chronic kidney disease; CVA, cardiovascular disease; eGFR, estimated glomerular filtration rate; ICD-10, International Statistical Classification of Diseases and Related Health Problems; HR, ischemic heart disease; PW, inverse probability weights; YH, myocardial infarction; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; PS, propensity score; T2DM, type 2 diabetes mellitus; THIR, The Health Improvement Research Network; TSH, thyroid-stimulating hormone

HR 0.25; 95% CI 0.10–0.80; p = 0.002, olanzapine HR 0.32; 95% CI 0.14–0.76; p = 0.008, quetiapine HR 0.23; 95% CI 0.07–0.73; p = 0.013) were also reduced relative to lithium. However, rates of greater than 15% weight gain on valproate, olanzapine, and quetiapine were higher (valproate HR 1.62; 95% CI 1.31–2.01; p < 0.001, olanzapine HR 1.84; 95% CI 1.47–2.30; p < 0.001, quetiapine HR 1.67; 95% CI 1.24–2.20; p < 0.001) than in individuals prescribed lithium, as were rates of hypertension in the olanzapine-treated group (HR 1.41, 95% CI 1.06–1.87; p = 0.017). We found no significant difference in rates of chronic kidney disease stage 4 or more severe, type 2 diabetes mellitus, cardiovascular disease, or hypertoxicity. Despite estimates being robust following sensitivity analyses, limitations include the potential for residual confounding and ascertainment bias and an inability to examine dose effects.

Conclusions
Lithium use is associated with more renal and endocrine adverse events but less weight gain than commonly used alternative mood stabilizers. Risks need to be offset with the effectiveness and anti-suicidal benefits of lithium and the potential metabolic side effects of alternative treatment options.

Author Summary

Why Was This Study Done?

- Although side effects of medications used for maintenance mood stabilizer treatment in bipolar disorder have been described, the relative risks of each drug are poorly understood.
- Randomised, controlled trials are unlikely to be able to give us this information because they include too few patients, with too short follow-up periods.
- We aimed to examine the rates of a number of adverse effects: renal failure, thyroid disease, hypercalcaemia, hepatotoxicity, weight gain, type 2 diabetes mellitus, hypertension, and cardiovascular disease in patients taking the most commonly prescribed mood stabilizers in the United Kingdom lithium, valproate, olanzapine, and quetiapine.

What Did the Researchers Do and Find?

- We conducted a cohort study of individuals with bipolar disorder prescribed lithium (n = 2148), valproate (n = 1670), olanzapine (n = 1477), or quetiapine (n = 1376), accounting for key predictors of treatment assignment, such as mental and physical health history.
- We found that renal failure was more common in patients taking lithium (approximately 9 in 100 person-years at risk) than in those taking the other drugs, but that rates of severe renal failure were similar in all groups (approximately 1 in 100 person-years at risk).
• The rate of thyroid disease was elevated in people taking lithium, compared to valproate and olanzapine, but not quetiapine.

• The rate of hypercalcemia was increased with lithium compared to all other drugs.

• The rates of greater than 7% and greater than 15% weight gain were significantly higher in individuals taking valproate, olanzapine, or quetiapine compared to lithium (for example, the rate of greater than 15% weight gain in those taking olanzapine was 0 in 100 person years at risk).

• The rate of new onset hypertension was higher in those treated with olanzapine, compared to patients treated with lithium.

• We found no statistically significant differences in rates of type 2 diabetes mellitus, cardiovascular disease, or hepatotoxicity.

What Do These Findings Mean?

• These findings highlight and quantify the relative risks of commonly used maintenance mood stabilizer treatments.

• The results for renal failure suggest that despite increasing rates of reduced renal function in those taking lithium compared to other drugs, severe renal failure is rare. We did not find differences in severe renal failure rates, which is in keeping with previous research. However, it is possible that this is due to the small number of people developing severe renal failure in each group.

• Although we did not find differences in rates of diabetes or cardiovascular disease, this may be due to relatively short follow-up times and the rarity of these adverse events. Weight gain is a significant risk factor for these health problems and was dramatically elevated in alternatives to lithium.

• Avidous monitoring of patients prescribed lithium should ameliorate some risk associated with effects on renal physiology and endocrine systems. Calcium monitoring was rare in this representative cohort of patients with bipolar disorder in UK primary care, and this needs to be improved in clinical practice.

• The potential adverse effects of each of these drugs need to be balanced with their individual therapeutic benefits, and we hope that our findings will permit informed, collaborative discussions with patients.

Introduction

Bipolar disorder (BPD) is a complex, recurrent, severe mental illness that affects over 350 million people worldwide [1]. Individuals with BPD will often require long-term drug treatment with the aim of preventing relapse or recurrence [2]. Much of the evidence for maintenance medication comes from relatively short-term randomized controlled trials, which are then extrapolated to longer-term use [3]. However, this fails to take into account the potential longer-term adverse effects of the recommended medications. In 2014, the update of the United
Kingdom National Institute for Health and Care Excellence (NICE) guidelines [1], a meta-analysis [4], and a network meta-analysis [5] all suggested that lithium should be seen as first-line monotherapy, whereas previous guidelines from around the world also recommended valproate, lamotrigine, carbamazepine, olanzapine, quetiapine, aripiprazole, ondansetron, and risperidone injection [6,7]. Prescribing in the UK has reflected the previous NICE guidance for first-line treatment [8], with lithium, valproate, olanzapine, and quetiapine being the most frequently prescribed maintenance treatments [9].

A number of adverse effects of lithium have been identified since its use as a mood stabilizer became established in the 1970s [10], but it is only recently that they have begun to be characterized and quantified [11-15]. Lithium’s adverse effects include renal, thyroid, and parathyroid dysfunction. Lithium is also recognized to cause weight gain, but the risk of weight gain relative to other potential maintenance therapies has not been widely investigated [11]. Alternatives, such as second-generation antipsychotics and valproate, have been found to be obeseogenic [16], especially olanzapine, which is the most commonly prescribed antipsychotic in BPD [9]. Weight gain is associated with a number of adverse events, such as hypertension, type 2 diabetes mellitus (T2DM), and cardiovascular disease (CVD) [17]. Valproate, olanzapine, and quetiapine are metabolized by the liver. Valproate has been found to be associated with a high risk of asymptomatic elevated transaminases and can cause idiosyncratic hepatic failure [15,18]. Olanzapine and quetiapine have also been associated with rare cases of hepatotoxicity [19-21]. Therefore, the balance of risks associated with maintenance mood stabilizer selection is not straightforward, and we are aware of no studies that make these comparisons across treatment options.

This study used a large electronic patient record database to compare rates of major recognized adverse outcomes amongst individuals prescribed lithium, valproate, olanzapine, or quetiapine for mood stabilization in BPD. The adverse events examined were chronic kidney disease (CKD), hypothyroidism, hyperthyroidism, hypercalcaemia, weight gain, hypertension, T2DM, CVD, and hepatotoxicity [15,18].

Methods

Study Design

A population-based longitudinal cohort from January 1, 1995, to December 31, 2013.

Setting

The Health Improvement Network (THIN) is a UK primary care database that contains anonymized patient information from routine clinical consultations [22]. The National Health Service (NHS) South-East Multicentre Research Ethics Committee approved THIN’s provision of anonymous patient data to researchers in 2005. Scientific approval for this study was obtained from the data provider’s Scientific Review Committee in March 2015.

THIN contained records of over 11 million people at the time of cohort extraction [22]. Included patients are broadly representative of the UK population, and physicians contributing data are representative in terms of consultation and prescribing statistics [23,24]. Approximately 98% of the UK population is registered with a primary care physician [25]. The incidence rate of BPD in THIN has been shown to be similar to European cohorts [26], and the validity of severe mental illness diagnoses held in primary care has been established [27]. NICE guidance recommends that any patient with suspected BPD should be referred to a psychiatrist for diagnosis and treatment planning [8]. Therefore, individuals in this cohort (psychiatrist-diagnosed BPD plus appropriate mood stabilizer treatment) are considered to fulfill
International Statistical Classification of Diseases and Related Health Problems (ICD-10) criteria for BFD.

In THEN, physicians use Read codes, a hierarchical coding system, to record information [28]. These codes include diagnoses made in primary and hospital care (which map onto ICD-10 codes), symptoms, examination findings, information from specialists, and test results [28]. In the UK, primary care physicians are responsible for drug prescriptions issued within the NHS, so this information is also complete and well recorded [30]. CKD, thyroid disease, T2DM, hypertension, CVD, and other chronic health condition diagnoses have been validated in THIN [23].

Participants

Patients with a diagnosis of BFD were included if they had at least one 28-day prescription of lithium, valproate, olanzapine, or quetiapine after January 1, 1995 or after the date at which the medical records met quality assurance criteria for data entry (based on computer usage and mortality recording rates [13,32]). Patients were excluded if they were prescribed another study drug at the start of follow-up or in the month before this. Diagnosis of BFD could occur at any time in the patient record. For each outcome requiring hematological or biochemical confirmation for diagnosis (CKD, thyroid disease, hypercalcemia, hepatotoxicity), patients were excluded from the primary analysis if they did not receive a specific blood test for the outcome, to reduce surveillance bias. For the weight gain outcome, patients were excluded if they did not have a baseline or pre-treatment weight and at least one other weight measurement. For the outcome of hyperthyroidism, patients taking thyroid hormone suppression [23]. Patients were also excluded if they had the outcome of interest at baseline (as we were interested in incident events). Therefore, each outcome has a different number of patients included.

Exposure

Date of first prescription was taken as the start of exposure time. The end of the prescription was calculated from the amount prescribed and dosage instructions coded by the physician. Patients were considered to have a period of continuous prescribing if another prescription for the drug was issued within 3 mo of the calculated end date. If this did not occur, the date of stopping the study drug was the end date of the final prescription. Three mo was added to this end date to account for late development of the adverse event or delayed recording. Each patient could only contribute exposure time to one of the study drugs (the first they received) and did not re-enter the cohort if they restarted the drug after more than 3 mo. Patients could be prescribed other psychiatric medications but not combinations of the study drugs. If they commenced another study drug, their outcomes were censored in the analysis (to ensure the outcome could be assigned to a particular drug).

Main Outcomes

All outcomes were defined by appropriate Read codes and/or lab results. Outcomes of interest were: CKD stage 3 or above (an estimated glomerular filtration rate [eGFR] of <60 ml/min/1.73 m²), CKD stage 4 or above (an eGFR <30 ml/min/1.73 m²) [32,33] if eGFR was unavailable we calculated it from available creatinine blood tests using the CKD-EPI equation [26], hyperthyroidism (a TSH of >10 mU/L), hyperthyroidism (or a TSH <0.1 mU/L) [23], hypercalcemia (adjusted calcium >2.65 mmol/L) [23], >7% and >13% weight gain from baseline [23], hypertension, T2DM (or HbA1c >48 mmol/mol) [23], CVD (defined as any ischemic heart disease [IH] or myocardial infarction [MI] or cerebrovascular event [CVE]), and

PLOS Medicine | DOI:10.1371/journal.pmed.1002059 August 2, 2016 5/18
Propensity Score Estimation Using Observed Pre-treatment Variables

A number of baseline patient characteristics were extracted from THIN. Physical and mental health conditions were considered present if referenced in patient notes and absent if they were not. If a patient had multiple entries of the same (or similar) codes, the start date of the condition was taken as the earliest date of entry.

A propensity score (PS) for each individual was estimated using variables defined a priori, based on existing research and clinical experience of factors influencing prescribing choice [44,45]. The PS is the conditional probability of receiving one study drug rather than another, given the variables included in the model [42,43]. Included variables were: sex, age at start of treatment with the study drug, year of entry to the cohort, ethnicity (grouped as White, Black, Asian, Mixed, other), with missing values coded as White [44], IHD diagnosis before baseline, history of MI, history of CVE, hypertension, CKD at baseline (defined by Read code or blood test), history of hypothyroidism (defined by Read code or blood test), history of liver disease or hepatotoxicity (defined by Read code or blood test), T2DM (defined by Read code or blood test), epilepsy, alcohol use (grouped as none/low, moderate, high/dependent), history of illicit drug use, smoking status (grouped as never-smoker, ex-smoker, current smoker), body mass index (BMI) (grouped as healthy weight, overweight [BMI 25 to 30], obese [BMI over 30]), anxiety symptoms or diagnosis before baseline, depressive symptoms or diagnosis, sleep disturbance before baseline, treatment with one of the study drugs at or before baseline, and clustering by practice in which the treating physician was working. The PS was checked by comparison of covariate balance across treatments, within strata. The variables in the PS excluded the outcome variable for that particular analysis. Although PS estimation cannot remove all bias, it has been postulated to also reduce confounding from unmeasured covariates, because of their association with measured variables [45–47]. In this way, use of a PS aims to replicate a randomized experiment as closely as possible by obtaining treatment groups with similar covariate distributions [48].

Statistical Analysis

Cox regression analyses were conducted, comparing the rates of adverse events in the four treatment groups. The proportional hazards model was tested formally with analysis of Schoenfeld residuals [49]. The PS was calculated using multinomial logistic regression, using drug treatment as the dependent variable and the covariates described as independent variables. The PS was then used as a linear term in a Cox regression analysis that also included age, calendar year, and clustering by practice [50]. In all cases, this model was shown to be superior to stratifying on PS using Akaike information criterion and Bayesian information criterion [41], and was a more efficient use of data than PS matching (because no patients were excluded). To account for the competing risk of each outcome with death, we plotted graphs of cumulative incidence function, adjusted for PS and age, following competing-risks regression [50,51]. We conducted sensitivity analyses in which individuals who did not receive blood tests or weight measurements were not dropped from the cohort, and in which individuals were assigned inverse probability weights (IPW) based on how likely they were to have blood test or
Results

For each outcome, 6,671 individuals with BPD diagnosis were potentially included in the analysis, 3,448 prescribed lithium, 1,670 prescribed valproate, 1,477 prescribed olanzapine, and 1,376 prescribed quetiapine (see SI Text). The median duration of drug treatment was 1.48 years (interquartile range 0.64–3.43). The characteristics of the potentially included cohort are shown in Table 1. The number of individuals included for each outcome by treatment group is shown in SI Table.

In unadjusted analysis and after adjustment for PS, age, calendar year, and clustering by practice in which the primary care physician worked, rates of CKD stage 3 or above in individuals prescribed valproate (HR 0.56; 95% CI 0.45–0.69; p < 0.001), olanzapine (HR 0.57; 95% CI 0.43–0.71; p < 0.001), or quetiapine (HR 0.62; 95% CI 0.47–0.80; p < 0.001) were reduced compared to lithium (Table 2, Fig 1).

Compared to lithium, rates of hypothyroidism were reduced in those prescribed valproate (HR 0.60; 95% CI 0.40–0.89; p = 0.012) or olanzapine (HR 0.48; 95% CI 0.29–0.77; p = 0.003), but not quetiapine (HR 0.63; 95% CI 0.38–1.05; p = 0.074) after adjustment. Rates of hypothyroidism were lower in those prescribed valproate (HR 0.24; 95% CI 0.09–0.66; p = 0.003) and

<table>
<thead>
<tr>
<th>Table 1. Patient characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Total, n</td>
</tr>
<tr>
<td>Female, n (%)</td>
</tr>
<tr>
<td>Age, median (IQR), years</td>
</tr>
<tr>
<td>Non-white ethnic background, n (%)</td>
</tr>
<tr>
<td>Duration of drug exposure, median (IQR), years</td>
</tr>
<tr>
<td>Primary care contacts per year, median (IQR)</td>
</tr>
<tr>
<td>CVD history, n (%)</td>
</tr>
<tr>
<td>CKD (or eGFR &lt;60 ml/min/1.73m²)</td>
</tr>
<tr>
<td>Hypothyroidism (or TSH &gt;10 mU/L)</td>
</tr>
<tr>
<td>Hypertension (or TSH &gt;10 mU/L)</td>
</tr>
<tr>
<td>T2DM (or HbA1c &gt;8% mmol/mol)</td>
</tr>
<tr>
<td>Hepatic impairment (ALT &gt;200 U/L, or AST &gt;250 U/L)</td>
</tr>
<tr>
<td>Obesity (BMI ≥30)</td>
</tr>
<tr>
<td>Hyperparathyroidism (adjusted calcium &gt;2.65 mmol/L)</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Epilepsy</td>
</tr>
<tr>
<td>Moderate/heavy alcohol use</td>
</tr>
<tr>
<td>Current smoker</td>
</tr>
<tr>
<td>Bipolar disorder characteristics at baseline, n (%)</td>
</tr>
<tr>
<td>Previous depressive episode</td>
</tr>
<tr>
<td>Previous record of taking study drug</td>
</tr>
</tbody>
</table>

CVD, cardiovascular disease; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; TSH, thyroid stimulating hormone; T2DM, type 2 diabetes mellitus; ALT, alanine transaminase; AST, aspartate aminotransferase; BMI, body mass index.
Table 2. Adverse effects during maintenance treatment.

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>n</td>
<td>Events</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PyAR 100%</td>
<td>51.97</td>
<td>29.85</td>
<td>25.64</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>9.41 (6.20–12.28)</td>
<td>4.32 (3.67–5.17)</td>
<td>4.72 (3.95–5.64)</td>
<td>4.20 (3.30–5.01)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.46 (0.38–0.58)</td>
<td>0.50 (0.41–0.61)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>0.56 (0.45–0.68)</td>
<td>0.67 (0.54–0.79)</td>
<td>0.42 (0.33–0.50)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2CKD stage 4 (n = 4617)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>1.43 (1.17–1.76)</td>
<td>1.04 (0.74–1.45)</td>
<td>0.72 (0.49–1.02)</td>
<td>0.65 (0.47–0.91)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.75 (0.54–0.99)</td>
<td>0.52 (0.33–0.81)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>0.84 (0.67–1.05)</td>
<td>0.84 (0.67–1.05)</td>
<td>0.67 (0.53–0.85)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.006</td>
<td>0.127</td>
<td>0.273</td>
</tr>
<tr>
<td>Hypothyroidism (n = 4,090)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>183</td>
<td>61</td>
<td>41</td>
<td>22</td>
</tr>
<tr>
<td>PyAR 100%</td>
<td>59.43</td>
<td>33.51</td>
<td>29.69</td>
<td>14.91</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>3.69 (2.77–5.75)</td>
<td>2.81 (1.71–2.62)</td>
<td>1.72 (1.27–2.34)</td>
<td>2.12 (1.53–2.88)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.89 (0.72–0.99)</td>
<td>0.80 (0.70–0.89)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>0.70 (0.59–0.85)</td>
<td>0.89 (0.78–1.00)</td>
<td>0.76 (0.67–0.86)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.012</td>
<td>0.003</td>
<td>0.074</td>
</tr>
<tr>
<td>Hyperthyroidism (n = 3,374)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>41</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>PyAR 100%</td>
<td>52.49</td>
<td>25.81</td>
<td>22.42</td>
<td>14.65</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>0.78 (0.59–0.99)</td>
<td>0.80 (0.68–0.92)</td>
<td>0.72 (0.58–0.87)</td>
<td>0.80 (0.67–0.93)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.73 (0.58–0.90)</td>
<td>0.68 (0.53–0.87)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>0.59 (0.49–0.73)</td>
<td>0.78 (0.62–0.95)</td>
<td>0.66 (0.50–0.87)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.003</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Hypercalcaemia (n = 2,094)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>55</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>PyAR 100%</td>
<td>35.10</td>
<td>17.1</td>
<td>13.44</td>
<td>6.59</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>1.52 (1.16–1.97)</td>
<td>0.38 (0.16–0.87)</td>
<td>0.45 (0.20–0.99)</td>
<td>0.33 (0.11–0.99)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.34 (0.19–0.58)</td>
<td>0.31 (0.15–0.68)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>0.28 (0.17–0.48)</td>
<td>0.46 (0.23–0.83)</td>
<td>0.32 (0.12–0.83)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.002</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>T2DM (n = 6,232)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>150</td>
<td>84</td>
<td>88</td>
<td>51</td>
</tr>
<tr>
<td>PyAR 100%</td>
<td>69.51</td>
<td>30.62</td>
<td>34.01</td>
<td>21.02</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>2.17 (1.25–3.65)</td>
<td>2.30 (1.91–2.71)</td>
<td>2.29 (1.80–2.89)</td>
<td>2.10 (1.50–2.91)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.91 (0.74–1.11)</td>
<td>1.07 (0.89–1.28)</td>
</tr>
<tr>
<td></td>
<td>1 (reference)</td>
<td>1.29 (0.74–2.12)</td>
<td>1.20 (0.80–1.49)</td>
<td>0.94 (0.65–1.35)</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.586</td>
<td>0.530</td>
<td>0.752</td>
</tr>
<tr>
<td>CVD (n = 6,365)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events, n</td>
<td>94</td>
<td>32</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>PyAR 100%</td>
<td>69.49</td>
<td>32.56</td>
<td>33.41</td>
<td>21.17</td>
</tr>
<tr>
<td>Rats, per 100 PyAR (95%CI)</td>
<td>1.38 (1.11–1.66)</td>
<td>0.88 (0.63–1.23)</td>
<td>0.78 (0.53–1.14)</td>
<td>0.08 (0.03–0.21)</td>
</tr>
<tr>
<td></td>
<td>Unadjusted HR (95%CI)</td>
<td>1 (reference)</td>
<td>0.67 (0.44–1.04)</td>
<td>0.61 (0.38–0.98)</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>0.91 (0.59–1.41)</td>
<td>0.88 (0.53–1.46)</td>
<td>1.13 (0.63–1.98)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.684</td>
<td>0.501</td>
<td>0.732</td>
<td></td>
</tr>
</tbody>
</table>

>7% weight gain (n = 4,458)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, n</td>
<td>497</td>
<td>410</td>
<td>396</td>
<td>299</td>
</tr>
<tr>
<td>PYAR (1000)</td>
<td>63.29</td>
<td>33.28</td>
<td>30.86</td>
<td>19.95</td>
</tr>
<tr>
<td>Rate, per 100 PYAR (95% CI)</td>
<td>7.39 (4.73–10.9)</td>
<td>12.32 (11.8–13.57)</td>
<td>12.96 (11.7–14.31)</td>
<td>13.75 (14.0–17.58)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.90 (1.64–2.26)</td>
<td>1.99 (1.72–2.30)</td>
<td>2.72 (2.33–3.16)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.27 (1.17–1.41)</td>
<td>1.43 (1.23–1.67)</td>
<td>1.37 (1.16–1.62)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

>15% weight gain (n = 4,458)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, n</td>
<td>179</td>
<td>182</td>
<td>189</td>
<td>130</td>
</tr>
<tr>
<td>PYAR (1000)</td>
<td>63.82</td>
<td>34.27</td>
<td>31.24</td>
<td>19.30</td>
</tr>
<tr>
<td>Rate, per 100 PYAR (95% CI)</td>
<td>2.86 (2.42–3.24)</td>
<td>5.21 (4.59–6.16)</td>
<td>6.06 (5.25–6.98)</td>
<td>6.74 (6.07–7.40)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>2.29 (1.87–2.82)</td>
<td>2.57 (2.06–3.22)</td>
<td>3.41 (2.61–4.44)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.62 (1.31–2.01)</td>
<td>1.84 (1.47–2.30)</td>
<td>1.67 (1.24–2.25)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Hypertension (n = 6,081)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, n</td>
<td>174</td>
<td>85</td>
<td>89</td>
<td>33</td>
</tr>
<tr>
<td>PYAR (1000)</td>
<td>64.25</td>
<td>35.66</td>
<td>32.80</td>
<td>29.20</td>
</tr>
<tr>
<td>Rate, per 100 PYAR (95% CI)</td>
<td>2.71 (2.32–3.14)</td>
<td>2.58 (1.93–3.16)</td>
<td>2.76 (2.24–3.35)</td>
<td>1.69 (1.14–2.42)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>0.98 (0.75–1.36)</td>
<td>1.11 (0.84–1.46)</td>
<td>0.75 (0.50–1.12)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>1.19 (0.90–1.58)</td>
<td>1.41 (1.06–1.87)</td>
<td>0.89 (0.58–1.34)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.274</td>
<td>0.017</td>
<td>0.590</td>
<td></td>
</tr>
</tbody>
</table>

Hepatotoxicity (n = 3,352)

<table>
<thead>
<tr>
<th></th>
<th>Lithium</th>
<th>Valproate</th>
<th>Olanzapine</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events, n</td>
<td>29</td>
<td>10</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>PYAR (1000)</td>
<td>59.13</td>
<td>29.98</td>
<td>20.30</td>
<td>12.63</td>
</tr>
<tr>
<td>Rate, per 100 PYAR (95% CI)</td>
<td>0.40 (0.28–0.62)</td>
<td>0.39 (0.21–0.72)</td>
<td>0.69 (0.41–1.10)</td>
<td>1.04 (0.60–1.79)</td>
</tr>
<tr>
<td>Unadjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>0.95 (0.64–1.40)</td>
<td>1.71 (0.86–3.40)</td>
<td>2.59 (1.25–5.43)</td>
</tr>
<tr>
<td>PS Adjusted HR (95% CI)</td>
<td>1 (reference)</td>
<td>0.65 (0.30–1.20)</td>
<td>1.23 (0.60–2.42)</td>
<td>1.21 (0.52–2.74)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.274</td>
<td>0.058</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CKD, chronic kidney disease; T2DM, type 2 diabetes mellitus; CVD, cardiovascular disease; PYAR, person-years at risk; HR, hazard ratio; PS, propensity score. Unadjusted hazard ratio accounts for clustering by primary care practice, adjusted hazard ratio is adjusted for propensity score age group and calendar period time varying variables and clustering by primary care practice. P-values for PS adjusted HR.

doi:10.1371/journal.pmed.1002058.t002

olanzapine (HR 0.31; 95% CI 0.13–0.73; p = 0.007), but not quetiapine (HR 0.45; 95% CI 0.18–1.18; 0.096), compared to lithium. Hypercalcaemia was less common in those prescribed valproate (HR 0.25; 95% CI 0.10–0.60; p = 0.002), olanzapine (HR 0.32; 95% CI 0.12–0.76; p = 0.008), or quetiapine (HR 0.25; 95% CI 0.07–0.73; p = 0.011) compared to lithium (Table 3, Fig 3).

After adjustment, rates of weight gain were higher with valproate, olanzapine, and quetiapine than lithium (≥15% weight gain: valproate HR 1.62; 95% CI 1.31–2.01; p < 0.001; olanzapine HR 1.84; 95% CI 1.67–2.03; p < 0.001; quetiapine HR 1.67; 95% CI 1.24–2.20; p < 0.001). Rates of hypertension were higher with olanzapine (HR 1.41; 95% CI 1.66–1.87; p = 0.007) than lithium (Table 3, Fig 3). We found no significant difference in rates of CKD stage 4 or above, T2DM, cardiovascular disease, or hepatotoxicity between groups (Table 3). The median number of GFR creatinine and TSH blood tests per year in treatment was higher in those taking lithium than the other drugs (see St Table). Weight measurement and blood tests for adjusted calcium and
ALT/AST were less frequent in patients prescribed lithium (see Table S2). For outcomes in which patients had been excluded because of missing tests (CKD, hyper- and hypothyroidism, hypercalcaemia, weight gain, and hepatotoxicity), sensitivity analyses including all patients resulted in reduced incident rate estimates compared to the primary analyses, but had little effect on HRs (see Table S3). Sensitivity analyses using IPW suggest results from the primary analyses are robust (see Table S4). From Schoenfeld residuals, there was no evidence against the assumption of proportional hazards for any outcome.

Discussion
In a large dataset of nearly 7,000 individuals treated for BPD with lithium, valproate, olanzapine, or quetiapine, with follow-up times of up to 17 y, we found differential rates of a number of adverse events. Those prescribed lithium were more likely to have a decline in renal function and develop hyper- or hypothyroidism and hypercalcaemia. However, they were less likely to gain significant weight. Individuals prescribed olanzapine had the highest rate of weight gain and new onset hypertension. We did not find any statistically significant differences in the rate of new T2DM, cardiovascular disease, or hepatotoxicity across drug treatment groups.

Severe CKD (stage 4 or above) was uncommon in the cohort (approximately 1 in 100 person years at risk), and we did not find differences by drug treatment, but less severe CKD
(stage 3 or above) occurred most frequently in patients prescribed lithium. Whilst many of these patients (i.e., those with CKD stage 3) would not progress to a clinically relevant decline in renal function, a number of them would be at increased risk of doing so. It remains unclear if this result is due to (1) lack of power to determine a true difference in rates of severe CKD, (2) surveillance bias due to increased monitoring of renal function in those taking lithium, which would lead to apparent increased rates of asymptomatic CKD stage 3, or (3) lithium treatment truly increasing the risk of reduced renal function without increasing severe CKD risk. Previous studies have found similar results and have not been able to account for this potential bias [12,14,56]. Clo et al. found no decline in eGFR in individuals taking lithium, using a similar active comparator design, but were also limited by potential ascertainment bias [12].

Rates of both hypothyroidism and hyperthyroidism were increased in individuals prescribed lithium compared to valproate and olanzapine (but not quetiapine). Increased hypothyroidism has been shown previously [11,55], but literature on the association between lithium and hyperthyroidism is inconsistent [13], and lithium-induced hyperthyroidism is considered rare [55]. Monitoring thyroid dysfunction in BPD is vital because of evidence that abnormalities are associated with longer time to remission and more symptoms during the maintenance period [60]. It is possible that thyroid function normalises on cessation of lithium, but only one study has investigated this [61]. Hypercalcaemia is also recognised to be associated with lithium prescribing [11,13,56,62]. Calcium monitoring in patients prescribed lithium was rare in one representative sample of primary care (37% had one or more calcium blood test result), despite it being recommended in the 2006 NICE guidance [3].

The rate of individuals gaining more than 7%, and more than 15% of their baseline weight, was greater in those prescribed olanzapine, quetiapine, or valproate than those prescribed
Lithium. This degree of weight gain represents a significant risk factor for a number of adverse physical health outcomes, including CVD and T2DM [38]. We may not have captured increased rates of CVD or T2DM because of the relatively brief median follow-up time, in relation to the time taken to develop these diseases. Olanzapine had the highest adjusted rate of greater than 15% weight gain compared to lithium, and the highest rate of new onset hypertension. This has been shown previously in comparisons of antipsychotic drugs [64] and in trials of olanzapine versus lithium or valproate [63].

Hepatotoxicity was rare in the cohort and, before PS adjustment rates, appeared to be elevated in the quetiapine group, compared to lithium. This association has been identified previously [18]. After adjustment, there was no evidence of between-group differences.

Strengths and Limitations

The major strength of this study, beyond size and length of follow-up, is the direct comparison between BP/D maintenance mood stabilizer treatment options for a number of adverse effects. The use of electronic health records also means it is possible to adjust for a number of demographic and physical health characteristics that may have influenced the clinician’s decision to treat with a particular medication or potentially confound the relationship between treatment and adverse outcome. Despite including numerous variables in the PS, it is possible that residual confounding remained, especially as those prescribed lithium were older and were more likely to have taken the drug previously, perhaps reflecting a more chronic illness course. It may be that important patient or clinician features were not captured by the score, and despite the balance of observed covariates, we cannot confirm balance of unobserved covariates [46-47]. We were also unable to consider dosage differences across the different treatment groups in this analysis. Periods of lithium toxicity may be particularly important with regards to developing renal failure, and we were unable to capture this information from the available data. Missing data can be a problem in studies utilising electronic patient records, especially as there may be a clinical reason why information is missing. Because of the way outcomes were defined, T2DM, cardiovascular disease, and diagnoses of hypertension had no missing data, and no covariates in the PS had missing values.

Patients prescribed lithium had no more physician contacts than those taking other mood stabilizer medication. In individuals that ever received tests during treatment exposure, testing frequency was similar in all study drugs for adjusted calcium, liver function, and weight (see S2 Table). Frequency of testing renal and thyroid function was higher in those taking lithium, which reflects the guidance for monitoring [51]. Patients prescribed lithium were also more likely to have at least one renal function, thyroid function, calcium, or liver function test compared to patients taking other drugs. This is likely to be due to both drug-related indications for monitoring and the longer drug exposure seen in those taking lithium. IPW sensitivity analysis to account for this difference did not alter our conclusions (see S3 Table). In the primary analysis, the likely effect of this differential missingness would be to reduce the hazard ratios for lithium compared to the other drugs, relative to their true values, as blood tests in the non-lithium group are more likely to be related to clinical symptoms than monitoring guidance (for instance, this is likely to represent an underestimation of the true hypercalcaemia hazard ratio for lithium versus other drugs). The median number of weight measurements was similar in each group, suggesting detection of weight gain was not related to differential monitoring. The sensitivity analyses including individuals irrespective of blood tests produced similar adjusted hazard ratios as the primary analyses for each outcome, but often with reduced incidence of the outcome in each treatment group (see S3 Table). These analyses may more accurately reflect testing occurring because of clinical indication.
Conclusions

Lithium remains an important treatment option for individuals with BPD. However, there is clear evidence that its use is associated with a number of adverse events. These risks need to be offset with the potentially superior effectiveness and anti-suicidal benefits of the drug compared to other treatment options [5,6,8]. It is also true that other recommended maintenance treatments can have serious side effects, often related to weight gain, and are not suitable for use in certain patient groups (such as the contraindication of valproate in women of childbearing potential [2]).

Assiduous monitoring of patients prescribed lithium should ameliorate some risk associated with effects on renal physiology and endocrine systems. Given the need to balance an array of risks and benefits, an individualised and collaborative approach to treatment choice is likely to be most appropriate. To achieve this, further research identifying patient characteristics that are risk factors for specific side effects and an understanding of the risks and benefits of stopping treatment in those who experience adverse effects is necessary.

Supporting Information

S1 Table. Patients included for each outcome, n (%).
(DOCX)

S2 Table. Median number (and interquartile range) of tests per year of drug exposure in patients included in analyses.
(DOCX)

S3 Table. Sensitivity analyses to account for missing blood tests by (1) including all individuals and (2) performing inverse probability weighting.
(DOCX)

S1 Text. Patient selection.
(DOCX)

S2 Text. STROBE statement.
(DOCX)

S3 Text. Prospective analysis plan.
(DOCX)

Author Contributions

Conceived and designed the experiments: JFH LM KW JRG MK DPJ0. Analyzed the data: JFH LM. Wrote the first draft of the manuscript: JFH. Contributed to the writing of the manuscript: JFH LM KW JRG MK DPJ0. Agree with the manuscript’s results and conclusions: JFH LM KW JRG MK DPJ0. All authors have read, and confirm that they meet, ICMJE criteria for authorship.

References


46. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46(3):399-432. DOI: 21819120

47. Senn JI. Kueh T, Walter A. Use of propensity score techniques to account for exposure-related covariates in an example of and lesson. Med Care. 2007;45(10):973-958. PMID: 17409274


5. Media

5.1 The Guardian (online 14.8.2016, print 15.8.2016) – Lithium should be more widely used for bipolar disorder, researchers say

...The lead author behind the research told the Guardian that widespread “lithium stigma” among patients is leading to them receiving the wrong treatment and ending up admitted to hospital unnecessarily because their condition is not as well controlled as it could be.

“Lithium is a drug with a bad reputation. It is seen by patients, and some psychiatrists, as a dangerous drug. People rightly have suspicions about it. Patients say that the downsides
include emotional numbing – feeling that you aren’t connected with your feelings – as well as tremors,” said Dr Joseph Hayes, a psychiatrist at University College London.

But lithium’s reputation is largely misplaced and based on the experiences of patients from the 1960s to the 1980s who were given too large a dose of the drug, he added. The new research, published in the medical journal PLOS Medicine, found that the side-effects of the mood-stabilising alternatives used by most patients are either the same as or worse than lithium, Hayes said.

The paper, co-written by Hayes and four colleagues from UCL and Oxford University, concluded that: “Lithium remains an important treatment for individuals with bipolar disorder.” It accepts that “there is clear evidence that its use is associated with a number of a owever, it adds: “These risks need to be offset with the potentially superior effectiveness and anti-suicidal benefits of the drug compared to other treatment options.”

The researchers studied a nationally representative sample of 6,671 patients across the UK who were treated for bipolar disorder between 1995 and 2013. Of those, 2,148 had taken lithium, 1,670 had used valproate, 1,477 had been on olanzapine and 1,376 had taken quetiapine. They experienced side-effects including chronic kidney disease, thyroid disease, weight gain and high blood pressure.

The researchers’ analysis bore out one of the two main criticisms of lithium, but found the other to be baseless. They found that patients on lithium had a higher risk of suffering kidney function problems and developing hypothyroidism or hyperthyroidism and also hypercalcemia. However, none of the drugs caused more severe kidney problems.

Meanwhile, lithium patients were less likely to put on weight than patients on the other drugs. While 15%-20% of those on the three other drugs were more likely to gain more than 15% of their body weight, just 10% of those on lithium put on the same amount of extra pounds. Those on olanzapine added the most weight and experienced high blood pressure as a result.

Separate research has shown that patients on the other three medications are 40% more likely to harm themselves than those on lithium. Bipolar disorder carries one of the highest
rates of suicide of any mental illness, alongside schizophrenia and alcohol and drug addiction.

The very limited use of lithium is despite the National Institute for Health and Care Excellence (Nice) advising in 2014 that it should be the standard treatment for bipolar disorder, which is also known as manic depression and is characterised by manic highs and bouts of depression. That superseded its previous view, outlined in 2006, that any of the four drugs were useful first lines of treatment for the condition, which affects about one in 100 people.

“Lithium stigma, which includes some people in the psychiatric community, leads to people using drugs that are less effective [than lithium]. To me as a doctor that’s a big worry because my main aim is to help people to be well and if you aren’t doing that with the best "I think that many patients are missing out quite commonly on the best available treatment. That means that people end up in hospital more often than they need to and end up achieving less in their lives than they could do if they were on lithium. The high suicide rate with bipolar disorder should encourage greater use of lithium. There should be more sensible use of it.”

Stephen Buckley, head of information at the charity Mind, said: “We welcome research which adds to our understanding of treatments and medications for people experiencing mental health problems, including bipolar disorder. But as with all areas of mental health there is still more research to be done.

“Different people will find that different treatments help with managing their mental health problems. This may be medication, talking therapies, or a mixture of both.”

5.2 The Mental Elf (online 2.6.15) – Premature mortality in bipolar disorder

People with bipolar disorder are more likely to die at a younger age compared with the general population (Crump et al 2013; Hoang et al 2013). This is often due to high rates of
suicide and violent crime, which I blogged about last year, but may also be attributed to a heightened risk for physical health problems.

In 2013, Rethink published a report (PDF) outlining a number of common factors which can lead to medical illnesses in this population such as smoking, side effects from antipsychotic medication, as well as poor physical health monitoring and stigma.

Hayes and colleagues (2015) conducted a systematic review and meta-analysis of large observational studies to estimate the mortality rate of people with bipolar disorder compared with the general population. They looked at a range of different reasons for mortality which included unnatural causes, such as suicide and violent crime, as well as natural causes, such as circulatory and respiratory problems.

Methods
The authors looked for eligible articles by searching three electronic databases, scanning reference lists of included studies and tracking citations using the Cochrane Database of Systematic Reviews and Google Scholar.

Studies were included if they reported data on deaths of people with a diagnosis of bipolar disorder (any criteria were accepted), due to any reason (all-cause mortality) or deaths due to specific reasons (cause-specific mortality) including:

- Natural deaths
- Unnatural deaths
- Suicide
- Other violent deaths
- Infection
- Neoplasm
- Respiratory
- Circulatory system disease

Studies were excluded if they had fewer than 50 participants or were not standardised by age.
Standardised mortality ratios (SMRs) with their 95% confidence intervals were used to calculate the ratio of participants with bipolar disorder who died for each reason compared with the general population. SMRs greater than 1 indicate increased mortality in people with bipolar disorder compared with the general population, whereas those smaller than 1 indicate decreased mortality in this population.

Heterogeneity within each meta-analysis was assessed using the I² index and chi-square test. Additionally a number of meta-regressions were carried out for all-cause mortality controlling for: decade of publication, cohort size, geographical region, mid-decade of cohort data collection and population type (inpatient or community based). Subgroup analyses were also performed for geographical region of study, patient population type and decade of the middle year of patient observation.

Results
In total, 31 studies of 305,859 people with bipolar disorder met the inclusion criteria. Studies were mainly inpatient cohorts (64%) and a large proportion were conducted in Scandinavian countries (45%).

All-cause mortality
The SMR for all-cause mortality was 2.05 (95% CI 1.98 to 2.23), ranging from 1.24 (95% CI 0.83 to 1.17) to 4.65 (95% CI 1.27 to 11.91). Heterogeneity between studies was significant and high (p < 0.001, I² =96.2%).

Cause-specific mortality
SMRs indicated increased rates of death for people with bipolar disorder for all cause-specific mortality categories. Estimates were highest for death due to suicide (SMR=14.44) and unnatural causes (SMR=7.42). As for all-cause mortality, heterogeneity was significant and high for most categories except for infection and neoplasm which were relatively homogenous.

Meta-regression and sub-group analyses
Meta-regression analyses revealed that none of the following variables could account for the heterogeneity in findings: decade of publication, cohort size, geographical region, mid-decade of cohort data collection or population type. Subgroup analyses also did not find
any effect on heterogeneity when stratifying for geographical region, population type and mid-decade of study.

**Strengths and limitations**

The authors carried out a thorough search strategy using multiple databases and tracking citations from reference lists of included studies. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) proposal for reporting were also followed. However, despite being an item in both reporting checklists, the authors did not carry out an assessment of study quality, which may have played a role in the high levels of heterogeneity found.

Another limitation in this review is the fact that included studies only adjusted SMRs for age and gender. It therefore was not possible to assess whether other factors may have accounted for the increased rate of deaths or heterogeneity in findings. For example, people with a serious mental illness are more likely to smoke which may have had a significant impact on deaths due to respiratory problems.

**Summary**

Overall this review provides evidence that people with bipolar disorder have increased mortality rates compared with the general population.

For all-cause mortality, there was a two-fold increase, whereas for suicide the SMR was as high as 14, which reflects previous results from large cohort studies. Although there was considerable heterogeneity in the summary estimate, all studies showed an increased risk for all-cause mortality of which only 4 out of 31 showed confidence intervals which crossed the line of no effect. These studies were all relatively small (74-440 participants) which explains the large confidence intervals. This finding indicates that we can be fairly certain that a true increased risk exists, although the precise estimate is still uncertain.

An interesting finding was that when controlling for year of publication, there was no effect on estimates indicating no change in mortality ratios from 60 years ago with those completed recently. This highlights the need for improvements in physical and mental
health monitoring, smoking cessation programmes and the provision of clear information about side effects of medication.

5.3 The Mental Elf (online 18.8.16) – Lithium for bipolar disorder: the best maintenance mood stabiliser protection against self-harm and suicide?

The prevention of self-harm and suicide is one of the primary goals of treatment across all psychiatric disorders. Bipolar disorder has particularly high rates of completed suicides, so prevention of suicide is especially important for this disorder. The question of whether lithium (or another medical treatment) best prevents suicide in bipolar disorder has long been asked, and the preponderance of indirect evidence suggests that it does, although there is great uncertainty about the relative benefit of lithium compared to other drugs when the goal is to prevent self-harm, injury and suicide.

The study under consideration here uses a different approach to answering the question about the relative benefit of different common treatments for bipolar disorder. It examines a large dataset derived from electronic health records (EHR) to determine whether exposure to lithium results in improved outcomes compared to sodium valproate, olanzapine or quetiapine, three commonly used acute and maintenance treatments for bipolar disorder. The authors used a propensity score to adjust for baseline clinical characteristics to try to make the study groups comparable.

The great strength to this approach is that it examines real patients, not those selected for participation in a randomised trial. There is a possibility of bias, however, that may make results from EHR studies difficult to interpret. Confounding by indication (the fact that certain treatments tend to look harmful because they are given to sicker patients, or visa versa) is the biggest barrier to knowing whether the results of this study are biased. Perhaps the patients in one group were inherently different from those in the other groups and therefore a medication was chosen for that reason (rather than that the medication itself caused the difference). Let’s have a look at the study and its results.

Methods
Cohort study using primary care EHRs (electronic health records) data collected between January 1, 1995, and December 31, 2013, by The Health Improvement Network (THIN) system.

Individuals (aged 16 and older) with diagnoses of bipolar disorder were included in the study if they received 2 or more consecutive prescriptions for treatment lasting 28 days or longer of lithium, valproate, olanzapine or quetiapine.

Patients were followed from the time of first prescription to 3 months after medication discontinuation (if that occurred). Patients prescribed any of the medications concurrently were excluded from the analyses.

**Outcomes**
The primary outcome of interest was emergency department or primary care attendance for self-harm during the period of drug exposure and the 3 months afterward (including intentional poisoning, intentional self-injurious behaviour, and self-harm acts of uncertain intent). Secondary outcomes were unintentional injury (e.g., falls or motor vehicle crashes) seen in primary or secondary care or a record of the patient’s suicide during this period.

**Propensity score**
Propensity score (PS) adjustment for sex, age at the start of treatment with the study drug, year of entry to the cohort, race/ethnicity, cardiovascular disease diagnosis before baseline, hypertension, chronic kidney disease at baseline, history of hypothyroidism or hyperthyroidism, history of liver disease, type 2 diabetes mellitus, epilepsy, alcohol use (grouped as none or low, moderate or heavy, or dependence), history of illicit drug use, smoking status, body mass index, anxiety symptoms or diagnosis before baseline, depressive symptoms or diagnosis before baseline, sleep disturbance before baseline, treatment with the study drug at or before baseline, and history of previous self-harm.

**Results**
The authors found a strong association between lithium prescribing and lower risk of self-harm:
Of 14,396 individuals with a diagnosis of bipolar disorder, 6,671 were included in the cohort, with:

- 2,148 prescribed lithium
- 1,670 prescribed valproate
- 1,477 prescribed olanzapine
- 1,376 prescribed quetiapine

Self-harm rates were lower in patients prescribed lithium compared with the other drugs:

- Lithium (205; 95% CI, 175 to 241 per 10,000 person-years at risk [PYAR])
- Valproate (392; 95% CI, 334 to 460 per 10,000 PYAR)
- Olanzapine (409; 95% CI, 345 to 483 per 10,000 PYAR)
- Quetiapine (582; 95% CI, 489-692 per 10,000 PYAR).

The authors also report:

People prescribed lithium tended to be older than those taking other study drugs, with more years of follow-up data. These individuals were less likely to have records of depression, anxiety, or self-harm before entry into the cohort. Individuals prescribed lithium had no more contacts with primary care services during follow-up than individuals prescribed other drugs.

This association [between lithium and self-harm] was maintained after PS adjustment (hazard ratio [HR], 1.40; 95% CI, 1.12 to 1.74 for valproate, olanzapine, or quetiapine vs lithium) and PS matching (HR, 1.51; 95% CI, 1.21 to 1.88). After PS adjustment, unintentional injury rates were lower for lithium compared with valproate (HR, 1.32; 95% CI, 1.10 to 1.58) and quetiapine (HR, 1.34; 95% CI, 1.07 to 1.69) but not olanzapine. The suicide rate in the cohort was 14 (95% CI, 9-21) per 10,000 PYAR. Although this rate was lower in the lithium group than for other treatments, there were too few events to allow accurate estimates.

**Is lithium viewed as a risky drug?**

Consistent with what has been reported in the literature, lithium is associated with lower rates of self-harm in patients with bipolar disorder compared to the rates found in patients treated with valproate, olanzapine, or quetiapine. There remains a great deal of
uncertainty, though, about all the differences as to why drugs were differentially prescribed and whether adjusting for them using propensity scores is adequate to statistically create groups that are as unbiased as those that would be arrived at through randomisation.

Clinicians likely make very complex determinations of risk and benefit when making prescribing decisions. It is quite possible, though, that prescribers view lithium as a risky drug (because of its risk of both accidental and intentional overdose) compared to the other drugs examined in this study (whose risks in overdose are considerably lower), and therefore are prescribing lithium to lower risk patients to begin with.

There certainly may be reasons for this. Lithium has a very narrow therapeutic window and must be monitored closely in the context of other medications (NSAIDs and diuretics, for example) and medical problems. Even if lithium is a drug that (as the authors suggest) reduces impulsive aggression, it is not likely being prescribed as readily to patients who are perceived by clinicians to have that problem in the first place. The baseline characteristics of the sample suggest as much: the sample, among other things, is older (younger people tend to be more impulsive), less likely to have had prior self-harm (in spite of being older and therefore having extra years before treatment to have harmed themselves), less likely to be cigarette smokers (itself associated with impulsiveness), and less likely to be anxious or depressed.

All of this suggests that lithium is being prescribed to a lower risk group to begin, and makes it very difficult (even with propensity score matching) to conclude that it is the lithium itself, rather than baseline differences in groups, that is reducing the risk. The propensity score matching markedly reduces the differences in risk between the lithium group and the others, and it is difficult to know whether the differences would be even further diminished (towards the null) if all factors actually involved in prescribing differences (such as actual measures of impulsiveness) were included in the propensity score.

Conclusion
The authors conclude that: Lower rates of self-harm in those prescribed lithium may be due either to improved mood stabilization compared with other treatments or specific effects on impulsive aggression and risk taking. An alternative conclusion, not addressed by the
authors, is that even at baseline, lithium is being preferentially prescribed to a lower risk group.

**Implications**

How this study might be most appropriately used is to increase the dissemination of evidence about the importance of more widespread lithium use. I, for one, am for more lithium prescribing. There is much more randomised data supporting its use in general in bipolar disorder than for many other drugs, and its prescribing is, contrary to what the evidence might necessitate, declining. People are afraid of it. There are no data, in fact, to support its being associated with more death from all causes; to the contrary, the opposite is true.

Lithium, in this cohort, is clearly being preferentially kept from more severely ill patients with higher risk of self-harm (as the baseline characteristics confirm), but much indirect evidence suggest that the converse should be true. Even if this study is not definitive, the preponderance of the evidence suggests that lithium mitigates risk of self-harm and suicide, and none, most importantly, suggests that it increases such risk. Continued education about lithium prescribing, the absolute and relative risks to patients of its use, and support to practitioners and patients alike must be stressed.

There is also a call to actually do comparative trials. The Veterans Administration, in the United States, is currently undertaking the largest randomised trial of lithium for suicide prevention ever undertaken, expecting to randomise 1,862 Veterans with depression (from both bipolar disorder and major depressive disorder) to either adjunctive lithium or placebo, and to follow them for up to a year. When completed, this study has the potential to add considerable information regarding just the question we are asking today.

**Primary paper**

