Cognitive function in people with psychiatric and neurological disorders in UK Biobank

Breda Cullen

BA (Hons), MSc, DClinPsy, PgDip

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

March 2018

Abstract

Cognitive impairment is a major cause of disability for a large number of working-age adults living with chronic psychiatric and neurological conditions. Although well recognised in schizophrenia spectrum disorders and in neurological diseases such as multiple sclerosis (MS), cognitive impairment has historically received less attention in mood disorders. The relative prevalence of cognitive impairment in bipolar disorder (BD) and major depression compared with other conditions has not been clearly established, and the risk factors that drive cognitive variation within and across conditions are not well understood.

The primary focus of this thesis was on BD, and the objectives were: (1) to investigate the prevalence of cognitive impairment in BD, compared with major depression, schizophrenia, MS and Parkinson's disease (PD); and (2) to develop causal models to quantify and explain variation in cognitive function in BD and in other conditions. The methods encompassed a systematic literature review, a prevalence study using cross-sectional data from the UK Biobank cohort, and a series of multivariable analyses of UK Biobank data using graphical methods, regression- and matching-based estimation, and mediation models.

The systematic review indicated that between 5% and 58% of adults with euthymic BD showed cognitive impairment. Prevalence was lower in the mania/BD group identified within the UK Biobank cohort, at around 7-10%, which was similar to rates seen in the MS and PD groups within the cohort. When causal models of cognitive performance in the mania/BD group took account of multiple potential confounders, performance on a short-term visuospatial memory test showed a small but reliable decrement. Mediation models provided evidence of indirect negative effects on cognitive performance via psychotropic medication, but not via cardiometabolic disease. A similar pattern of results was seen in the major depression group, though with smaller effect sizes.

This thesis emphasises the importance of cognitive function as a fundamental phenotype in psychiatric and epidemiological research. There is scope to build on this work in future follow-up waves in UK Biobank, as well as in other UK and international cohort studies and through linkage with routine healthcare data.

Table of Contents

Abstract 2
List of Tables 6
List of Figures 7
Publications and Conference Presentations 9
Acknowledgements 10
Author's Declaration 11
Chapter 1 Introduction 12
1.1 Cognitive function and its assessment 12
1.2 Cognitive impairment in psychiatric and neurological conditions 15
1.3 Population cohorts in cognitive and mental health research 20
1.4 Thesis aims 21
1.5 Overview of thesis structure 22
Chapter 2 Systematic review of the prevalence of cognitive impairment in bipolar disorder 23
2.1 Background 23
2.2 Methods 24
2.3 Results 28
2.4 Discussion 49
2.5 Next steps 52
Chapter 3 Description of the UK Biobank resource 53
3.1 Cohort composition 53
3.2 Assessment procedures 55
3.3 Sociodemographic data 57
3.4 Lifestyle and physical measures 57
3.5 Medical history and mental wellbeing 57
3.6 Cognitive assessment 60
3.7 Genetic data 66
3.8 Summary 68
Chapter 4 Identifying psychiatric and neurological groups in UK Biobank 70
4.1 Information sources 70
4.2 Ascertaining exposures of interest 71
4.3 Comparison of group characteristics by each ascertainment method 73
4.4 Overlap between ascertainment methods 82
4.5 Discussion 84
4.6 Next steps 85
Chapter 5 Prevalence of cognitive impairment in psychiatric and neurological groups in UK Biobank 86

5.1 Background 86
5.2 Methods 87
5.3 Results 91
5.4 Discussion 103
5.5 Conclusion 106
5.6 Next steps 106
Chapter 6 Causal inference with observational data 108
6.1 Conceptualisations of causality 108
6.2 Sources of statistical association in data 110
6.3 Assumptions necessary for valid causal inference 115
6.4 Total causal effects 118
6.5 Decomposition into direct and indirect effects 122
6.6 Validity and generalisability 128
6.7 Criticisms of the causal inference approach 130
6.8 Overview of analyses in UK Biobank 131
Chapter 7 Explaining variation in cognitive function in bipolar disorder 133
7.1 Methods 133
7.2 Results 143
7.3 Discussion 166
Chapter 8 Explaining variation in cognitive function in other psychiatric conditions 170
8.1 Major depression 170
8.2 Schizophrenia 190
8.3 Discussion 194
Chapter 9 Discussion 197
9.1 Summary of findings 197
9.2 Contribution of this work to the literature 199
9.3 Critical appraisal 201
9.4 Clinical and public health implications 204
9.5 Future directions 205
Appendices 208
A – PROSPERO registration 209
B – Search strategy for systematic review 214
C – Eligibility checklist for systematic review 215
D – Data extraction form for systematic review 217
E – Eligible articles excluded from systematic review synthesis 220
F – Critical appraisal ratings for systematic review 222
G – STROBE ratings for systematic review 224
H – Characteristics of the study samples in the systematic review 226
I – Cognitive impairment in BD-I vs BD-II 228

J – Factors associated with cognitive impairment in the systematic review 229
K – UK Biobank project approval 232
L – NHS Research Ethics Committee approval 234
M – University Research Ethics Committee approval 242
N – NHS R&D acknowledgement 244
O – Touchscreen mood questionnaire classifications in UK Biobank 245
P – Psychiatric and neurological diagnoses in linked hospital records 246
Q – Self-reported psychiatric and neurological diagnoses in UK Biobank 248
R – Self-reported medications in UK Biobank 249
S – Characteristics of the depression sub-groups in UK Biobank 253
T – Additional characteristics of the exposed and unexposed groups in UK Biobank 256
U – Population attributable prevalence results in UK Biobank 258
V – Sensitivity analysis results for prevalence study in UK Biobank 260
W – Genome-wide polygenic score results 267
X – Testable independency results for mania/bipolar DAG 272
Y – Explanation of DAG used in causal analyses 273
Z – Supplementary results for mania/bipolar mediation analyses 280
AA – Equivalence class for mania/bipolar DAG 283
BB – Testable independency results for major depression DAG 284
CC – Supplementary results for major depression mediation analyses 285
DD – Equivalence class for major depression DAG 289
EE – Testable independency results for schizophrenia DAG 290
References 293

List of Tables

Table 2.1 - Characteristics of the included articles 31
Table 2.2 - Prevalence of cognitive impairment in BD-I samples 36
Table 2.3 - Prevalence of cognitive impairment in mixed BD samples 38
Table 4.1 - Information sources for each exposure 71
Table 4.2 - Mania/bipolar disorder group characteristics 75
Table 4.3 - 'Any depression' group characteristics 76
Table 4.4 - Schizophrenia group characteristics 78
Table 4.5 - Multiple sclerosis group characteristics 79
Table 4.6 - Parkinson's disease group characteristics 80
Table 4.7 - Unexposed comparison group characteristics 81
Table 5.1 - Characteristics of the exposed and unexposed groups 92
Table 5.2 - Prevalence of cognitive impairment across groups 95
Table 7.1 - Summary of cognitive outcome measures in the mania/bipolar and comparison groups 145
Table 7.2 - Summary of covariates in the mania/bipolar and comparison groups 147
Table 7.3 - Summary of covariates in matched mania/bipolar and comparison groups 155
Table 7.4 - Mediation of the effect of mania/bipolar disorder on cognitive outcome via cardiometabolic disease 162
Table 7.5 - Mediation of the effect of mania/bipolar disorder on cognitive outcome via psychotropic medication 163
Table 8.1 - Summary of cognitive outcome measures in the major depression and comparison groups 174
Table 8.2 - Summary of covariates in the major depression and comparison groups 175
Table 8.3 - Summary of covariates in matched major depression and comparison groups 180
Table 8.4 - Mediation of the effect of major depression on cognitive outcome via cardiometabolic disease 187
Table 8.5 - Mediation of the effect of major depression on cognitive outcome via psychotropic medication 188

List of Figures

Figure 2.1 - PRISMA flow diagram showing results of literature search and screening.....29
Figure 2.2 - Executive function impairment prevalence across different thresholds.....41
Figure 2.3 - Attention/working memory impairment prevalence across different thresholds 42
Figure 2.4 - Speed/reaction time impairment prevalence across different thresholds 43
Figure 2.5 - Verbal and visual memory impairment prevalence across different thresholds 45
Figure 2.6 - Funnel plots with pseudo 95% confidence limits.....48
Figure 3.1 - UK Biobank invitation and appointment process.....54
Figure 3.2 - Layout of UK Biobank assessment centre 56
Figure 3.3 - Reaction time test in UK Biobank.....61
Figure 3.4 - Numeric memory test in UK Biobank.....62
Figure 3.5 - Visuospatial memory (‘pairs matching’) test in UK Biobank.....62
Figure 3.6 - Prospective memory test in UK Biobank.....63
Figure 4.1 - Overlap between mania/bipolar disorder ascertainment sources 82
Figure 4.2 - Overlap between 'any depression' ascertainment sources 82
Figure 4.3 - Overlap between schizophrenia ascertainment sources 83
Figure 4.4 - Overlap between multiple sclerosis ascertainment sources 83
Figure 4.5 - Overlap between Parkinson's disease ascertainment sources.....83
Figure 5.1 - Standardised prevalence estimates for cognitive impairment 99
Figure 5.2 - Standardised prevalence ratios for cognitive impairment 101
Figure 6.1 - Directed acyclic graph showing a causal effect of X on Y 110
Figure 6.2 - Causal chains.....112
Figure 6.3 - Confounding.....112
Figure 6.4 - Collider conditioning.....114
Figure 6.5 - Simple mediation model.....123
Figure 6.6 - Intermediate confounding 125
Figure 6.7 - Linear mediation model.....127
Figure 7.1 - Mania/bipolar disorder analysis sample flowchart.....144
Figure 7.2 - Original directed acyclic graph for the effect of mania/bipolar disorder on cognitive outcome 151
Figure 7.3 - Modified directed acyclic graph for the effect of mania/bipolar disorder on cognitive outcome 153
Figure 7.4 - Total causal effect of mania/bipolar disorder on reasoning 156
Figure 7.5 - Total causal effect of mania/bipolar disorder on reaction time.....157
Figure 7.6 - Total causal effect of mania/bipolar disorder on numeric memory 158
Figure 7.7 - Total causal effect of mania/bipolar disorder on visuospatial memory 159

Figure 7.8 - Total causal effect of mania/bipolar disorder on prospective memory 160
Figure 7.9 - Comparison of missing data approaches in mania/bipolar disorder total effects analyses 165
Figure 8.1 - Major depression analysis sample flowchart.....173
Figure 8.2 - Directed acyclic graph for the effect of major depression on cognitive outcome 178
Figure 8.3 - Total causal effect of major depression on reasoning 181
Figure 8.4 - Total causal effect of major depression on reaction time.....182
Figure 8.5 - Total causal effect of major depression on numeric memory 183
Figure 8.6 - Total causal effect of major depression on visuospatial memory 184
Figure 8.7 - Total causal effect of major depression on prospective memory.....185
Figure 8.8 - Comparison of missing data approaches in major depression total effects analyses 189
Figure 8.9 - Schizophrenia analysis sample flowchart.....193

Publications and Conference Presentations

The following publications and presentations have resulted from the research described in this thesis:

Publications

Cullen, B., Smith, D. J., Deary, I. J., Evans, J. J., & Pell, J. P. (2017). The 'cognitive footprint' of psychiatric and neurological conditions: Cross-sectional study in the UK Biobank cohort. *Acta Psychiatrica Scandinavica, 135,* 593-605. doi:10.1111/acps.12733

Cullen, B., Ward, J., Graham, N. A., Deary, I. J., Pell, J. P., Smith, D. J., & Evans, J. J. (2016). Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder: A systematic review. *Journal of Affective Disorders, 205,* 165-181. doi:10.1016/j.jad.2016.06.063

Conference presentations

Cullen, B., Smith, D. J., Deary, I. J., Evans, J. J., & Pell, J. P. (2017). The 'cognitive footprint' of psychiatric and neurological conditions: Cross-sectional study in the UK Biobank cohort. Oral presentation at the International Federation of Psychiatric Epidemiology congress, Melbourne, 17-20 October

Cullen, B., Ward, J., Graham, N. A., Deary, I. J., Pell, J. P., Smith, D. J., & Evans, J. J. (2016). Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder: A systematic review. Poster presentation at the International Society for Bipolar Disorders & International Society for Affective Disorders joint conference, Amsterdam, 13-16 July

Acknowledgements

The research described in this thesis was conducted using the UK Biobank resource, under approved application 11332. UK Biobank was established by the Wellcome Trust, Medical Research Council, UK Government Department of Health, Scottish Government and Northwest Regional Development Agency. UK Biobank has also had funding from the Welsh Government, British Heart Foundation and Diabetes UK. I am very grateful to the UK Biobank participants for their time and effort.

I would like to express my gratitude to the many people who made this thesis possible:

- My supervisors, for their wisdom and guidance; particular thanks to Daniel Smith for getting me involved in psychiatric epidemiology research in UK Biobank from an early stage, to Ian Deary for lending his expertise in cognitive epidemiology, to Jill Pell for supporting my methodological training, and especially to Jonathan Evans for supervising me through not one but two doctorates and for mentoring me in my clinical and academic career for the past fifteen years
- My assessors, Tom McMillan and Jim Lewsey, and postgraduate convenor Craig Melville, for advice and support throughout this process
- My colleagues Joey Ward, Nicholas Graham, Daniel Mackay and Carlos Celis Morales, for their invaluable contributions to the data analyses
- The UK Biobank research group in the Institute of Health and Wellbeing, for their advice, Stata coding tips and good company - Jana Anderson, Mark Bailey, Amy Ferguson, Keira Johnston, Donald Lyall, Laura Lyall, Daniel Martin, Barbara Nicholl, Claire Niedzwiedz, Robert Pearsall, Richard Shaw, Rona Strawbridge and Zia Ul-Haq
- Karl Ferguson, for sharing my enthusiasm for all things causal inference-related
- Katherine Keyes at Columbia University, for expert advice and encouragement.

I gratefully acknowledge the financial support of the Scottish Executive Chief Scientist Office and The Dr Mortimer and Theresa Sackler Foundation. Thanks are due to Jonathan Cavanagh and Andrew McIntosh for enabling me to share work in progress at the annual Sackler PhD conference.

I am especially thankful to Harry and all my family and friends, for their constant support.

Author's Declaration

I declare that I am the sole author of this thesis, except where the assistance of others has been acknowledged. The work in this thesis has not been submitted in any form for another degree or professional qualification.

Breda Cullen
March 2018

Chapter 1 Introduction

Cognitive impairment is a core feature of many different behavioural and brain disorders, contributing to the significant burden attributed to these conditions. In addition to rising dementia prevalence in the ageing population, cognitive dysfunction is a major cause of disability for a large number of working age adults living with chronic psychiatric and neurological conditions. Although well recognised in schizophrenia spectrum disorders and in neurological diseases such as multiple sclerosis (MS), cognitive impairment has historically received less attention in mood disorders. The relative prevalence of cognitive impairment in bipolar disorder (BD) and major depression compared with other conditions has not been clearly established, and the risk factors that drive cognitive variation within and across conditions are not well understood.

This chapter summarises previous research findings regarding the nature and impact of cognitive impairment in chronic behavioural and brain disorders, and highlights important gaps in knowledge. The potential to use large population cohorts and routine healthcare data to address some of these gaps is considered, and the aims and structure of this thesis are outlined.

1.1 Cognitive function and its assessment

Cognitive function refers to a range of mental processes that enable humans to perceive and make sense of the world and other people, and to acquire and use knowledge and skills that facilitate independence and success in everyday life. General cognitive ability, or intelligence, has been conceptualised as an underlying level of intellectual capability that can be captured empirically as a unitary latent factor contributing to performance across diverse cognitive domains and tasks (Deary, 2013; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). Such ability varies widely in the population and is substantially heritable (Plomin & Deary, 2015), and its development and decline over the life course has been the subject of much research. The distinct cognitive domains that reflect general intelligence include processing speed, memory, reasoning, language and spatial ability, and their neurological basis has been elucidated through structural and functional brain imaging studies (Deary, Penke, & Johnson, 2010). For example, magnetic resonance imaging (MRI) tractography has highlighted the importance of white matter tract integrity for speed of information processing across distributed networks involved in general cognitive performance (Penke et al., 2012).

Research on cognitive ability in the general population has primarily been situated within the fields of cognitive and differential psychology. Within neuropsychology, there has been a related but distinct emphasis on the effects of brain injury and illness on cognitive function. Neuropsychological research has focused on identifying patterns or profiles of cognitive dysfunction within and across different domains, and relating these to underlying neuropathology in clinical populations (Vakil, 2012). Neuropsychological assessments are typically structured around the core cognitive domains of processing speed, attention, memory, executive function (i.e. higher-order abilities such as planning, problem-solving and error monitoring), language and visuospatial performance (Burrell & Piguet, 2015; Goldstein & McNeil, 2013; Vakil, 2012). Overall assessment of general intelligence is also important, although often as a premorbid contextual factor against which current performance in each cognitive domain is interpreted (Schoenberg, Lange, Marsh, & Saklofske, 2011). Domain-specific deficits in cognitive performance have been documented following focal brain injuries (e.g. Stuss & Levine, 2002), and typical patterns of impairment across domains have been found to associate with clinical syndromes such as sub-types of dementia (Burrell & Piguet, 2015). The origins of modern clinical neuropsychology can be traced to work with brain injury survivors following the second world war and with neurosurgical patients in the mid-20th century, and the discipline has since developed to encompass assessment and intervention in a wide range of neurological, psychiatric and developmental conditions across the life course.

Whether working in cognitive, differential or neuropsychology traditions, psychologists make use of a wide range of standardised psychometric tests to measure cognitive function. Intelligence tests yield an overall score called the intelligence quotient (IQ), and can also produce separate domain-focused and task-specific scores, e.g. the verbal, perceptual, working memory and processing speed scales of the Wechsler Adult Intelligence Scale (Wechsler, 2010). Multi-domain test batteries, such as the Kaplan-Baycrest Neurocognitive Assessment (Richards, Rewilak, Kaplan, Proulx, & Leach, 2000), are often used in clinical practice to examine the profile of strengths and weaknesses across different cognitive domains. Computerised cognitive batteries such as the Cambridge Neuropsychological Test Automated Battery (CANTAB; Cambridge Cognition, 2017) and the National Institutes of Health (NIH) Toolbox (Weintraub et al., 2013) are now widely used in research, especially where efficient administration and scoring is required at scale in large studies.

Separate tasks or batteries are also available for the assessment of each cognitive domain (although it should be noted that there are no pure tests of any specific domain, because of the multiple demands inherent in understanding and responding during any assessment). Processing speed is typically assessed using psychomotor tests such as the Trailmaking Test Part A (Tombaugh, 2004) and the Symbol Digit Modalities Test (A. Smith, 1982), and reaction time can be measured with high precision using computerised tasks (e.g. Deary, Liewald, & Nissan, 2011). Batteries such as the Wechsler Memory Scale (Wechsler, 2011) assess various aspects of memory function, including learning, recall and recognition, in visual and verbal modalities, while other memory tests focus in detail on one modality (e.g. California Verbal Learning Test; Delis, Kramer, Kaplan, & Ober, 2000). Higher order or 'executive' functions, such as planning, problem-solving, attentional switching, abstract reasoning and response inhibition, are assessed by various batteries including the Delis-Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001) and the Behavioural Assessment of the Dysexecutive Syndrome (Wilson, Emslie, Evans, Alderman, & Burgess, 1996), or by standalone tasks such as the Trailmaking Test (Tombaugh, 2004), the Stroop task (Trenerry, Crosson, DeBoe, & Leber, 1989) and the Wisconsin Card Sorting Test (Grant & Berg, 1948). Separate tests are available for attention (e.g. Test of Everyday Attention; Robertson, Nimmo-Smith, Ward, & Ridgeway, 1994), language (e.g. Multilingual Aphasia Examination; Benton, Hamsher, & Sivan, 1994) and visuospatial function (e.g. Visual Object and Space Perception battery; Warrington & James, 1991). Tests have also been developed for social cognition skills, focusing on the ability to understand and behave appropriately in social situations (e.g. the Awareness of Social Inference Test; McDonald, Flanagan, & Rollins, 2002).

These standardised tests have usually been validated against similar established tasks or relevant clinical criteria, and they are expected to demonstrate satisfactory reliability, quantified as the proportion of variance in test scores that is true variance rather than measurement error (Crawford, 2012). Most tests produce standardised scores, whereby raw scores are transformed to a standard distribution with reference to a representative sample. Scoring metrics in common use include z-scores (mean 0 and standard deviation 1), standard scores (mean 100 and standard deviation 15) and percentiles; standardisation typically takes into account age-group and sometimes gender and education level. There is no specific score threshold below which performance is deemed to be impaired: the 2nd percentile is used as a guide in many tests (e.g. <70 on the Wechsler scales), but less stringent thresholds such as the 7th percentile are also employed (Strauss, Sherman, & Spreen, 2006, Ch. 5). These thresholds can be used to classify an individual as impaired

relative to his or her peer group, but they do not take into account individual contextual factors (e.g. the person's own premorbid level of ability).

1.2 Cognitive impairment in psychiatric and neurological conditions

Adults with chronic psychiatric and neurological disorders experience functional disability, reduced wellbeing and quality of life, and restricted social and economic participation. Cognitive impairment plays an important role in this. The overall detrimental impact of cognitive impairment on society has been termed the 'cognitive footprint' (Rossor & Knapp, 2015); this is conceptualised as a marker of the effect of cognitive impairment (from any cause) on such outcomes as education, employment, earnings and instrumental activities of daily living, which in turn influence national income distribution and economic growth. Because many neurological and mental health conditions have a young age of onset and a chronic course, and some are common in the population, their societal 'cognitive footprint' may be substantial and long-lasting. The cognitive burden of these chronic conditions is of great public health importance: echoing previous discussions on the 'mental wealth of nations' (Beddington et al., 2008), it has been argued that interventions to prevent or manage associated cognitive impairment have the potential "to foster cognitive health and to preserve cognitive capital" (Rossor & Knapp, 2015, p. 1008) at both individual and societal levels. This thesis will focus on psychiatric and neurological conditions with an average age of onset below 65 years, and whose chronic course means that affected individuals may live with cognitive impairment for many years.

1.2.1 Psychiatric conditions

Psychiatric conditions that are commonly associated with persistent cognitive impairment include bipolar disorder, major depression and schizophrenia. Cognitive impairment is also a feature of attention deficit hyperactivity disorder, but this will not be considered further here because of its distinct classification status as a neurodevelopmental disorder.

1.2.1.1 Mood disorders

Mood disorders are very common in the population, and follow a chronic or recurrent course in many of those affected. BD is characterised by episodes of abnormally elevated or irritable mood, encompassing symptoms such as grandiosity, hyperactivity and talkativeness, with resultant functional impairment or need for treatment (Gajwani, 2017);

many people with BD also experience episodes of depressed mood. BD may be sub-divided into type I and type II, with the latter denoting depressive and hypomanic episodes in the absence of frank mania (Gajwani, 2017). The core features of major depressive disorder are persistent low mood or loss of interest and pleasure (anhedonia), which are present for prolonged periods together with symptoms such as appetite or sleep disturbance, feelings of worthlessness, and loss of energy, causing functional impairment (D. F. MacKinnon, 2017). In the United Kingdom (UK), lifetime prevalence per 100,000 population is estimated at up to 10,000 for major depression (NICE, 2011) and approximately 1,000 for BD (Fajutrao, Locklear, Priaulx, & Heyes, 2009).

Research into cognitive impairment in mood disorders has increased in recent years, with studies investigating the pattern and persistence of cognitive dysfunction and its relationship to disorder subtype, severity and remission status. Several systematic reviews and meta-analyses have reported largely consistent findings in adults with BD. Impairment is typically found on tests of attention, working and episodic memory, processing speed and executive function, even in euthymia, with group differences of medium to large effect size compared with adults without a history of psychiatric illness (Arts, Jabben, Krabbendam, & van Os, 2008; Bortolato, Miskowiak, Koehler, Vieta, & Carvalho, 2015; Bostock, Kirkby, Garry, & Taylor, 2017; Bourne et al., 2013; Dickinson, Becerra, & Coombes, 2017; Mann-Wrobel, Carreno, & Dickinson, 2011; Raucher-Chene, Achim, Kaladjian, & Besche-Richard, 2017; Robinson et al., 2006). Meta-analyses of longitudinal studies in BD have found no evidence of accelerated cognitive decline over follow-up periods of up to five years (Bora & Ozerdem, 2017; Samamé, Martino, & Strejilevich, 2014), although other meta-analytic evidence indicates that BD history increases the odds of dementia diagnosis in later life (Diniz et al., 2017). Estimates of the prevalence of cognitive impairment in BD have ranged from 14-70% (Burdick et al., 2014; Gualtieri & Morgan, 2008; Kessing, 1998; Martino et al., 2014; Reichenberg et al., 2009; Szmulewicz, Samame, Martino, & Strejilevich, 2015), varying according to the sample composition, test used, and threshold applied to define impairment. Cognitive impairment has been highlighted as a key determinant of instrumental, occupational and social outcomes and quality of life in BD (Baune & Malhi, 2015; Duarte, Becerra, & Cruise, 2016; Gitlin & Miklowitz, 2017; P. D. Harvey, Wingo, Burdick, & Baldessarini, 2010). The growing recognition of this impact has sparked increasing interest in developing pharmacological and behavioural interventions to prevent, reduce or manage cognitive impairment in BD patients (Kluwe-Schiavon et al., 2015; Miskowiak, Carvalho, Vieta, & Kessing, 2016; Solé et al., 2017).

Studies that have directly compared cognitive performance between BD and major depression have generally found similar patterns across cognitive domains, although typically at a less severe level in the groups with major depression (Szmulewicz et al., 2017). The magnitude of differences between groups with remitted major depression and those with no psychiatric history is typically small, with greater deficits seen in late-onset depression patients (Bora, Harrison, Yucel, & Pantelis, 2013). Approximately one-third to one-half of patients with remitted major depression are estimated to show clinically significant levels of impairment (Rock, Roiser, Riedel, & Blackwell, 2014). As with BD patients, greater cognitive impairment in adults with major depression is associated with worse functional and psychosocial outcomes (Buist-Bouwman et al., 2008; Evans, Iverson, Yatham, & Lam, 2014), and the development of interventions for cognitive enhancement or remediation has been a key focus of recent research (Bortolato et al., 2016; Miskowiak, Ott, Petersen, & Kessing, 2016; Rosenblat, Kakar, & McIntyre, 2015; Salagre et al., 2017; Solé, Jimenez, Martinez-Aran, & Vieta, 2015).

1.2.1.2 Schizophrenia

Schizophrenia is a chronic and often highly disabling illness involving delusions, hallucinations, disordered thinking or speech, and so-called negative symptoms such as apathy and social withdrawal, together with lack of insight (Picchioni & Murray, 2007). The broader term ‘schizophrenia spectrum’ encompasses related conditions such as schizoaffective disorder (Bhati, 2013). The lifetime prevalence of schizophrenia is estimated to be approximately 400-1,000 per 100,000 (McGrath, Saha, Chant, & Welham, 2008), of whom up to 80% are thought to have cognitive impairment (Keefe & Fenton, 2007). Studies have consistently shown marked and persistent deficits in multiple cognitive domains (Fioravanti, Bianchi, & Cinti, 2012), with evidence of greater severity of impairment in patients with prominent negative symptoms (Bora, Akdede, & Alptekin, 2017). Many studies have directly compared cognitive functioning in schizophrenia and BD; domain profiles tend to be similar, but impairment is typically more severe in schizophrenia, and is reported to be evident from a younger age, often pre-dating illness onset (Bora, 2016; Bortolato et al., 2015; Trotta, Murray, & MacCabe, 2015; Vohringer et al., 2013). Poor functional outcomes are strongly associated with persistent cognitive deficits (Lepage, Bodnar, & Bowie, 2014), especially problems with social cognition (Fett et al., 2011). Cognitive remediation interventions have been extensively studied in this population (Kluwe-Schiavon, Sanvicente-Vieira, Kristensen, & Grassi-Oliveira, 2013;

Revell, Neill, Harte, Khan, & Drake, 2015; Wykes, Huddy, Cellard, McGurk, & Czobor, 2011).

1.2.2 Chronic neurological conditions

Many individuals in the working-age population are living with neurological disorders that affect cognitive functioning. It is of interest to compare the prevalence and pattern of cognitive impairment in mood disorders and schizophrenia with that in chronic neurological conditions, because they each contribute to the 'cognitive footprint' at the societal level, and they may share risk factors that contribute to cognitive outcome. The focus here will be on MS and Parkinson's disease (PD), as their average age of onset is below 65 years and they share psychiatric phenotypic features with the disorders discussed above.

1.2.2.1 Multiple sclerosis

MS is a degenerative disease of the central nervous system, which is thought to be autoimmune in origin (Dendrou, Fugger, & Friese, 2015). It causes characteristic demyelinating lesions—primarily in the brain white matter and the spinal cord but also evident in cortical and subcortical grey matter—together with axonal damage and atrophy (Matthews et al., 2016). Clinically, this leads to symptoms including sensory and motor dysfunction, fatigue, pain and depression, which may follow a relapsing-remitting or progressive course (Kamm, Uitdehaag, & Polman, 2014). The lifetime prevalence of MS in the UK population is approximately 200 per 100,000 (MacDonald, Cockerell, Sander, & Shorvon, 2000). The typical cognitive profile involves prominent processing speed deficits, as well as impairment of attention, executive function, memory and visuospatial skills (Brito Ferreira, 2010; Guimaraes & Sa, 2012). Cognitive impairment severity is only partly explained by severity of structural brain changes (Bobholz & Rao, 2003). Prevalence of impairment has been estimated at 40-80%, depending on the threshold applied (Fischer et al., 2014; Patti et al., 2015; Rao, Leo, Bernardin, & Unverzagt, 1991). Over and above the impact of physical disability and illness duration, patients with cognitive impairment experience generally worse domestic, occupational and social outcomes (Rao, Leo, Ellington, et al., 1991). The evidence base for cognitive rehabilitation programmes in MS is mixed (Mitolo, Venneri, Wilkinson, & Sharrack, 2015), and more research is also needed regarding the potential efficacy of stimulant medications such as methylphenidate and L-amphetamine (Lovera & Kovner, 2012). There is some evidence that the disease modifying treatments interferon β-1a and natalizumab have a beneficial impact on

cognitive function in the short term, but long term evidence is lacking (Lovera & Kovner, 2012).

1.2.2.2 Parkinson's disease

PD is a progressive disease involving the degeneration of dopaminergic neurons in the substantia nigra pars compacta and the proliferation of α-synuclein inclusions, known as Lewy bodies, within the neuronal cell bodies. The resultant dopamine deficiency leads to a clinical picture of movement disorder, including slowing (bradykinesia), rigidity, resting tremor, postural instability and gait impairment (Kalia & Lang, 2015). Non-motor symptoms such as mood disturbance, apathy and sleep disorders are also common (Poewe, 2008). The lifetime prevalence of PD is similar to that of MS, at around 200 per 100,000 (MacDonald et al., 2000). Approximately 50-55% are estimated to experience cognitive impairment (Svenningsson, Westman, Ballard, & Aarsland, 2012), about half of whom meet criteria for dementia (Aarsland, Zaccai, & Brayne, 2005). Cognitive impairment is associated with increasing age, disease duration and disease severity (Litvan et al., 2011), and the domains affected typically include executive function, attention, processing speed and visuospatial skills. The severity of cognitive impairment has been found to be associated with measures of quality of life (A. J. Mitchell, Kemp, Benito-Leon, & Reuber, 2010). There is some evidence for cognitive benefits from acetylcholinesterase inhibitors (Pagano et al., 2015), and modest benefits have also been reported from non-pharmacological interventions such as cognitive training (Hindle, Petrelli, Clare, & Kalbe, 2013; Leung et al., 2015).

1.2.3 Limitations of current research evidence

The literature on cognitive function in psychiatric and neurological conditions primarily documents average between-group score differences on cognitive tests, with prevalence of cognitive impairment reported much less frequently. Where prevalence estimates are available for a particular condition, these are difficult to synthesise, owing to heterogeneity in the populations studied (e.g. clinic- or community-based), in the tests used, and in the thresholds applied to define impairment. Systematic reviews of cognitive impairment prevalence are lacking, and there are also challenges in comparing prevalence estimates across different conditions, again because of heterogeneity in recruitment, assessment and impairment definitions. In order to quantify and understand the cognitive footprint of psychiatric and neurological conditions, it is essential that we obtain clearer estimates of cognitive impairment prevalence within and across these conditions; this will contribute to

a greater appreciation of the scale of clinical and service need within each clinical group, as well as highlighting the population-level impact of cognitive impairment associated with different conditions, some of which (e.g. major depression) are considerably more common than others.

In studies that have investigated differences in cognitive test performance between groups, approaches to accounting for confounding influences have been inconsistent. Some studies have used groups that are matched on characteristics such as age and gender, while others have used statistical methods to adjust between-group difference estimates, or to measure the strength of association between cognitive performance and sociodemographic, clinical and lifestyle factors. The selection of variables used for matching and statistical adjustment has differed from study to study, and systematic distinctions have not been drawn between potential confounders (i.e. background factors that are not themselves a consequence of the clinical disorder) and other kinds of variables that may co-vary with clinical status and cognitive outcome in other ways (e.g. intermediate factors such as mood state or medication use). The collective inclusion of such variables without distinction in multiple regression analyses, for example, makes it difficult to interpret the coefficient estimates, and to move towards causal explanations for variation in cognitive function in psychiatric and neurological conditions. Such causal explanations require careful modelling of a range of potential confounding, mediating and moderating factors, acknowledging the complexity of their inter-relationships with the clinical disorder, the cognitive outcome, and each other. Although path analysis and mediation modelling are commonly used in psychological research (Gelfand, Mensinger, & Tenhave, 2009), the application of formal causal inference frameworks and related statistical techniques in psychiatric epidemiology is not yet widespread (Kendler, 2017). This endeavour requires large samples with data on a wide range of relevant variables, which is difficult to achieve in clinic-based studies.

1.3 Population cohorts in cognitive and mental health research

There is increasing interest in taking a ‘data science’ approach to mental health research (McIntosh et al., 2016), which has been facilitated by computational and statistical advances that permit analysis of high dimensional datasets. Relevant data sources include dedicated research cohorts as well as routine healthcare and administrative records (Stewart & Davis, 2016). A recent review by the UK Medical Research Council (MRC, 2014) described 34 UK-led large population research cohorts, of which 21 contain data on

mental health together with direct measures of cognitive function. By far the largest of these is UK Biobank, which recruited more than 502,000 adults at baseline (Sudlow et al., 2015).

An obvious advantage of population cohorts is their typically large size, which means they are more likely to provide sufficient statistical power to conduct multivariable analyses with high precision. General population cohorts may also provide broader representation than those recruited from clinics, in that they will include individuals with a psychiatric or neurological history who are not necessarily in contact with specialist clinical services. Conversely, more severely affected individuals may be less likely to join population cohort studies. Other benefits of population cohorts include the breadth of data collected, which typically encompasses demographic, lifestyle, clinical, physiological and psychosocial measures, as well as the necessary resources and administrative capability for re-contact and linkage to other records and registries.

The empirical studies in this thesis use data from the UK Biobank cohort. As described in detail in Chapter 3, UK Biobank collected mental health and cognitive data on an unprecedented scale, together with a broad range of other relevant measures including genotyping and linkage to hospital records. These permit comprehensive investigation of the complex relationship between psychiatric and neurological health and cognitive function. Initial analyses of cognitive function in the whole cohort and in clinical sub-groups have previously been reported by our group (Cullen et al., 2015; Lyall et al., 2016), and this thesis develops this work by investigating prevalence and by applying a formal causal inference approach to understand patterns of cognitive performance.

1.4 Thesis aims

The aims of this thesis were to investigate the prevalence of cognitive impairment in psychiatric and neurological conditions, and to understand the risk factors that may explain variation in cognitive function in these conditions. The primary focus was on BD, as this disorder is relatively common in the population, and cognitive impairment is thought to be a major cause of chronic disability even in euthymia. Comparisons were also made with major depression, schizophrenia, MS and PD.

1.4.1 Research questions

Two research questions were addressed, via a systematic literature review and a series of primary data analyses.

1.4.1.1 What is the prevalence of cognitive impairment in BD, and in other psychiatric and neurological conditions?

This question was investigated by (i) undertaking a systematic review of studies reporting the prevalence of cognitive impairment in euthymic BD and (ii) conducting an empirical study of baseline data from the UK Biobank cohort, to estimate the prevalence of cognitive impairment in BD, major depression, schizophrenia, MS and PD.

1.4.1.2 What explains variation in cognitive function in BD and other conditions?

This question was investigated by conducting a series of multivariable analyses of baseline UK Biobank data, using a causal inference framework to estimate the overall effect of BD and other disorders on cognitive function, and the magnitude of the effect that is transmitted through potentially modifiable intermediate factors.

1.5 Overview of thesis structure

Chapter 2 presents the systematic review of cognitive impairment prevalence in euthymic BD. Chapter 3 describes the UK Biobank resource, Chapter 4 outlines the ascertainment and characteristics of the psychiatric and neurological groups within this resource, and Chapter 5 reports the prevalence of cognitive impairment in those groups. An overview of causal inference methods, as applied to observational data, is given in Chapter 6. Chapter 7 presents estimates of the overall effect of BD on cognitive performance and indirect effects through potentially modifiable intermediate factors, and Chapter 8 presents similar analyses for other psychiatric conditions. The key results, implications and future directions are discussed in Chapter 9.

Chapter 2 Systematic review of the prevalence of cognitive impairment in bipolar disorder

This chapter focuses on cognitive impairment in bipolar disorder. As outlined in Chapter 1, it is particularly important to study BD because it is common and chronic, and risk factors for poor cognitive function in people with this condition are currently not well understood. This chapter describes a systematic review carried out to establish the prevalence of cognitive impairment in people in the euthymic phase of BD, and to identify factors which may be associated with cognitive impairment.[1]

2.1 Background

Adults with BD show impairment in cognitive domains including attention, memory, processing speed and executive function, even when in remitted or euthymic phases (Arts et al., 2008; Bortolato et al., 2015; Bostock et al., 2017; Bourne et al., 2013; Dickinson et al., 2017; Mann-Wrobel et al., 2011; Raucher-Chene et al., 2017; Robinson et al., 2006). Although group-level differences of medium to large effect size have been consistently reported, the proportion of adults with BD who have clinically relevant levels of cognitive impairment has not yet been clearly established. There are a number of reasons why it would be beneficial to establish the prevalence of cognitive impairment in the BD population. From a clinical point of view, cognitive impairment is a major contributor to the overall burden of disability in mood disorders, and is a target in its own right for therapeutic intervention. Service planning would be helped by clearer information about the numbers and characteristics of those who are likely to need more health or social care input to manage the disabling effects of cognitive impairment. From a research perspective, shifting our focus to identifying subgroups with cognitive impairment will facilitate efforts to understand why some people with BD experience significant problems with cognitive function while others remain unimpaired. This, in turn, may help to identify particular risk factors for clinically significant cognitive impairment.

[1] The work described in this chapter has been published in Cullen, B., Ward, J., Graham, N. A., Deary, I. J., Pell, J. P., Smith, D. J., & Evans, J. J. (2016). Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder: A systematic review. *Journal of Affective Disorders, 205*, 165-181. doi:10.1016/j.jad.2016.06.063

2.1.1 Objectives

The objectives of the review were: (1) to determine the prevalence of cognitive impairment in euthymic adults with a history of BD; and (2) to describe sociodemographic, clinical and other factors that are associated with cognitive impairment in BD.

2.1.2 Scope of review

The population of interest was community-dwelling adults with a history of BD (the exposure), who were euthymic at the time of assessment. The outcome of interest was cognitive impairment, measured using standardised tests; presence or absence of impairment was defined with reference to published test norms or scores obtained by a comparison group without BD. Since the aim was to determine prevalence, only cross-sectional results were considered (cross-sectional studies, or baseline results from cohort studies or trials).

2.2 Methods

The review was conducted according to a structured protocol which followed Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidance (Moher et al., 2015). The protocol was registered on the International Prospective Register of Systematic Reviews (PROSPERO) database on 16 March 2015 (reference number CRD42015017558; see Appendix A). Reporting is in accordance with PRISMA and Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009; Stroup et al., 2000).

2.2.1 Eligibility criteria

The following inclusion criteria were applied during the search and screening process: original research published in peer-reviewed journals from 1994 onwards (the year that the Diagnostic and Statistical Manual of Mental Disorders [DSM-IV] and International Classification of Diseases [ICD-10] diagnostic classifications came into use); articles published in English; studies of community-dwelling adults (not hospital in-patients) aged 18 to 70 years inclusive (to minimise the additional contribution of age-related cognitive decline); cross-sectional studies or baseline results from cohort studies or trials; clinical samples must have been recruited consecutively from clinics or via a method that ensured eligible individuals in the target population had an equal chance of being approached (so

that prevalence estimates would be based on representative samples); primary diagnosis of BD; euthymic at time of assessment; assessed using at least one direct, standardised, objective cognitive measure. Articles were excluded if samples were selected on the basis of presence/absence of cognitive impairment (known or suspected).

2.2.2 Concepts and definitions

The following definitions were applied for each construct:

2.2.2.1 Bipolar disorder

History of bipolar disorder type I, II or not otherwise specified, meeting defined criteria (e.g. DSM or ICD).

2.2.2.2 Euthymia

Not meeting defined criteria for a depressive or manic episode at time of cognitive assessment; or as otherwise defined by the study authors based on an appropriate clinical measure.

2.2.2.3 Cognitive impairment

Evidence of impaired performance on one or more objective cognitive tests. Impairment was defined as the fail range on a pass/fail test, or as otherwise defined by the study authors with reference to the score distribution of an unexposed (“healthy”) comparison group (e.g. from published test norms, or an appropriate comparison group recruited to the study). Results based on any threshold that was less strict than 1 standard deviation (SD) below the comparison mean were not considered.

2.2.2.4 Prevalence

Assessments must have been conducted at a single time point, yielding a point prevalence estimate of cognitive impairment, reported as the proportion of the sample falling below the cut-off for impairment.

2.2.2.5 Correlates

Any sociodemographic, clinical or other factor that was reported by the authors to be statistically associated with the presence or severity of cognitive impairment.

2.2.3 Search strategy

2.2.3.1 Information sources

The following electronic databases were searched on 24 February 2015: Web of Science (Thomson Reuters), including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Current Contents Connect, Data Citation Index, MEDLINE and SciELO Citation Index; PubMed (NCBI), including MEDLINE, PubMed Central and in-process/ahead-of-print citations; EBSCOhost (EBSCO), including CINAHL and PsycINFO. Additional articles published up to the search date were sought via: the 'cited by' function within individual electronic records of relevant articles; hand searching of reference lists of relevant articles and recent review papers; and email contact with study authors.

2.2.3.2 Process for study identification and selection

A detailed search strategy was developed and tailored for each electronic database. Controlled vocabulary and free text variations were used, including synonyms, abbreviations and spelling variants. Appendix B shows the search strategy as implemented in Web of Science. Search outputs were managed using EndNote software (http://endnote.com/).

Duplicate records were removed, and study titles and/or abstracts were screened for relevance by B.C. Screening was carried out with reference to a detailed checklist of eligibility criteria; this was piloted by B.C. and a second researcher independently against a sample of initial search results, and refined as required (see Appendix C for the final checklist). The sensitivity of the search strategy was checked by testing whether key papers that were known to be relevant were detected by the search. Reproducibility was assessed by the second researcher, who independently ran the search in one electronic database (Web of Science) and screened the first 200 titles and/or abstracts for relevance. Agreement was 93% (100% following consensus discussion). Full text was obtained for all potentially relevant papers that remained. These were assessed by B.C. using the eligibility checklist, with the second researcher independently assessing the first 100 papers for comparison. Agreement was 95% (100% following consensus discussion). Reasons for exclusion were documented.

2.2.4 Data extraction

A spreadsheet template was used for extracting data from included papers, having been piloted by B.C. and a second researcher independently. The list of data fields is given in Appendix D. Data extraction was carried out by B.C., following which the second researcher compared four randomly-chosen data extraction records against the source papers to check for accuracy and completeness; no discrepancies were identified. Where authors appeared to have collected data that could be used to report the prevalence of impairment, but had not reported the prevalence explicitly in the paper (e.g. articles only reporting group mean differences), the authors were contacted via email to request prevalence results using an appropriate cut-off of their choice.

2.2.5 Assessment of risk of bias

2.2.5.1 Risk of bias within studies

Each included study was assessed for risk of bias using a critical appraisal tool for systematic reviews addressing questions of prevalence (Munn, Moola, Riitano, & Lisy, 2014). Reporting bias was assessed using the STROBE checklist for cross-sectional studies (von Elm et al., 2007). B.C. and a second researcher independently rated randomly-chosen articles for comparison, followed by consensus discussion. Initial rating concordance was 83-95% across four articles, and 93% when one further article was independently assessed following the consensus discussion exercise. Subsequent ratings were made by B.C. only. These assessments were considered in the synthesis and discussion, in order to comment on the quality of the literature in this field and to aid interpretation of the results.

2.2.5.2 Risk of bias across studies

Funnel plots were generated using the `metafunnel` package in Stata software version 13 (StataCorp, 2013) to allow visual inspection of the relationship between the magnitude and precision of the prevalence estimates. These are scatter plots depicting a measure of study size (e.g. sample size or standard error of the effect estimate) on the vertical axis against the study's effect estimate on the horizontal axis. Larger (more precise) studies are expected to have effect estimates close to the centre on the horizontal axis, and smaller studies are expected to have effect estimates scattered symmetrically about the centre. Asymmetry in this characteristic inverted funnel shape indicates "small study bias", for example resulting from publication bias (Egger, Smith, Schneider, & Minder, 1997).

2.2.6 Data synthesis

Where one study population was analysed in two or more eligible articles, the article reporting the largest sample was included in the data synthesis. Additional articles were only included if they contributed unique relevant information (e.g. additional cognitive measures). A narrative synthesis was conducted, alongside summary tables of extracted data, and forest plots of impairment prevalence estimates and 95% confidence intervals (CI) by cognitive domain. Forest plots were generated using the `metan` package in Stata v13. Sociodemographic, clinical and other variables that were reported to be associated with cognitive impairment were summarised, and consistency in these findings was compared across studies. Only variables that were potential risk factors for cognitive impairment were included; variables that were viewed as consequences of that impairment (e.g. occupational status, instrumental functioning) were not considered, on the basis that they are not potential causal, mediating or moderating factors in explaining the association between BD status and cognitive impairment.

2.3 Results

2.3.1 Article selection

Figure 2.1 shows a PRISMA flow diagram of the article selection process. Titles and/or abstracts of 5,412 records were screened for eligibility, followed by full text evaluation of 658 papers. Forty-six articles were deemed eligible. The most common reasons for exclusion were lack of evidence of consecutive sample recruitment, inclusion of in-patients in study samples, and inclusion of non-euthymic participants. Examples of acceptable sample recruitment methods in the eligible articles were systematic invitation of: consecutively attending eligible patients at out-patient clinics; all eligible patients on a database of open records at a specific clinical service; all eligible persons identified via national registers during a specific period. Of the 46 eligible articles, 16 were omitted from the data synthesis (see list in Appendix E): 11 reported on overlapping samples without contributing relevant additional information, and for a further five, results directly addressing either of the two research questions of this review were unavailable.

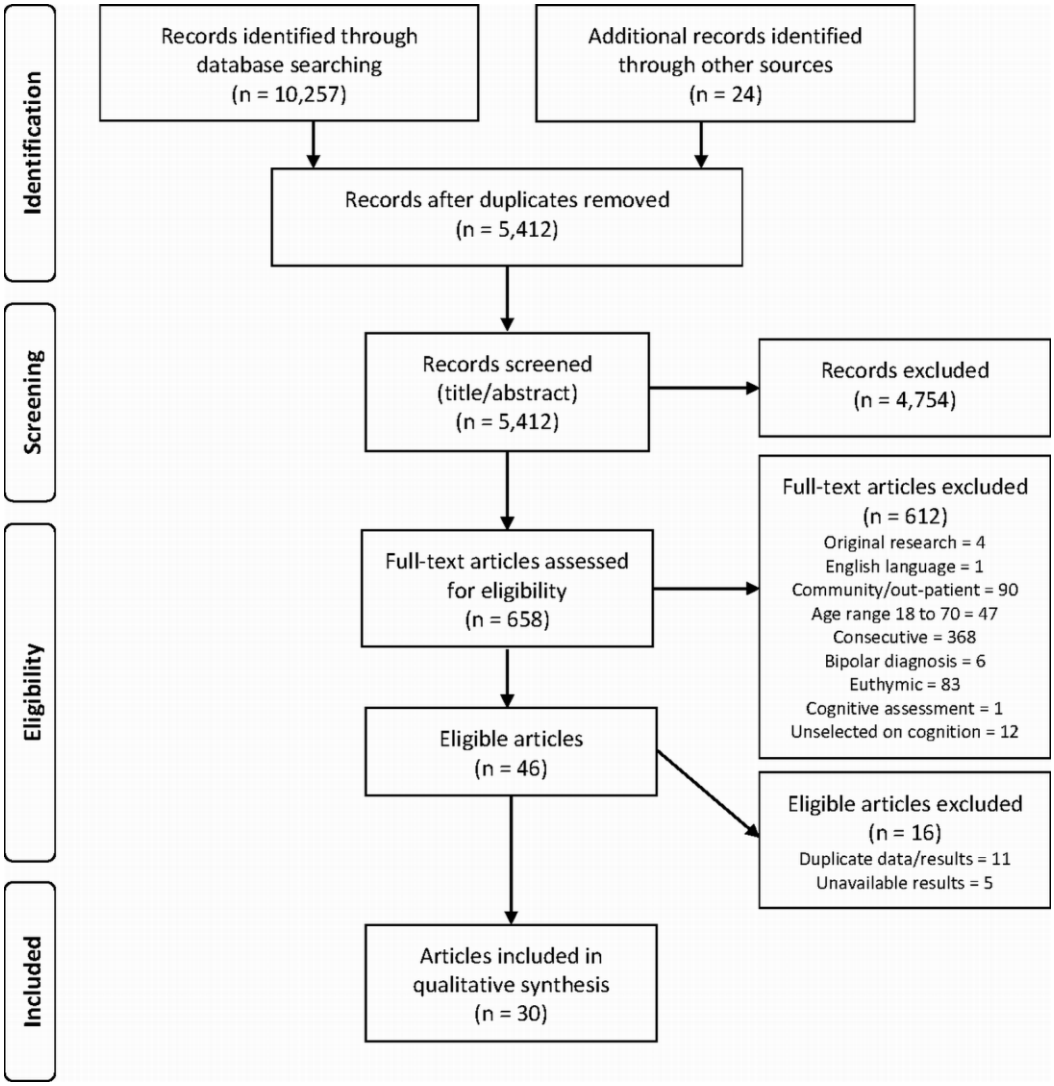


Figure 2.1 - PRISMA flow diagram showing results of literature search and screening

The possible reasons for exclusion are explained in the eligibility checklist in Appendix C.

2.3.2 Characteristics of included studies

Key characteristics of the 30 included articles are summarised in Table 2.1. The majority included BD-I samples only (Altshuler et al., 2004; Arslan, Tiryaki, & Ozkorumak, 2014; Cavanagh, Van Beck, Muir, & Blackwood, 2002; Cheung, Halari, Cheng, Leung, & Young, 2013; Doganavsargil-Baysal et al., 2013; Fakhry, El Ghonemy, & Salem, 2013; Ferrier, Stanton, Kelly, & Scott, 1999; Frangou, Donaldson, Hadjulis, Landau, & Goldstein, 2005; Goswami et al., 2009; Ibrahim, Rahman, & Shah, 2009; Jamrozinski, Gruber, Kemmer, Falkai, & Scherk, 2009; Juselius, Kieseppa, Kaprio, Lonnqvist, & Tuulio-Henriksson, 2009; Kieseppa et al., 2005; Lopera-Vasquez, Bell, & López-Jaramillo, 2011; Lopez-Jaramillo et al., 2010; Normala et al., 2010; Osher, Dobron, Belmaker, Bersudsky, & Dwolatzky, 2011; Pirkola et al., 2005). A further eight articles

reported on mixed BD samples (Barrera, Vazquez, Tannenhaus, Lolich, & Herbst, 2013; Daban et al., 2012; Elshahawi et al., 2011; Martino et al., 2014; Martino et al., 2008; Mur, Portella, Martinez-Aran, Pifarre, & Vieta, 2007; Sánchez-Morla et al., 2009; van der Werf-Eldering, Burger, Holthausen, Aleman, & Nolen, 2010) and four articles included separate BD-I and BD-II samples (Martino, Igoa, Marengo, Scapola, & Strejilevich, 2011a; Martino, Strejilevich, Fassi, Marengo, & Igoa, 2011b; Martino, Strejilevich, Torralva, & Manes, 2011c; Sparding et al., 2015). Three articles reported on samples recruited from population registers of twin births and hospital discharges (Juselius et al., 2009; Kieseppa et al., 2005; Pirkola et al., 2005) and the rest recruited from specialist psychiatry clinics. Definitions of euthymia differed across studies; many used the Hamilton Rating Scale for Depression (HRSD) and the Young Mania Rating Scale (YMRS), but score thresholds varied. Most studies excluded participants with major psychiatric, neurological or medical comorbidity or learning disability, and many also excluded those with recent substance misuse or electroconvulsive therapy.

Ratings of methodological and reporting bias are shown in Appendix F and Appendix G, respectively. Although all the studies aimed to recruit representative participants using consecutive or random methods, nine of 30 articles included samples which were unrepresentative of the BD population with regard to gender balance, and two did not report gender composition. Most articles did not report numbers of patients initially considered or deemed eligible, or information about comparability of eligible patients who did and did not participate; there was evidence of adequate coverage of the intended population in only four articles. Sample sizes were generally small. All articles reported on objective cognitive measures, but 13 did not report sufficient information to allow appraisal of measurement reliability (e.g. qualifications and training of assessors; inter-rater reliability data). Most did not report adequate consideration of sources of bias or imprecision in their procedures or interpretation.

Table 2.1 - Characteristics of the included articles

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
Altshuler 2004	USA	40	22	BD-I DSM-III-R	HRSD <6 and YMRS <7 for 3 consecutive months	Head injury with LOC >1 hour; learning disability; migraine; liver function abnormalities; alcoholic dementia; abuse of alcohol in past 6 months; history of cocaine abuse/dependence; diabetes; hypertension; seizure disorder; any other neurologic illness; left-handedness; ECT in past 2 years; other current DSM-III-R Axis I disorder
Arslan 2014	Turkey	30	32	BD-I DSM-IV	HRSD <7 and MADRS <12 and YMRS ≤12	DSM Axis I comorbidities; mental retardation; hearing/visual loss interfering with clinical interview; alcohol/substance abuse in past 6 months; any disease affecting CNS; head trauma with LOC
Barrera 2013	Argentina	12	12	BD-I or BD-II Not stated	HRSD (17 items) <7 and YMRS <8	Significant medical diseases; neurological disorders; mental deficiency; drug abuse
Cavanagh 2002	UK	20	20	BD-I DSM-IV	HRSD ≤7 and MMS ≤2	Significant physical or neurological illness; stroke or head trauma; significant alcohol and/or drug misuse; ECT in past 6 months; comorbid psychiatric disorder; neurodegenerative disorder; learning disability; endocrine abnormalities
Cheung 2013	China (Hong Kong)	52	52	BD-I ICD-10 and DSM-IV	HRSD (21 items) <7 and YMRS <7 on two occasions 4 weeks apart	Mental retardation; change in psychotropic medication in past 4 weeks; DSM-IV alcohol/substance abuse in past 12 months; head injury with LOC; neurological disorder; history of psychiatric illness other than BD-I; significant physical health problem which could affect cognition
Daban 2012	France	53	60	BD DSM-III-R	MADRS <6 and BR-MRS <5	History of severe head trauma; learning difficulties; neurological disorder; current alcohol/drug abuse
Doganavsargil-Baysal 2013	Turkey	54	18	BD-I DSM-IV-TR	HRSD ≤7 and YMRS ≤5	Comorbid psychiatric or neurological disorders; IQ score <80; infectious or autoimmune diseases; on anti-inflammatory or antibiotic medication; biochemical values not within normal range
Elshahawi 2011	Egypt	50	50	BD-I or BD-II; history of ≥3 affective episodes ICD-10	HRSD <8 and YMRS <6	Comorbid psychiatric disorder; ECT in past 3 months; neurological disorder; mental retardation; substance abuse; organic cause of cognitive impairment
Fakhry 2013	UAE	30 (recent manic episode) 30 (recent depressive episode)	30	BD-I; history of ≤3 affective episodes; illness duration <5 years DSM-IV	MES and MAS <6; free from symptoms for at least 8 weeks and not fulfilling DSM-IV criteria for an affective episode	Comorbid psychiatric disorders; ECT in past 6 months; lithium-receiving patients in a trial

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
Ferrier 1999	UK	21 ('good' outcome) 20 ('poor' outcome)	20	BD-I; at least 5 years illness duration DSM-IV	HRSD ≤8 and MSS <20	Dementing disorder; learning disability; history of substance misuse, cerebrovascular disease, neurodegenerative disorders, head injury with concussion, clinical epilepsy, systemic illness with known cerebral consequences, severe hypertension, severe hepatic or renal disorder, or endocrine disorders other than corrected hypothyroidism
Frangou 2005	UK	44	44	BD-I DSM-IV	Syndromal remission: not meeting DSM-IV criteria for a mood episode for at least 3 months; no change in medication type/dose over the same period. Symptomatic remission: HRSD and MRS-SADS <10	None
Goswami 2009	India	22 (on medication) 22 (not on medication)	NA	BD-I DSM-IV	HRSD <8 and MSS <20 on two occasions 4 weeks apart	Other DSM Axis I or II diagnoses; cardiorespiratory, gastrointestinal, neurological and endocrine disorders (other than corrected hypothyroidism); substance misuse/dependency disorders; other medications e.g. anticholinergics, hypnotics or steroids
Ibrahim[a] 2009	Malaysia	40	40	BD-I DSM-IV	No active manic or depressive symptoms as reflected by YMRS and HRSD scores	Overtly disturbed/aggressive; severe mental retardation; dementia; significant CNS disease; head injury; comorbid psychiatric disorders; substance abuse/dependence; use of anticholinergics or benzodiazepines
Jamrozinski 2009	Germany	40	40	BD-I DSM-IV	MADRS ≤10 and YMRS ≤12	Other medical disorders
Juselius[b] 2009	Finland	26	114	BD-I DSM-IV (past diagnosis using ICD-8 or DSM-III-R)	In remission according to DSM-IV criteria	Other psychotic disorders; neurological disorders; brain injury; current alcohol dependence
Kieseppä[b] 2005	Finland	26	114	BD-I DSM-IV (past diagnosis using ICD-8 or DSM-III-R)	In full symptom remission according to DSM-IV criteria	Other psychotic disorders; neurological disorders; brain injury; current alcohol dependence
Lopera-Vásquez 2011	Colombia	40 (on medication) 31 (not on medication)	28	BD-I DSM-IV	ZSDS <8 and YMRS <6	Illicit substances or benzodiazepines in past 4 weeks; other psychiatric or neurological disorders; mental retardation; any treatment with ECT

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
López-Jaramillo 2010	Colombia	24 (1 manic episode) 27 (2 manic episodes) 47 (≥3 manic episodes)	66	BD-I DSM-IV	HRSD <8 and YMRS <6	Illicit substances or benzodiazepines in past 4 weeks; other psychiatric or neurological disorders that could affect cognition; mental retardation; any treatment with ECT; physical/sensory limitations that could affect performance
Martino[c] 2008	Argentina	50	30	BD-I or BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 6 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011a	Argentina	48 (BD-I) 39 (BD-II)	39	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011b	Argentina	45 (BD-I) 36 (BD-II)	34	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011c	Argentina	48 (BD-I) 37 (BD-II)	34	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2014	Argentina	100	40	BD-I or BD-II DSM-IV	HRSD ≤9 and YMRS ≤8 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Mur 2007	Spain	44	46	BD-I or BD-II DSM-IV	HRSD (17-item) <8 and YMRS <6 for at least 3 months; on same treatment regimen and clinically stable for 3 months	Significant physical or neurologic illness; substance abuse/dependence in the past year; ECT in the past year; any mood-stabilising medication other than lithium
Normala[a] 2010	Malaysia	40	40	BD-I DSM-IV	No active manic or depressive symptoms as reflected by YMRS and HRSD scores	Overtly disturbed/aggressive; severe mental retardation; dementia; significant CNS disease; head injury; comorbid psychiatric disorders; substance abuse/dependence; use of anticholinergics or benzodiazepines
Osher 2011	Israel	51	495	BD-I DSM-IV	Consensus judgement by two clinicians based on full history and evidence of stability for at least three months[d]	Serious physical illness or substance abuse
Pirkola[b] 2005	Finland	22	100	BD-I DSM-III-R or DSM-IV	Not stated	Schizoaffective disorder; psychotic disorder other than BD-I; neurological disease; clinically significant head injury; mental retardation
Sánchez-Morla 2009	Spain	73	67	BD DSM-IV	HRSD <7 and YMRS <6 for 3 consecutive monthly evaluations	Neurological or medical diseases that can affect cognition; mental retardation; history of alcohol or other substance abuse/dependence in past 2 years; ECT in past 2 years; history of head injury with LOC
Sparding 2015	Sweden	64 (BD-I) 44 (BD-II)	86	BD-I; BD-II DSM-IV	MADRS and YMRS <14	None stated
van der Werf-Eldering 2010	The Netherlands	46	75	BD-I, BD-II or BD-NOS DSM-IV	IDS-SR <14 and YMRS <8	Mental retardation; systemic or neurological disease which could affect cognition; alcohol use disorder currently needing treatment in a specialised setting

BD, bipolar disorder; BD-I, bipolar disorder type I; BD-II, bipolar disorder type II; BD-NOS, bipolar disorder not otherwise specified; BR-MRS, Bech–Rafaelsen Mania Rating Scale; CNS, central nervous system; DSM, Diagnostic and Statistical Manual of Mental Disorders; DSM-III-R, Diagnostic and Statistical Manual of Mental Disorders third edition revised; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders fourth edition; DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders fourth edition text revision; ECT, electroconvulsive therapy; HC, healthy comparison; HRSD, Hamilton Rating Scale for Depression; ICD-8, International Classification of Diseases eighth revision; ICD-10, International Classification of Diseases tenth revision; IDS-SR, Inventory of Depressive Symptomatology - Self Rating; IQ, intelligence quotient; LOC, loss of consciousness; MADRS, Montgomery–Åsberg Depression Rating Scale; MAS, Bech–Rafaelsen Mania Scale; MES, Bech–Rafaelsen Melancholia Scale; MRS-SADS, Mania Rating Scale from the Schedule for Affective Disorders and Schizophrenia (Change Version); MSS, Bech's modification of Beigel's Mania State Rating Scale; NA, not applicable; YMRS, Young Mania Rating Scale; ZSDS, Zung Self-Rated Depression Scale.
a. Studies contain overlapping samples.
b. Studies contain overlapping samples.
c. Studies contain overlapping samples.
d. Information provided by author.

2.3.3 Prevalence of cognitive impairment

Prevalence was available for 15 articles, reporting on 16 BD samples. Tables 2.2 and 2.3 show prevalence results in BD-I only and mixed BD samples, respectively. Characteristics of these samples are provided in supplementary tables within Appendix H. Prevalence was available for one study with separate BD-I and BD-II samples (Appendix I).

Studies applied a variety of impairment thresholds: some were simple pass/fail cut-offs, and others were based on score distributions from published test norms or from a healthy comparison group. Distribution-based thresholds ranged from 1 SD to 2 SD below the comparison mean, with the most common being 1.5 SD (approximately 7th percentile), 1.64 SD (approximately 5th percentile) and 2 SD (approximately 2nd percentile). At every threshold and on almost all cognitive measures, the prevalence of impairment in BD samples was higher than in the comparison group. Heterogeneity in prevalence across studies did not clearly relate to study quality/risk of bias. Studies differed in whether they used comparison group score distributions or published norms as the reference for impairment, but there was no clear relationship between choice of reference and magnitude of impairment prevalence. For example, on the same tests at the same thresholds, Mur et al. (2007) used published norms and reported lower prevalence estimates than Juselius et al. (2009), who used their own comparison group. On the other hand, Cheung et al. (2013) used published norms and reported some of the highest prevalence estimates across several cognitive domains. Prevalence of impairment did not differ consistently between BD-I only (Table 2.2) and mixed BD samples (Table 2.3), although direct comparison is difficult owing to the variation in measures and thresholds used. In the only study where BD-I and BD-II samples could be directly compared (Sparding et al., 2015) (results given in Appendix I), prevalence was higher in the BD-I participants on several measures, but there was considerable overlap in estimates between the two samples.

Prevalence of cognitive impairment was further considered according to cognitive domain. Results within domains are presented graphically using forest plots, but pooled estimates are not reported because of the wide variation in cognitive tests used and in cut-offs applied to define presence of impairment. The classification of tests by domain was guided by the classifications used by the authors of the original articles. Where tests were thought to cross multiple domains, this is indicated in Tables 2.2 and 2.3.

Table 2.2 - Prevalence of cognitive impairment in BD-I samples

Author Year	Sample *n*[a]		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Altshuler 2004	40	22	WCST categories (executive)	Score 0 to 3	(42%)	(0%)	-0.98
			CVLT total recall 1-5 (verbal memory)	1.75 SD from published normative mean[b]	(22%)	(0%)	-0.99
Cavanagh 2002[c]	20	20	Stroop Color-Word Test (executive)	1.64 SD from HC mean	7 (36.8%)	(5%)[d]	-0.61
			Letter fluency (executive)		2 (10%)	(5%)[d]	-0.31
			BADS Six Elements (executive)		1 (5.3%)	(5%)[d]	-0.31
			CVLT trial 1 (verbal memory)		7 (35%)	(5%)[d]	-1.24
			CVLT total recall 1-5 (verbal memory)		5 (25%)	(5%)[d]	-1.06
			CVLT delayed recall (verbal memory)		8 (42.1%)	(5%)[d]	-0.96
			CVLT delayed recognition total (verbal memory)		4 (21.1%)	(5%)[d]	-0.62
			CVLT delayed recognition minus false positives (verbal memory)		4 (21.1%)	(5%)[d]	-0.66
			Stroop Color-Word Test (executive)	2 SD from HC mean	7 (36.8%)	(2.275%)[d]	-0.61
			Letter fluency (executive)		2 (10%)	(2.275%)[d]	-0.31
			BADS Six Elements (executive)		1 (5.3%)	(2.275%)[d]	-0.31
			CVLT trial 1 (verbal memory)		2 (10%)	(2.275%)[d]	-1.24
			CVLT total recall 1-5 (verbal memory)		4 (20%)	(2.275%)[d]	-1.06
			CVLT delayed recall (verbal memory)		5 (26.3%)	(2.275%)[d]	-0.96
			CVLT delayed recognition total (verbal memory)		4 (21.1%)	(2.275%)[d]	-0.62
			CVLT delayed recognition minus false positives (verbal memory)		4 (21.1%)	(2.275%)[d]	-0.66
Cheung 2013	52	52	CNS-VS neurocognition (overall)	5th percentile of published norm	(46.2%)	(0.0%)	-1.64
			CNS-VS executive function		(53.8%)	(0.0%)	-1.69
			CNS-VS cognitive flexibility		(57.7%)	(0.0%)	-1.66
			CNS-VS complex attention		(51.9%)	(1.9%)	-1.36
			CNS-VS processing speed		(26.9%)	(0.0%)	-1.21
			CNS-VS psychomotor speed		(30.8%)	(1.9%)	-1.15
			CNS-VS reaction time		(44.2%)	(13.5%)	-0.90
			CNS-VS memory composite		(30.8%)	(5.8%)	-0.80
			CNS-VS verbal memory		(28.8%)	(5.8%)	-0.71
			CNS-VS visual memory		(11.5%)	(3.8%)	-0.65
			1 SD from published normative mean on ≥2 index scores		(61.5%)	(1.9%)	NA
			2 SD from published normative mean on ≥2 index scores		(40.4%)	(0.0%)	NA
Fakhry 2013 S1: recent manic episode	30	30	MMSE (global)	Score <25	23 (76.7%)	0 (0%)	-4.62
			MTS (global)	Score <27	18 (60%)	0 (0%)	-2.10
			CDT (executive/visuospatial)	Score <6	0 (0%)	0 (0%)	-3.31
Fakhry 2013 S2: recent depressive episode	30	30	MMSE (global)	Score <25	6 (20%)	0 (0%)	-2.74
			MTS (global)	Score <27	5 (16.7%)	0 (0%)	-0.84
			CDT (executive/visuospatial)	Score <6	0 (0%)	0 (0%)	-2.89

Author Year	Sample *n*[a]		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Ibrahim 2009[e] & Normala 2010[e]	40	40	Category fluency (executive/language)	Score ≤30	3 (7.5%)	0 (0%)[c]	-1.01
			TMT part A (speed/attention)[c]	>40/45/50 seconds[f]	19 (47.5%)	11 (27.5%)	-0.52
			TMT part B (executive)[c]	>90/100/135 seconds[f]	25 (62.5%)	13 (32.5%)	-0.81
			Digit span forward (attention)[c]	Span <5	3 (7.5%)	1 (2.5%)	-0.97
			Digit span backward (working memory)[c]	Span <4	18 (15.0%)	5 (12.5%)	-1.10
			RAVLT trial 1 (verbal memory)	Score <7	31 (77.5%)	13 (32.5%)	NR
			RAVLT trial 5 (verbal memory)	Score <12	23 (57.5%)	1 (2.5%)	NR
			RAVLT trials 1 to 5 (verbal memory)	Score increment <5	16 (40%)	3 (7.5%)	NR
			RAVLT list B (verbal memory)	Score <7	37 (92.5%)	14 (35%)	NR
Juselius 2009	26[g]	114	WCST categories (executive)	1.5 SD from HC mean	12 (50%)	(6.68%)[d]	-0.78
			WCST perseverative (executive)		13 (54%)	(6.68%)[d]	-1.74
			Stroop interference (executive)		15 (68%)	(6.68%)[d]	-3.58
			TMT B minus A (executive)		10 (42%)	(6.68%)[d]	-0.33
			Letter fluency (executive/language)		15 (63%)	(6.68%)[d]	-1.75
			Category fluency (executive/language)		18 (78%)	(6.68%)[d]	-3.40
Osher 2011[c]	51	495	Mindstreams global cognition	1.5 SD from HC mean	25 (49.0%)	(6.68%)[d]	-1.19
			Mindstreams executive function		13 (25.5%)	(6.68%)[d]	-0.83
			Mindstreams attention		20 (39.2%)	(6.68%)[d]	-1.04
			Mindstreams information processing speed		15 (29.4%)	(6.68%)[d]	-0.91
			Mindstreams memory		22 (43.1%)	(6.68%)[d]	-0.96
			Mindstreams verbal function		11 (21.6%)	(6.68%)[d]	-0.51
			Mindstreams visual-spatial		16 (31.4%)	(6.68%)[d]	-0.67
			Mindstreams motor skills		12 (23.5%)	(6.68%)[d]	-0.58

BADS, Behavioural Assessment of the Dysexecutive Syndrome; BD, bipolar disorder; BD-I, bipolar disorder type I; CDT, Clock Drawing Test; CNS-VS, Central Nervous System Vital Signs computerised battery; CVLT, California Verbal Learning Test; HC, healthy comparison; MMSE, Mini-mental State Examination; MTS, Mental Test Score; NA, not applicable; NR, unable to calculate as mean and SD not reported in article; RAVLT, Rey Auditory Verbal Learning Test; SD, standard deviation; TMT, Trailmaking Test; WCST, Wisconsin Card Sorting Test.
d is the standardised mean difference between BD and HC groups, calculated from unadjusted results in the article; negative values indicate worse performance in BD group.
a. Sample characteristics are reported in Supplementary Table H.1 (Appendix H)
b. T-score <32; impairment definition not explicit in article but inferred from bar graph of results.
c. Prevalence data provided by author.
d. By definition, according to impairment threshold applied.
e. Same sample; RAVLT reported in Ibrahim 2009 and other tests reported in Normala 2010.
f. Age groups 18-39, 40-49 and 50-59, respectively.
g. Sample denominator for prevalence results ranges from 22 to 24.

Table 2.3 - Prevalence of cognitive impairment in mixed BD samples

Author Year	Sample *n*[a]		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Barrera 2013[b]	12	12	Reading the Mind in the Eyes test (theory of mind)	Score <21	6 (50%)	2 (16.7%)	-0.61
			Faux Pas Recognition Test cognitive items (theory of mind)	Score <0.75	7 (58.3%)	4 (33.3%)	-0.77
Daban 2012	53	60	WAIS-III Digit Symbol Coding (processing speed)	1.64 SD from HC mean	(30.2%)	(5%)[c]	-0.89
Martino 2014	100	40	Various tests (executive, attention/working memory, verbal memory, naming)	1.5 SD from published normative mean in ≥1 cognitive domain	(70%)	(27.5%)	NA
				2 SD from published normative mean in ≥2 cognitive domains	(30%)	(7.5%)	NA
Mur 2007[b]	44	46	TMT part B (executive)	1.5 SD from published normative mean	0 (0%)	0 (0%)	-0.72
			Letter fluency (executive/language)		6 (13.6%)	0 (0%)	-0.71
			WCST categories (executive)		18 (40.9%)	7 (15.2%)	-0.87
			WCST perseverative (executive)		15 (34.1%)	5 (10.9%)	-0.49
			Stroop inhibition (executive)		11 (25.0%)	1 (2.2%)	-1.30
			Digit span (attention/working memory)		3 (6.8%)	0 (0%)	NR
			TMT part A (speed/attention)		0 (0%)	0 (0%)	-0.28
			CVLT trial 1 (verbal memory)		11 (25.0%)	7 (15.2%)	0.19
			CVLT total words (verbal memory)		17 (38.6%)	6 (13.0%)	0.01
			CVLT immediate recall (verbal memory)		13 (29.5%)	4 (8.7%)	0.12
			CVLT delayed recall (verbal memory)		12 (27.3%)	6 (13.0%)	-0.33
			RCFT immediate (visual memory)		13 (29.5%)	0 (0%)	-0.52
			RCFT delayed (visual memory)		16 (36.4%)	4 (8.7%)	-0.55
Sánchez-Morla 2009	73	67	Executive composite z-score	1.64 SD from HC mean	33 (45.2%)	(5%)[c]	-1.80
			Sustained attention composite z-score		10 (13.7%)	(5%)[c]	-0.65
			Verbal memory composite z-score		21 (28.8%)	(5%)[c]	-1.18
			Visual memory composite z-score		24 (32.9%)	(5%)[c]	-1.10
			WCST % conceptual level response (executive)		(19.2%)	(5%)[c]	-1.02
			WCST % perseverative errors (executive)		(19.2%)	(5%)[c]	-1.01
			Stroop interference (executive)		(35.6%)	(5%)[c]	-0.98
			TMT part B (executive)		(32.9%)	(5%)[c]	-0.97
			Letter fluency (executive/language)		(16.4%)	(5%)[c]	-1.00
			Animal fluency (executive/language)		(24.7%)	(5%)[c]	-0.89
			Tower of Hanoi no. of movements (executive)		(19.0%)	(5%)[c]	-0.64
			Digit span backward (working memory)		(11.0%)	(5%)[c]	-0.53

Author Year	Sample *n*[a]		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
			CPT hits (attention)		(9.6%)	(5%)[c]	-0.52
			CPT sensitivity A (attention)		(9.6%)	(5%)[c]	-0.58
			CPT reaction time (attention/speed)		(23.3%)	(5%)[c]	-0.72
			CVLT total recall 1-5 (verbal memory)		(19.2%)	(5%)[c]	-0.97
			CVLT short free-recall (verbal memory)		(27.4%)	(5%)[c]	-0.96
			CVLT long free-recall (verbal memory)		(15.1%)	(5%)[c]	-0.97
			CVLT short cued-recall (verbal memory)		(20.5%)	(5%)[c]	-1.11
			CVLT long cued-recall (verbal memory)		(23.3%)	(5%)[c]	-0.97
			CVLT recognition discriminability (verbal memory)		(8.2%)	(5%)[c]	-0.67
			CVLT semantic strategies trial A (verbal memory)		(41.1%)	(5%)[c]	-0.82
			RCFT copy (visuospatial)		(16.4%)	(5%)[c]	-0.51
			RCFT short-term (visual memory)		(31.5%)	(5%)[c]	-0.98
			RCFT long-term (visual memory)		(32.9%)	(5%)[c]	-1.01
van der Werf-Eldering 2010	46	75	Various tests (executive, attention/working memory, reaction time, verbal and visual memory)	2 SD from HC mean in ≥1 cognitive domain	6 (13%)	(2.275%)[c]	NA

BD, bipolar disorder; BD-I, bipolar disorder type I; CPT, Continuous Performance Test; CVLT, California Verbal Learning Test; HC, healthy comparison; NA, not applicable; NR, unable to calculate as mean and SD not reported in article; RCFT, Rey Complex Figure Test; SD, standard deviation; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition; WCST, Wisconsin Card Sorting Test.
d is the standardised mean difference between BD and HC groups, calculated from unadjusted results in the article; negative values indicate worse performance in BD group.
a. Sample characteristics are reported in Supplementary Table H.2 (Appendix H).
b. Prevalence data provided by author.
c. By definition, according to impairment threshold applied.

2.3.3.1 Executive function, reasoning and social cognition

Figure 2.2 shows the prevalence of impairment in studies that used a normative distribution-based threshold for impairment. Additional score-based threshold results from three studies (Altshuler et al., 2004; Barrera et al., 2013; Normala et al., 2010) are reported in Tables 2.2 and 2.3. Measures that are likely to be significantly influenced by performance speed are considered separately from those that are not, to minimise the overlap between underlying processing speed ability and instrumental executive function. The former category included timed fluency measures, Stroop test, Trailmaking test, and composite scores primarily influenced by these. The latter category included Tower tests, non-time-dependent aspects of fluency tasks (e.g. category switching accuracy), reasoning tests, Wisconsin Card Sorting test, BADS Six Elements task, and composite scores primarily influenced by these. These timed/non-timed distinctions are not absolute: some tasks in the latter category have time limits imposed during administration, although speed of responding *per se* is not the key determinant of performance success.

Figure 2.2 shows that impairment prevalence tended to be slightly higher on speed-sensitive tasks (lower panel) than on those that depend less on speed (upper panel), though this pattern was not evident in all studies. The estimates did not follow a clear gradient according to the different threshold strata: for example, the estimates from Cavanagh et al. (2002) were the same at the 5^{th} and 2^{nd} percentile levels. This may be a consequence of the small sample size, or may indicate that impaired individuals were strongly clustered at the extreme low end of the score distribution, such that less strict thresholds made little difference to the absolute numbers considered impaired. The estimates from Juselius et al. (2009) were somewhat higher than expected in the context of the other studies. This may be related to study size and quality, but it should also be noted that this study included several twin pairs who were concordant for BD. BD-II-only results are not shown in Figure 2.2, but the supplementary table in Appendix I indicates that fewer BD-II participants were impaired, in comparison with BD-I participants, on most executive function measures in Sparding et al. (2015). Only one study provided prevalence data for social cognition tasks (Barrera et al., 2013): in a small mixed BD sample (n = 12), prevalence of impairment on emotional and cognitive theory of mind measures was higher compared with the healthy comparison sample (Table 2.3).

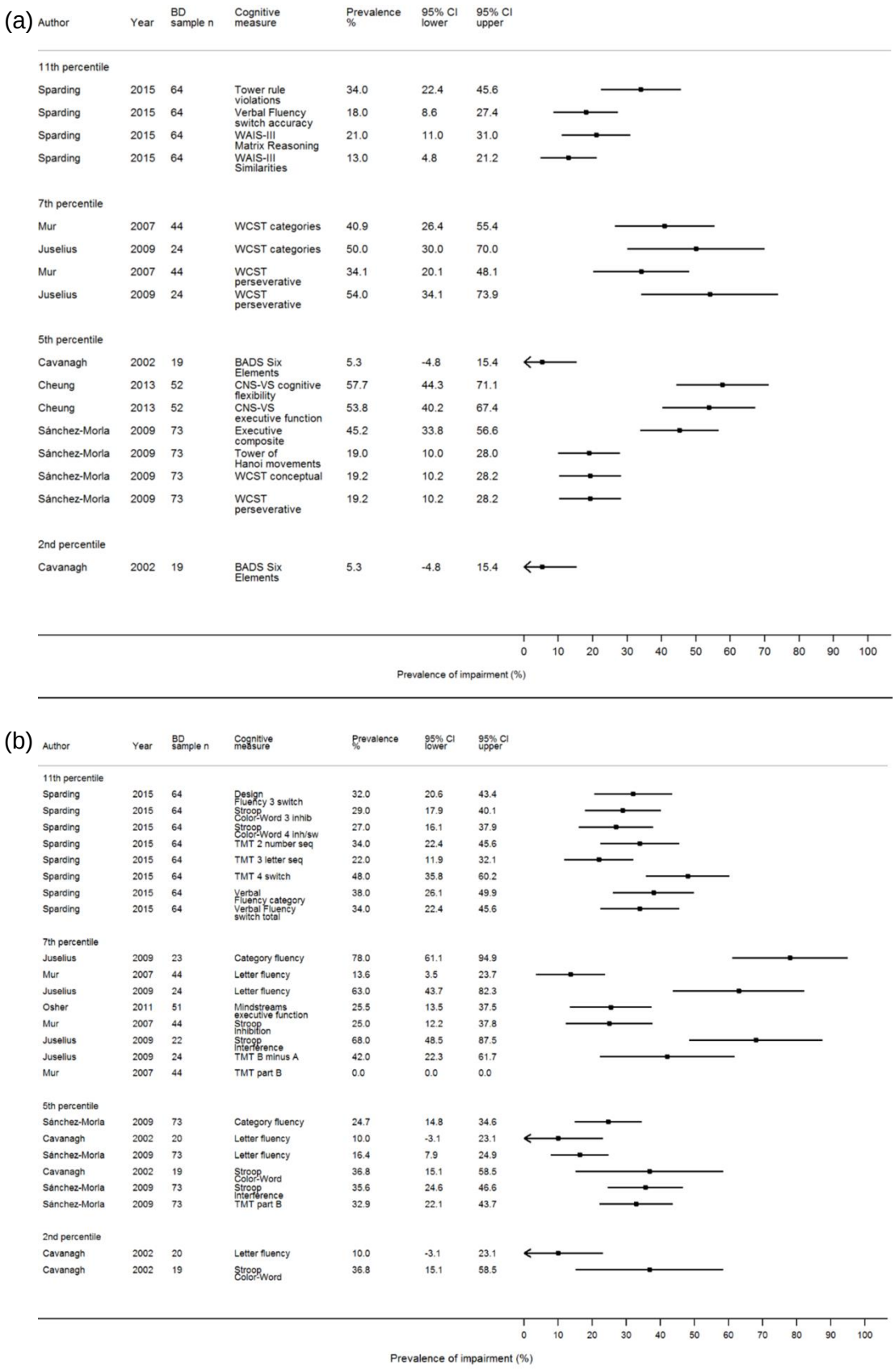


Figure 2.2 - Executive function impairment prevalence across different thresholds

BADS, Behavioural Assessment of the Dysexecutive Syndrome; BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition; WCST, Wisconsin Card Sorting Test. Results include mixed BD and BD-I

samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: √((prevalence*(100-prevalence))/*n*). Results are stratified by impairment threshold (percentile), in descending order from least to most strict. Panel (a) shows executive measures whose scores do not have a prominent timed/speed contribution, and panel (b) shows executive measures whose scores are influenced by speed.

2.3.3.2 Attention and working memory

Figure 2.3 shows the prevalence of impairment in five studies, of similar quality, that reported attention/working memory measures. Estimates were generally higher than in the healthy comparison population, and this was most striking on the CNS-VS complex attention score reported by Cheung et al. (2013) and the Mindstreams attention score from Osher et al. (2011). These scores are composites of several demanding tasks, more akin to the executive measures presented in Figure 2.2. Additional measures from Normala et al. (2010) are reported in Table 2.2, showing a slightly elevated percentage of BD participants with reduced forward and backward digit span. The study by Sparding et al. (2015) allows comparison between BD-I and BD-II samples on two attention/working memory measures, indicating that the proportions with impairment were similar (Appendix I).

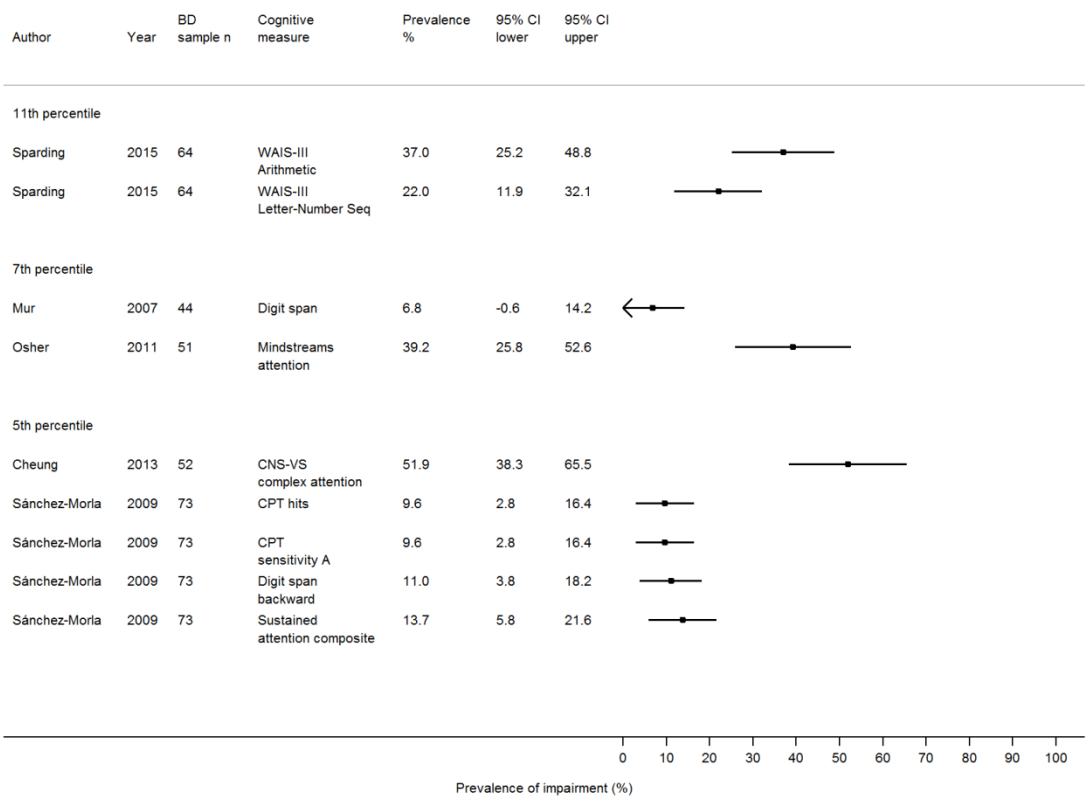


Figure 2.3 - Attention/working memory impairment prevalence across different thresholds

BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CPT, Continuous Performance Test;

WAIS-III, Wechsler Adult Intelligence Scale third edition. Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: √((prevalence*(100-prevalence))/*n*). Results are stratified by impairment threshold (percentile), in descending order from least to most strict.

2.3.3.3 Speed and reaction time

Figure 2.4 shows that the prevalence of impairment on speed and reaction time measures was similar across different impairment thresholds. However, Daban et al. (2012) reported that 30.2% were impaired on the WAIS-III Digit Symbol Coding task at the 5th percentile threshold, whereas Sparding et al. (2015) reported 19% impairment prevalence on the same task at the less strict threshold of 11th percentile. Daban et al. assessed a mixed BD sample but did not report subtypes or illness characteristics, making it difficult to infer reasons for the disparity with Sparding et al.'s BD-I sample. It was also evident from the Sparding et al. study that fewer BD-II participants were impaired on these tasks; in the case of WAIS-III Digit-Symbol Coding, the proportion impaired (11%) was in line with the normative score distribution (supplementary table in Appendix I).

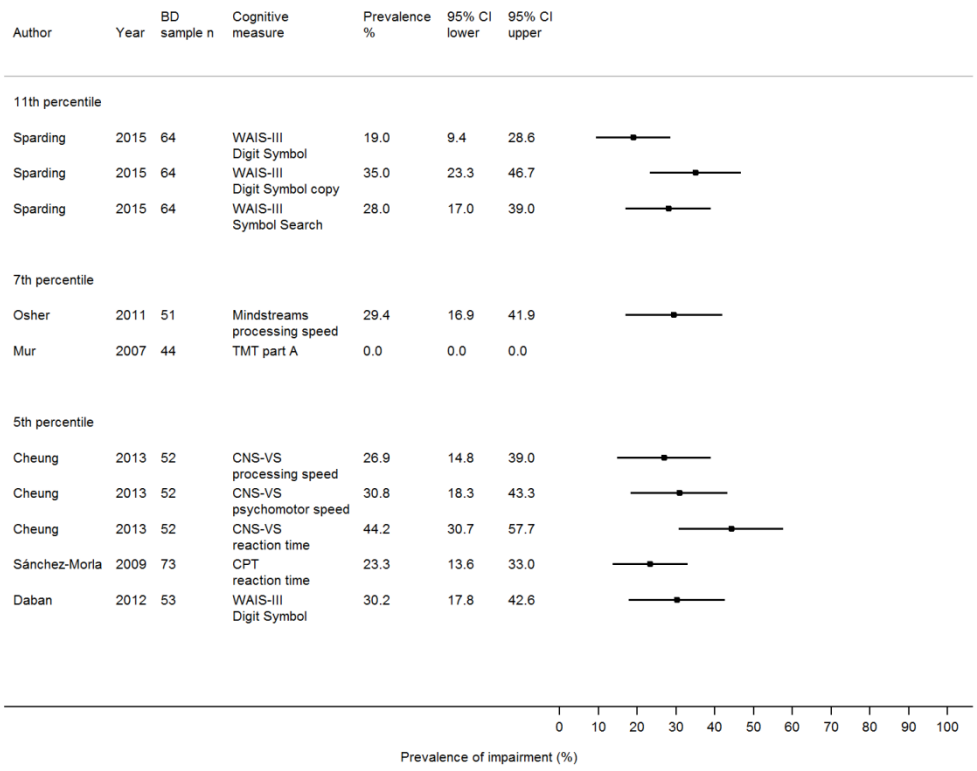


Figure 2.4 - Speed/reaction time impairment prevalence across different thresholds

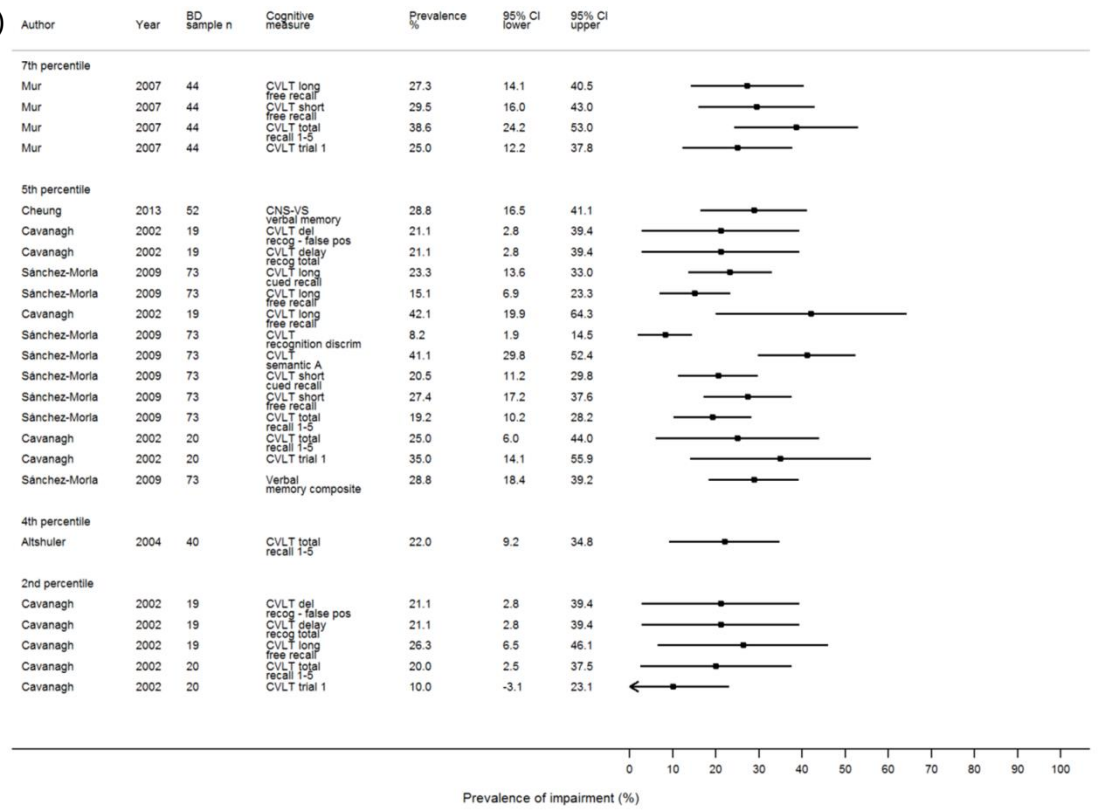
BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CPT, Continuous Performance Test; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition. Results include mixed BD and BD-I samples. Some studies reported results for several cognitive

scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: √((prevalence*(100-prevalence))/*n*). Results are stratified by impairment threshold (percentile), in descending order from least to most strict.

2.3.3.4 Memory

Figure 2.5 shows impairment prevalence results for verbal memory (upper panel) and visual memory (lower panel). Additional score-based threshold results from Ibrahim et al. (2009) are shown in Table 2.2. Two studies of similar quality that reported composite verbal and visual measures separately (Cheung et al., 2013; Sánchez-Morla et al., 2009) showed contradictory findings regarding relative prevalence of impairment: both studies reported that 28.8% were impaired on verbal memory at the 5th percentile threshold, but the proportions impaired on visual memory were 11.5% in Cheung et al. (2013) versus 32.9% in Sánchez-Morla et al. (2009). The proportions impaired on overall memory composite measures were 43.1% at the 7th percentile threshold (Osher et al., 2011) and 30.8% at the 5th percentile (Cheung et al., 2013).

(a)



(b)

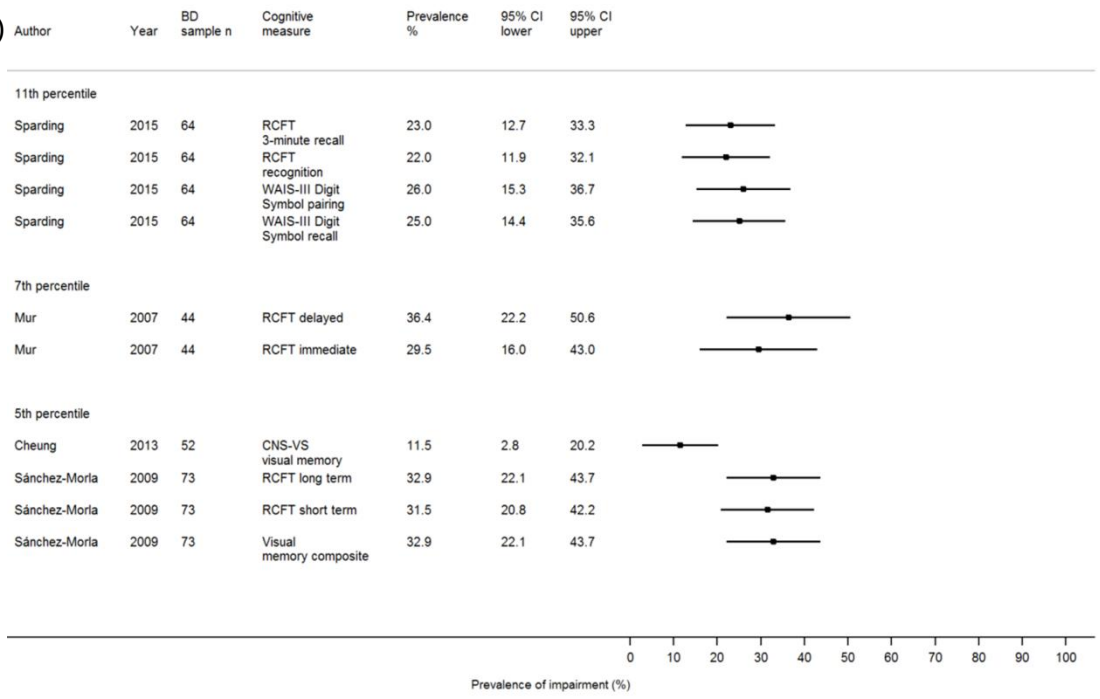


Figure 2.5 - Verbal and visual memory impairment prevalence across different thresholds

BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CVLT, California Verbal Learning Test; RCFT, Rey Complex Figure Test; WAIS-III, Wechsler Adult Intelligence Scale third edition. Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: √((prevalence*(100-prevalence))/*n*). Results are stratified by impairment threshold (percentile), in descending order from least to most strict. Panel (a) shows verbal memory measures and panel (b) shows visual memory measures.

The California Verbal Learning Test (CVLT) was the most common of the verbal memory measures, used in four studies with different thresholds. Results from Cavanagh et al. (2002) and Altshuler et al. (2004) indicated a threshold-related gradient, with fewer participants falling below the stricter 2nd percentile level for CVLT learning and recall, though not for recognition performance. Sánchez-Morla et al. (2009) reported lower impairment prevalence than Cavanagh et al. using the same 5th percentile threshold for the same CVLT measures (total trials 1 to 5, and long delay recall). This may be explained by the larger sample size and mix of BD-I and BD-II participants in the former study. No verbal memory results were available for BD-II separately.

Visual memory results were available from four studies of similar quality. Three (Mur et al., 2007; Sánchez-Morla et al., 2009; Sparding et al., 2015) used the Rey Complex Figure Test (RCFT) at different impairment thresholds; prevalence on this test was lowest in Sparding et al. (2015) despite the less strict threshold and more severe clinical characteristics of their sample. Prevalence of visual memory impairment was similar between BD-I and BD-II samples in that study (Appendix I).

2.3.3.5 Visuospatial function

Three studies (Osher et al., 2011; Sánchez-Morla et al., 2009; Sparding et al., 2015) reported visuospatial measures (Tables 2.2 and 2.3, and Appendix I). Impairment prevalence was lower for visuospatial tasks than for other cognitive domains, though still somewhat higher than would be expected from the normative distribution. Prevalence was highest on the WAIS-III Block Design task—reported as 40% by Sparding et al. (2015) at the 11th percentile threshold—which may reflect the executive and speed components that contribute to success on this task. Prevalence was similarly high among BD-II participants on this task (Sparding et al., 2015).

2.3.3.6 Any domain, multi-domain and global impairment

Fakhry et al. (2013) used the Mini-mental State Examination (MMSE), Mental Test Score (MTS) and Clock Drawing Test (CDT)—typically used as global measures in dementia settings—to assess BD-I participants, grouped by the polarity of their most recent illness episode. Table 2.2 shows that the proportions falling below the impairment cut-off were markedly higher in the group whose most recent episode was manic. No BD participant scored below the cut-off on the CDT, however.

Osher et al. (2011) reported that 49% of their BD-I sample fell below the 7th percentile (1.5 SD) on the global cognition measure of the Mindstreams computerised battery. Also in BD-I, 46.2% of the Cheung et al. (2013) sample were below the 5th percentile on the CNS-VS overall measure of neurocognition. Furthermore, 61.5% were at least 1 SD below the normative mean on at least two CNS-VS index scores, and 40.4% met the stricter criterion of being at least 2 SD below the normative mean on at least two index scores.

Two studies reported overall results from mixed BD samples. Van der Werf-Eldering et al. (2010) found that 6 of 46 participants (13%) were at least 2 SD above the healthy comparison mean (where higher scores indicated worse performance) in at least one cognitive domain. The sample of 46 was a euthymic sub-group from a larger study, for whom demographic and clinical characteristics were not available. It is therefore unclear why the proportion impaired was relatively low in this study. Martino et al. (2014) assessed a larger sample (n = 100), and reported that 70% were impaired using "soft" criteria (1.5 SD below the normative mean in at least one cognitive domain) and 30% were impaired using "hard" criteria (at least 2 SD below the normative mean on at least two domains).

2.3.4 Risk of bias across studies

Figure 2.6 shows funnel plots of the relationship between the prevalence estimates and their precision (standard error), presented separately by cognitive domain, for studies reporting measures at the 5th percentile impairment threshold. Visual inspection suggested a degree of asymmetry for measures of verbal memory, and to a lesser extent for speed-sensitive measures (both within the executive domain and on specific tests of speed/reaction time). Relatively fewer estimates in the lower left quadrant of these plots may indicate publication bias, or reflect other factors such as different sample characteristics or assessment methods in the smaller/less precise studies. The small number of independent measures meant it was not possible to apply statistical tests of asymmetry.

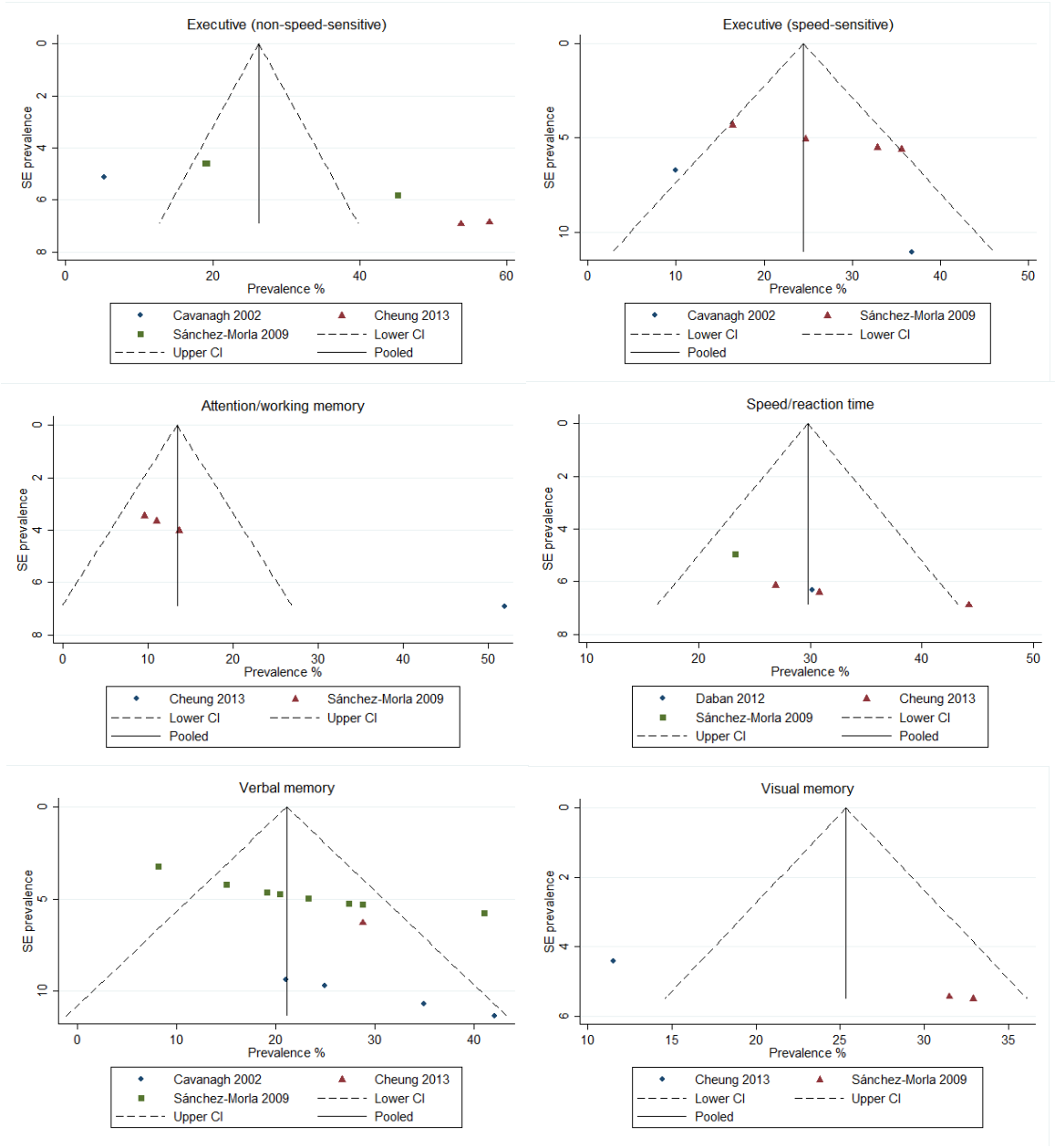


Figure 2.6 - Funnel plots with pseudo 95% confidence limits

SE, standard error. Plots show cognitive impairment prevalence estimates and standard errors by cognitive domain, using results reported at the 5th percentile impairment threshold. Standard errors were calculated as follows: √((prevalence*(100-prevalence))/*n*). Some studies contributed more than one data point to the plot; data points are labelled by study.

2.3.5 Factors associated with cognitive impairment

Twenty-eight articles provided information regarding the relationship between various sociodemographic, clinical or other variables, and the presence or severity of cognitive impairment. Articles were not always clear about which associations had been tested statistically, and they varied in the extent to which they adjusted for other potential confounders. Appendix J shows an overview of the types of variables that were

investigated, with reportedly significant findings highlighted across studies. This synthesis is presented qualitatively rather than quantitatively, owing to the wide variety of ways in which differences or associations were reported across studies.

Associations with demographic variables and premorbid ability were often not investigated. In some cases this was because key background factors had been frequency-matched in a between-group study design, or had been adjusted for when calculating standardised cognitive scores. Other analyses included these background factors as covariates (e.g. in multiple regression), without reporting results for these covariates separately. For the remainder, greater cognitive impairment was associated with older age, lower education and lower premorbid ability in some studies, but others reported no statistically significant findings.

Illness characteristics—such as duration since onset, number of affective episodes and hospitalisations, history of psychotic symptoms, and residual depressive or manic symptoms—were more frequently investigated. Where statistically significant results were reported, they indicated that more severe illness characteristics were associated with worse cognitive function. An exception was history of psychotic symptoms, for which one study reported both positive and negative associations. Several studies investigated associations with psychotropic medication, with mixed findings. The most frequently reported association was between antipsychotic medication and worse cognition, although some studies reported null findings. By contrast, mood stabilisers (lithium or anticonvulsants) were less frequently associated with impairment.

Although two studies examined history of alcohol/substance use disorder, none investigated the relationship between frequency/amount of alcohol or recreational drug consumption and cognitive impairment. No study examined associations with smoking or other cardiovascular risk factors that may be relevant to cognitive impairment.

2.4 Discussion

2.4.1 Summary of findings

The aims of the review presented in this chapter were to determine the prevalence of cognitive impairment in euthymic adults with BD, and to ascertain which clinical, sociodemographic or other factors were associated with cognitive impairment in this population. Thirty articles contributed to the findings, of which 15 provided prevalence

data. Impairment prevalence was similar between BD-I only and mixed BD samples. One study with separate results for BD-I and BD-II participants indicated that impairment was more common in those with BD-I, though considerable overlap was apparent. Examination of impairment proportions across different cognitive domains indicated wide variation both within and between domains. For example, taking the 5th percentile threshold as the reference, impairment prevalence ranges were as follows: non-speed-sensitive executive function 5.3% to 57.7%; speed-sensitive executive function 10.0% to 36.8%; attention/working memory 9.6% to 51.9%; speed/reaction time 23.3% to 44.2%; verbal memory 8.2% to 42.1%; visual memory 11.5% to 32.9%. Generally small sample sizes resulted in wide CIs for most estimates. A recent review of neuropsychological function in BD (Szmulewicz et al., 2015) highlighted impairment prevalence as an issue of particular interest, and reported estimates between 30% and 57% from six studies. Four of these studies were not eligible for the present review, either because participants were not euthymic or because the recruitment method did not meet our criteria. The fact that the lower bounds of the prevalence estimates reported in the present review are below the previous estimate of 30% can be understood in light of our exclusion of non-euthymic participants and samples recruited by convenience, either of which may bias prevalence estimates upwards.

There was some evidence that more severe or longstanding illness was associated with greater cognitive impairment. Several studies reported an association with antipsychotic medication but less so with other types of psychotropic medication; it should be noted, however, that medication associations are likely to be confounded by illness severity as well as treatment adherence and responsiveness. A previous individual participant data meta-analysis of 2,876 euthymic patients with BD (Bourne et al., 2013) also reported statistically significant associations between cognitive performance and some illness severity indices (e.g. number of manic episodes and total hospitalisations), and reported an association on a verbal memory test for antipsychotic medication only, but not lithium, antidepressants or anticonvulsants.

2.4.2 Limitations of included studies

Valid prevalence estimates depend on representative samples, but representativeness was questionable in many of the studies included here. Although all appear to have used an appropriate recruitment method (e.g. consecutive or random sampling), details were scant in published papers regarding exact recruitment processes and numbers considered at each

stage. Exclusion on the basis of comorbidity such as substance misuse was common, but numbers excluded were rarely reported. Definitions of euthymia varied; even when these were based on common measures (e.g. HRSD and YMRS), cut-off scores differed across studies. A wide range of cognitive tests was used, and even within specific tests, many different scores were reported (e.g. various CVLT sub-scores). This made direct comparison across studies difficult. The use of different thresholds to define cognitive impairment also limited synthesis at the outcome level. Most studies focused on the cognitive domains of executive function, memory and attention, with other areas of function such as visuospatial ability and language studied rarely if at all. Articles were sometimes unclear regarding which demographic, clinical or other variables were statistically analysed in conjunction with the cognitive measures.

2.4.3 Limitations of review

Recommended practices in systematic review methodology were followed, but reproducibility of screening, data extraction and bias appraisal processes was checked for only a proportion of records. Judgements about study eligibility relied solely on information contained in the articles, and authors were not contacted to request missing information during the selection process. A large number of articles were excluded on the sample recruitment criterion, in some cases because this information was not contained in the article; it is possible that some of these did in fact employ an appropriate sampling procedure. The requirement for information within the article indicating an acceptable recruitment procedure meant that several articles included in previous reviews of cognitive function in BD are not included here, including some that reported prevalence estimates. Despite repeated attempts to obtain additional prevalence results from authors of eligible articles, prevalence data were available for only 15 articles. In particular, there was little information regarding impairment prevalence in BD-II samples. Heterogeneity of cognitive measures and thresholds meant that it was not feasible to meta-analyse the prevalence estimates obtained, or to conduct statistical tests of funnel plot asymmetry, and so the results are limited to graphical and narrative synthesis only. This was organised by cognitive domain, though it is acknowledged that many tests make multiple cognitive demands across domains. Regarding the second review question, variation in the way that correlates were analysed across studies meant it was not possible to comment on the nature of any inter-relationships between the potential risk factors reported here. Risk of bias was considered carefully, but it should be noted that the appraisal tool used was developed for questions of prevalence, whereas many of the studies included here were not originally

designed to investigate prevalence. The literature search results were restricted to English-language publications only, although studies from a wide range of international settings were found.

2.4.4 Conclusions and implications

The review presented in this chapter is the first to systematically examine the prevalence of cognitive impairment in euthymic bipolar disorder. It complements and extends the findings of previous reviews, which have focused on the magnitude of between-group differences on cognitive measures. Although group differences are important for understanding the nature and extent of cognitive impairment in this population, quantifying the number who have clinically relevant cognitive impairment is essential if we wish to identify risk factors for a cognitively impaired subtype of euthymic BD, and to target clinical resources towards neuropsychological rehabilitation and support for those who need it most. Despite the heterogeneity demonstrated in this chapter, some tentative conclusions can be drawn. Cognitive impairment affects patients across the BD spectrum; impairment appears to be more common in BD-I but there is considerable overlap with BD-II. It is also evident that even at the lower ends of the prevalence ranges reported here, the proportion of patients whose affective illness is in remission but who continue to show cognitive impairment substantially exceeds the expected proportion in the general population. With BD diagnosis typically being made in early adulthood, this means that the excess burden of cognitive impairment will affect this population over several decades. This ‘cognitive footprint’ effect (Rossor & Knapp, 2015) is considered further in Chapter 5.

2.5 Next steps

The review presented here has highlighted the challenges of synthesising prevalence estimates from studies with heterogeneous methods. By using the UK Biobank data resource, there is potential to estimate prevalence reliably in a large community-based BD sample who have been assessed in a standardised way, and to directly compare estimates between BD and other clinical groups within the cohort. Chapter 3 describes the UK Biobank resource in detail, and Chapter 4 describes how self-reported and linked health records data were used to identify groups within the cohort with a history of the conditions of interest. Chapter 5 reports the results of the prevalence analyses across these groups.

Chapter 3 Description of the UK Biobank resource

This chapter describes the recruitment, cohort composition and relevant data available in the UK Biobank resource. Detailed information about the resource is publicly accessible at http://www.ukbiobank.ac.uk/.

UK Biobank is a general population-based prospective cohort study, established with the aim of elucidating the genetic and non-genetic determinants of diseases of middle and old age (Sudlow et al., 2015). It has received funding from the Wellcome Trust, Medical Research Council, UK Government Department of Health, Scottish Government, Northwest Regional Development Agency, Welsh Assembly Government, British Heart Foundation and Diabetes UK. Data are made available, upon application, to bona fide researchers for health-related research that is in the public interest. The studies reported in this thesis were conducted under approved application 11332 (Appendix K). UK Biobank has approval from the National Health Service (NHS) National Research Ethics Service as a research tissue bank (reference 16/NW/0274 and 11/NW/0382; see Appendix L), and separate project-specific ethical approval is not required by researchers using data released by UK Biobank. For administrative completeness, however, the proposal for this thesis was approved by the College of Medical, Veterinary and Life Sciences Ethics Committee at the University of Glasgow (reference 200150023; Appendix M), and was acknowledged by the NHS Greater Glasgow & Clyde Research and Development Department (reference GN14NE132; Appendix N).

3.1 Cohort composition

3.1.1 Population and recruitment

UK Biobank aimed to recruit 500,000 participants, in order to achieve adequate statistical power to reliably detect odds ratios of 1.3 to 1.5 for the main effects of different exposures on risk of disease over a follow-up period of 10-20 years (Sudlow et al., 2015). Adults aged 40 to 69 years who were registered with the NHS and living within 25 miles of a study assessment centre were invited by post to participate in UK Biobank. No exclusion criteria were applied. The age range was chosen so that participants would be expected to experience incident disease outcomes in the early years of follow-up, while allowing baseline assessment of exposures with minimal influence of incipient disease (Sudlow et

al., 2015). Twenty-two assessment centres were in operation across England, Scotland and Wales at different times between 2006 and 2010. Figure 3.1 shows the invitation and appointment process, which was managed centrally using national NHS records. Invitation mailings were stratified according to age, gender and postcode area (as a measure of social deprivation). Approximately 9 million invitations were issued to achieve the eventual cohort size of ~502,000, indicating an overall response rate of around 5.6% (Manolio et al., 2012).

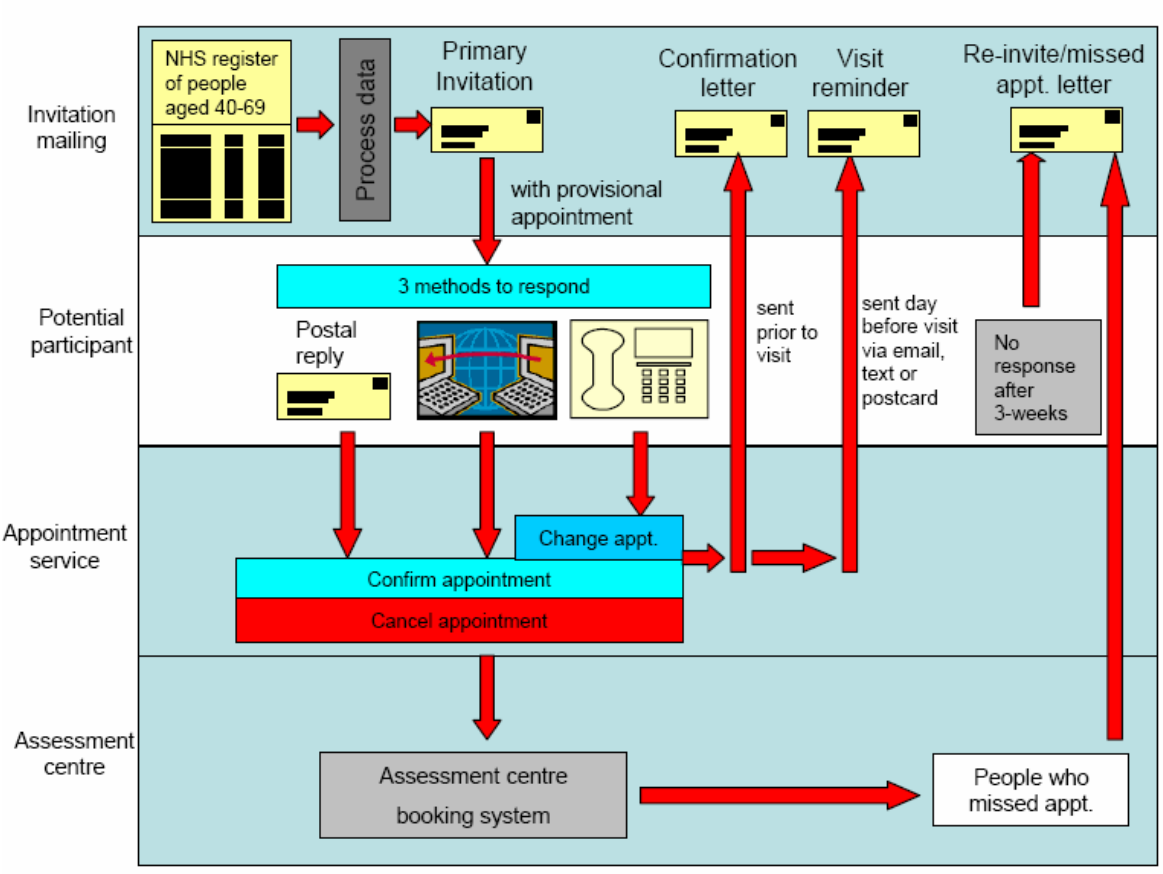


Figure 3.1 - UK Biobank invitation and appointment process

A schematic representation of the invitation and appointment process is shown. Image taken from the UK Biobank protocol UKBB-PROT-09-06, dated 21 March 2007.

3.1.2 Cohort characteristics

Available data as at 3rd November 2016 (n = 502,639)[2] indicated that participant ages ranged from 37 to 73 years, with the majority (n = 500,205; 99.5%) being between 40 and 69 years as per the recruitment strategy (M = 56.5; SD = 8.1). More women than men took

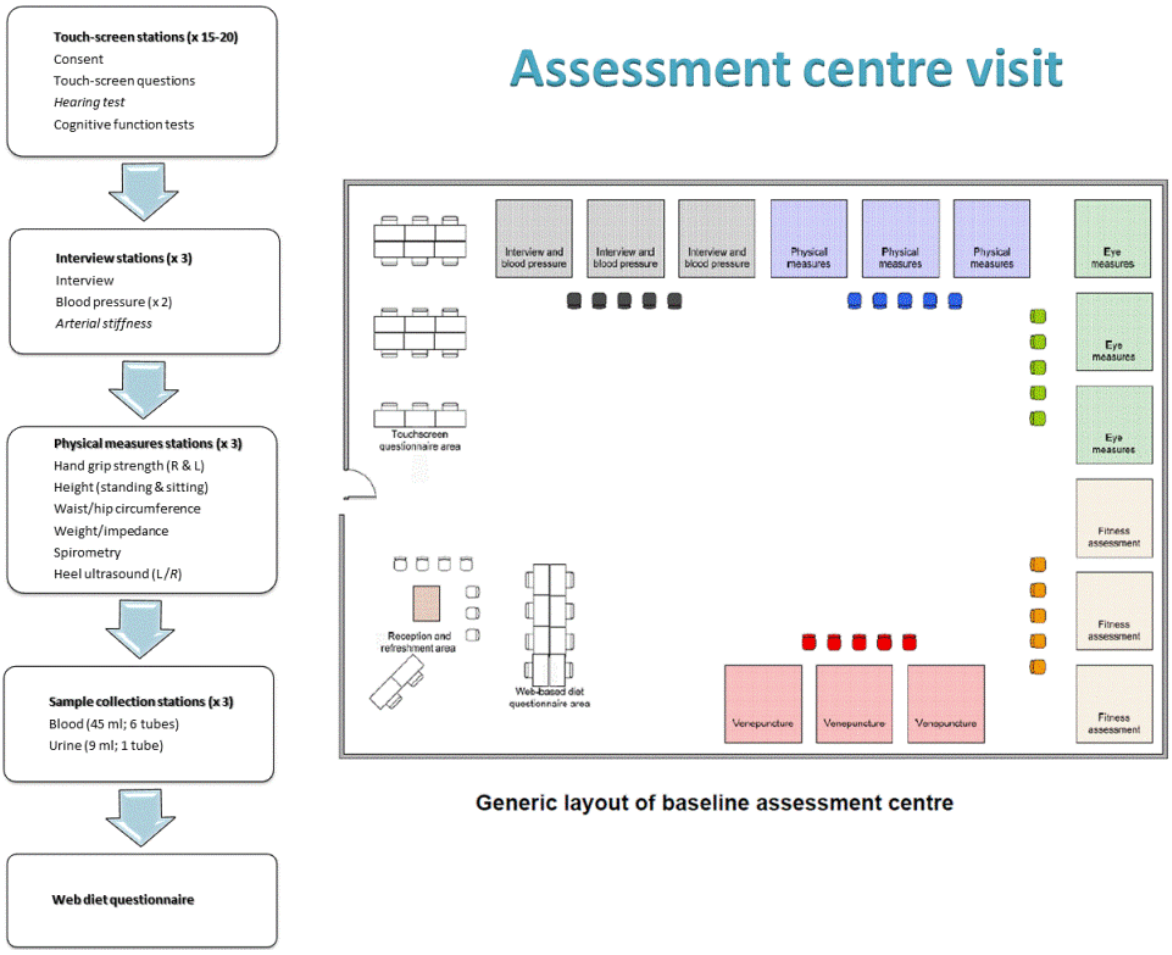
[2] The number with available baseline data changes slightly over time due to participant withdrawals.

part (*n* = 273,465; 54.4%). The majority self-reported white ethnicity (white British *n* = 442,698, 88.1%; white Irish *n* = 13,213, 2.6%; white other *n* = 16,911, 3.4%), followed by Asian/Asian British (*n* = 9,882; 2.0%), black/black British (*n* = 8,065; 1.6%) and mixed ethnic background (*n* = 2,958; 0.6%). Most were in paid employment or were self-employed (*n* = 287,231; 57.1%), with 177,485 (35.3%) retired and 20,386 (4.1%) unable to work due to sickness or disability. Almost one-third reported having a university or college degree (*n* = 161,208; 32.1%).

A recent comparison of the UK Biobank cohort with UK Biobank invitees who did not participate, and with findings from nationally representative surveys, confirmed that the cohort is not representative with regard to gender and deprivation; the proportion reporting white ethnicity is representative of the 2001 UK census but is higher than that reported in the 2011 census (Fry et al., 2017). Another study (Hill, Hagenaars, et al., 2016), focusing on the subset of the cohort that was included in the first release of genomic data, compared the distribution of deprivation scores with the distribution in England and Wales in the 2001 census; this showed that the most deprived areas were not represented in the UK Biobank sample, but the shape of the distribution was otherwise similar. It has been argued that measures of association within the study population may still be valid (Collins, 2012), but caution must be exercised when non-representativeness is due to self-selection, because spurious associations may arise when the characteristics under study have themselves influenced selection into the cohort (Ebrahim & Smith, 2013; Swanson, 2012). This issue, known as collider stratification bias (Greenland, Pearl, & Robins, 1999; Munafò, Tilling, Taylor, Evans, & Smith, 2017), is considered in more detail in later chapters.

3.2 Assessment procedures

Participants attended baseline assessment visits lasting about two hours at UK Biobank assessment centres between 2006 and 2010. Written consent was provided, following which participants completed computerised touchscreen questionnaires and cognitive assessments, then underwent a verbal interview and physical measures with a trained staff member. They also provided urine and blood samples for biomarker and genomic analysis. The assessment process was standardised, as depicted in Figure 3.2.



Generic layout of baseline assessment centre

Figure 3.2 - Layout of UK Biobank assessment centre

Each assessment centre was laid out approximately as shown, allowing standardised participant flow between stations. Image taken from http://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Clinic_layout

Approximately 20,000 participants attended a repeat visit in 2012-2013, at which the same assessments were re-administered. Imaging visits commenced in 2014, with the aim of collecting MRI scans of brain, heart and abdomen from up to 100,000 participants. Additional web-based data collection has been undertaken since 2013, including questionnaires about diet, work environment and lifetime mental health experiences. Additional cognitive data have also been collected via web-based assessment, and physical activity has been measured using wrist-worn accelerometers supplied and returned by post. Data linkage has been established between UK Biobank and NHS medical records, cancer registers and death registers, providing information about medical history pre-dating recruitment to UK Biobank as well as incident outcomes and mortality on an ongoing basis.

The studies reported in this thesis primarily used data from the baseline assessment visit, together with linked NHS records and web-based mental health data. Relevant measures and data sources are described below.

3.3 Sociodemographic data

Age was calculated from assessment date minus date of birth, and truncated to whole years. Gender, ethnic group and country of origin were self-reported via the computerised touchscreen interface. The Townsend deprivation index (Townsend, 1987) was calculated for the participant's postcode of residence immediately prior to the baseline assessment date. This score combines data from census output areas (approximately 125 households per area in England/Wales and 50 households per area in Scotland) regarding unemployment, car ownership, home ownership and household overcrowding, to produce a standardised score, where higher values indicate greater relative deprivation. Information about occupation, income, household factors and car ownership was also self-reported in the touchscreen questionnaire. Participants were asked to report which educational qualifications they held (if any), from a list encompassing college/university degree, higher and intermediate secondary school qualifications, and vocational qualifications.

3.4 Lifestyle and physical measures

The touchscreen questionnaire recorded information on current and past smoking habits, from which variables were derived by UK Biobank to code for current, former and never smokers. Current frequency of alcohol consumption was recorded using categories from 'never' to 'daily/almost daily', and those responding 'never' were asked about previous consumption. Touchscreen questions also recorded self-reported estimates of typical dietary intake and physical activity levels, as well as typical sleep patterns. Physical measures were performed by trained staff, and included blood pressure, height, waist and hip circumference, weight, and body mass index (BMI).

3.5 Medical history and mental wellbeing

Information regarding medical and psychiatric history and medications was primarily self-reported, via touchscreen questions and an interview conducted by a trained staff member at baseline. Additional self-reported data regarding lifetime mental health-related experiences were collected several years later via an online questionnaire. Hospital

admissions data were available for most participants from linked NHS records. These sources of information are described below.

3.5.1 Self-reported information

3.5.1.1 Self-reported diagnoses, medications and family history

Touchscreen questions asked about diagnoses within certain pre-specified categories, e.g. cardiovascular disease. All participants were later asked during the interview whether they had ever been told by a doctor that they had any serious illness or disability, and they were asked to name all current regular prescribed medications (doses and formulations were not recorded). Responses were recorded by the interviewer and were subsequently assigned unique codes. Diagnoses and medications of interest for the studies in this thesis were manually coded, as described in the relevant chapters. The touchscreen questionnaire also asked about history of specific illnesses in parents and siblings, including cancers, cardiovascular disease, dementia, Parkinson's disease and severe depression.

3.5.1.2 Baseline mental health questionnaire

Additional self-reported information about mental health was elicited as part of the touchscreen questionnaire. Some of the questions were added in the final two years of the baseline phase, and so data are not available for all participants. Neuroticism was assessed using 12 yes/no items from the Eysenck Personality Questionnaire Revised short form (Eysenck, Eysenck, & Barrett, 1985), summed to produce a total score from 0 to 12, where higher scores indicate greater neuroticism. Four questions were administered regarding frequency of depressive symptoms in the past two weeks: depressed mood or hopelessness; lack of interest or pleasure; tenseness or restlessness; and tiredness or low energy. These were based on items from the Patient Health Questionnaire (PHQ; Kroenke, Spitzer, & Williams, 2001). Participants self-rated each symptom on a four-point scale from 'not at all' to 'nearly every day', summed to produce an overall score ranging from 0 to 12, with higher scores indicating more frequent depressive symptoms.

Past experiences of depressive and manic symptoms, and medical help-seeking for mental health, were assessed using touchscreen questions agreed by an expert group and informed by the Structured Clinical Interview for DSM-IV Axis I disorders (First, Spitzer, Gibbon, & Williams, 2002). Responses to these items have been used to classify UK Biobank

participants according to probable lifetime features of mania/BD and major depression (D. J. Smith et al., 2013).

3.5.1.3 Web-based mental health questionnaire

A web-based mental health questionnaire was disseminated to all participants in 2016. This included a range of items taken from standardised instruments, to assess lifetime and current experiences relating to depression, mania, anxiety, psychotic-like phenomena, self-harm, post-traumatic stress and substance use. The questionnaire and the results from the first data release (n = 157,366) are described in detail in Davis et al. (2018). This questionnaire also asked about adverse or traumatic experiences in adulthood and in childhood. The childhood items were taken from the brief Childhood Trauma Questionnaire (Bernstein et al., 2003). Participants were asked how frequently certain statements, reflecting abuse or neglect, applied to them while they were growing up ('never true' to 'very often true'): that they felt loved; were hit resulting in bruises or marks; felt hated within their family; were molested sexually; and had someone to take them to the doctor if needed.

3.5.2 Linked NHS records

NHS records were obtained by UK Biobank from the Health and Social Care Information Centre (HSCIC) in England, the Information and Statistics Division (ISD) of NHS Scotland and the Secure Anonymised Information Linkage (SAIL) system in Wales, and were linked centrally to UK Biobank data using participant identifiers. A mixture of exact and probabilistic matching was used (http://biobank.ctsu.ox.ac.uk/~bbdatan/matching_algorithms_documentation_v1.pdf). Data were available from in-patient and day case admissions to NHS hospitals from the mid-1990s onwards (dates varied by country). Records from England (Hospital Episode Statistics; HES) and Wales (Patient Episode Database for Wales; PEDW) included both acute and psychiatric hospitals. Records from Scotland were from the Scottish Morbidity Record - General Acute Inpatient and Day Case dataset (SMR01), which does not cover psychiatric hospitals.

3.5.3 Adjudicated diagnoses

UK Biobank analysts developed algorithms to classify likely history of certain medical conditions, using data from the baseline assessments (self-reported medical conditions,

operations and medications), linked hospital records (diagnoses and procedures) and death registers. These were developed by the UK Biobank Outcome Adjudication Group, with the aim of maximising the positive predictive value of the classification. To date, algorithmic classifications are available for myocardial infarction and stroke. Details of the algorithms can be found at http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=42.

3.6 Cognitive assessment

A short cognitive assessment battery was administered during assessment centre visits, as described below.

3.6.1 Materials and procedure

Brief cognitive tests were administered via touchscreen, during the first stage of the assessment centre visit. The tests were designed specifically for UK Biobank but share some characteristics with other established tests of cognitive function. Two tests were included in the protocol throughout the UK Biobank baseline phase (reaction time and visuospatial memory), two tests were introduced in the final two years of recruitment (reasoning and prospective memory), and one test was introduced in the final two years and then subsequently removed for reasons of time (numeric memory). Sample size therefore varies across tests. The total time to complete all five tests was approximately 15 minutes.

3.6.1.1 Reasoning test

Thirteen questions were presented sequentially via touchscreen on a self-paced basis with an overall time limit of two minutes. Responses were selected from a multiple-choice array. Any questions not attempted during the two-minute time limit were scored as zero. The items included both verbal and numerical reasoning tasks. An example of a verbal item is “Bud is to flower as child is to? (Select from: Grow / Develop / Improve / Adult / Old / Do not know / Prefer not to answer)”, and an example of a numerical item is “150…137…125…114…104… What comes next? (Select from: 96 / 95 / 94 / 93 / 92 / Do not know / Prefer not to answer)”. The score for analysis was an unweighted total ranging from 0 to 13, with higher scores indicating better performance.

3.6.1.2 Reaction time test

Participants were asked to press a button with their dominant hand as quickly as possible each time a matching pair of symbols was presented on the computer screen. This type of 'Go/No-Go' reaction time paradigm was described by Donders (1868/1969), cited in Gomez, Ratcliff, and Perea (2007). Five practice trials were administered, followed by seven test trials. An example of a non-matching trial is shown in Figure 3.3. The score for analysis was the mean time (in milliseconds) taken to press the button, derived from the four trials on which a matching pair occurred. Higher scores indicate slower (i.e. worse) performance.

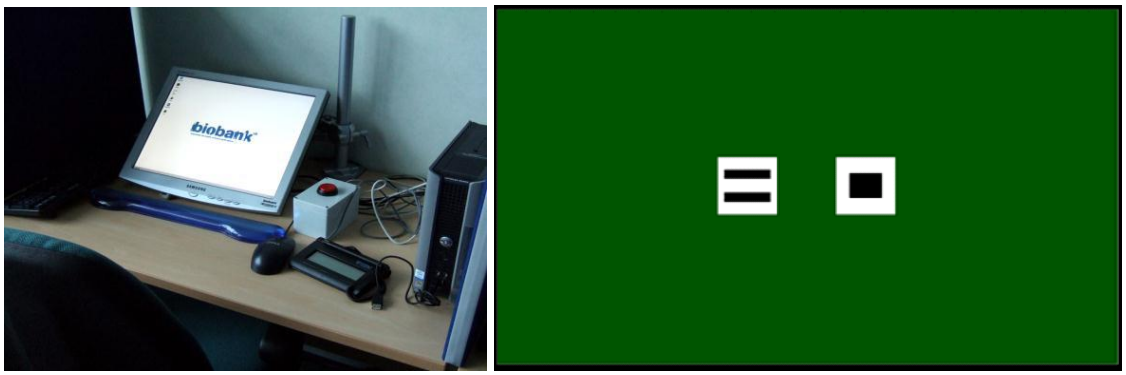


Figure 3.3 - Reaction time test in UK Biobank

The touchscreen and response button are shown in the left panel. The right panel shows a screenshot of a non-matching trial. Images taken from the UK Biobank manual v1.1, dated 19th March 2013.

3.6.1.3 Numeric memory test

A string of numbers was presented on-screen and, after a brief delay, participants were asked to enter it from memory, in reverse order, via a touchscreen numeric keypad (see Figure 3.4). Each string was presented on the screen for 2000ms, plus an additional 500ms multiplied by the string length. A delay of 3000ms occurred between clearing the screen and activating the response keypad. All participants began with a string length of two, and successive strings increased by one, up to a maximum string length of 12. The test was discontinued after five successive incorrect responses at a string length of two, or after two successive incorrect responses at string lengths of three or more. The score for analysis was the maximum string length recalled correctly, with higher scores indicating better performance.

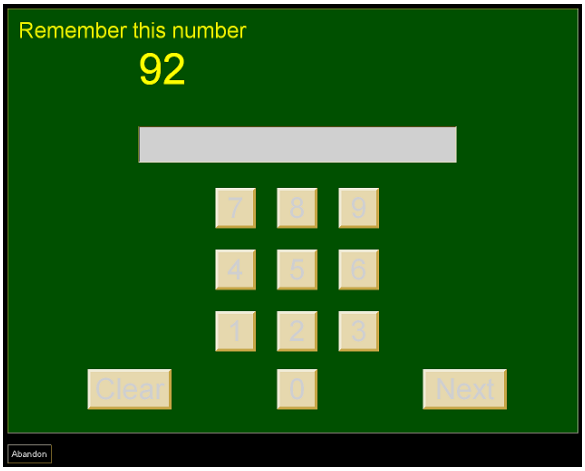


Figure 3.4 - Numeric memory test in UK Biobank

Screenshot of the first trial of the numeric memory test. The touchscreen keypad activated after the number was cleared from the screen. Image taken from http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=10

3.6.1.4 Visuospatial memory test

In the 'pairs matching' test, symbol cards were presented on-screen in a random array. Participants were asked to memorise the position of as many matching pairs as possible. The cards were then turned face down on the screen and participants were asked to touch as many matching pairs as possible in the fewest tries (see Figure 3.5). Two trials of this task were administered, one with three pairs of symbols and one with six pairs; the presentation times for memorisation were 3000ms and 5000ms seconds respectively. No time limit was applied during the recall phase. The score for analysis was the number of errors made while attempting to select the pairs, with a higher score indicating worse performance.

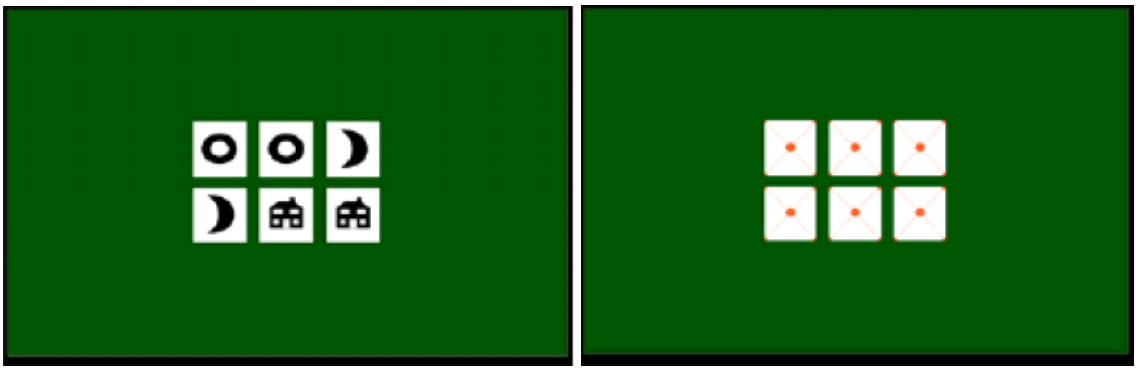


Figure 3.5 - Visuospatial memory ('pairs matching') test in UK Biobank

The left panel shows a three-pair array during the memorisation phase. The right panel shows the cards face down, ready for the participant to respond. Images taken from the UK Biobank manual v1.2, dated 17th December 2013.

3.6.1.5 Prospective memory test

The following instruction appeared on the touchscreen: “At the end of the games we will show you four coloured symbols and ask you to touch the blue square. However, to test your memory, we want you to actually touch the orange circle instead”. After a delay during which participants completed the other cognitive tasks described above, a screen appeared showing four coloured shapes with the instruction to touch the blue square (Figure 3.6). If the participant touched the orange circle, their response was recorded as ‘correct on first attempt’. If they touched the blue square, they were given a prompt on-screen to try to recall what the original instruction was, and were asked to respond again. If they correctly selected the orange circle after receiving this prompt, their response was recorded as ‘correct on second attempt’. All other responses (including no response) were recorded on the system as incorrect.

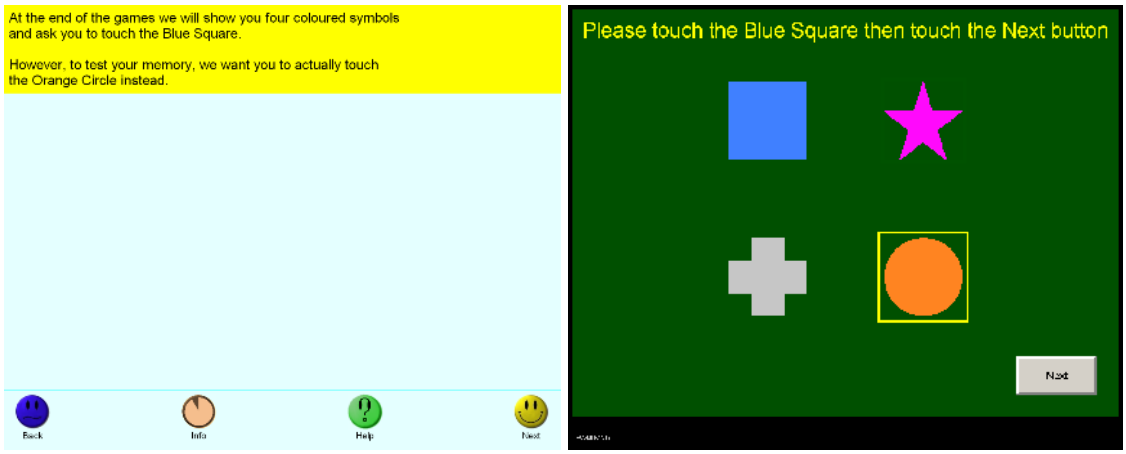


Figure 3.6 - Prospective memory test in UK Biobank

The left panel shows the instruction screen, and the right panel shows the response screen presented after a filled delay. Images taken from UK Biobank manual v1.1, dated 28th February 2012.

3.6.2 Psychometric characteristics

The cognitive tests were developed specifically by UK Biobank to be brief and feasible for self-administration at scale using a touchscreen. Pilot testing for feasibility purposes was conducted before the main baseline phase commenced, which resulted in some original tasks being removed and others being shortened. UK Biobank did not publish information about the psychometric properties of the pilot or main phase tests, but the data that are

available to researchers from the pilot, baseline and repeat assessment visits enable some limited examination of score stability and covariance structure.

3.6.2.1 Validity

Although the UK Biobank cognitive tests share some face validity with existing tests of cognitive function, there are some points of difference worth noting. The reasoning test is referred to in all UK Biobank documentation as a 'fluid intelligence' test, but performance does not appear to decline with age as would be expected on a fluid ability task, and it is likely that successful performance on some of the items would be influenced by crystallised abilities (Hagenaars et al., 2016). For this reason, it is referred to simply as a reasoning test throughout this thesis. In the numeric memory test, the digit strings were presented on-screen in their entirety, rather than one digit at a time. This means that the apparent demand of recalling the string in reverse order will have been lessened if participants chose to read the numbers in reverse when they were first presented, thus turning the task into a forward rather than reverse span test. The 'pairs matching' test of visuospatial memory differed from existing computerised visual paired associate learning tests (such as the CANTAB Paired Associate Learning test; www.cambridgecognition.com) by presenting the full array of symbol pairs on screen simultaneously, rather than a series of individual symbols whose spatial location must be memorised and recalled. The prospective memory test required memorisation and execution of an instruction after a filled delay, in common with other tests of this type, but it also required the participant to inhibit a competing response (to touch the blue square). These idiosyncrasies in the bespoke UK Biobank cognitive tests must be considered when interpreting patterns of performance observed in the data.

The validity of the tests can also be considered in terms of their covariance structure, which indicates to what extent they are measuring shared or unique underlying abilities. Lyall and colleagues (2016) reported a principal components analysis of the reasoning, reaction time, numeric memory and visuospatial memory scores, whose first unrotated principal component accounted for 39.9% of the variance. When only reasoning, reaction time and visuospatial memory were included, the first component accounted for 44.4% of the variance. All individual tests had moderate to high loadings on the first unrotated principal component in each model (loadings ranged from |0.49| to |0.77| in the four-test model). The overall results indicated a one-factor solution, but the magnitude of variance explained was lower than that reported in other cohorts with more detailed cognitive

assessments, e.g. approximately 50% in the Lothian Birth Cohort-1936 (Lyall et al., 2013). It would therefore be informative to analyse the UK Biobank cognitive tasks separately as well as, or instead of, constructing a composite score or using the observed data to estimate a latent general factor of cognitive function.

3.6.2.2 Reliability

Lyall and colleagues (2016) reported that test-retest reliability between the UK Biobank baseline and the first repeat visit (mean interval = 4.33 years, SD = 0.93), excluding participants with diseases affecting brain function at baseline, varied across tests. Intraclass correlation (ICC) coefficients were 0.65 for reasoning, 0.57 for reaction time and 0.16 for visuospatial memory (numeric memory was not administered at the repeat visit), and Cohen's kappa for prospective memory was 0.36, with 97.9% of participants scoring correctly at both time points. Additional data were available from the pilot phase of UK Biobank for the visuospatial memory test, when participants were administered a six-pair trial twice in immediate succession: the ICC was again low at 0.17.

Internal consistency can be examined for tests containing multiple items or multiple trials. In UK Biobank, only the reasoning and reaction time tests could be examined in this way. Hagenaars and colleagues (2016) reported Cronbach's alpha coefficients in a subsample of the cohort who were included in a genetic analysis. Alpha was 0.85 for the four 'Go' trials on the reaction time test. Although this is considered acceptable (Streiner, 2003), other computerised reaction time tests such as the Deary-Liewald task (Deary et al., 2011) have yielded alpha coefficients >0.90 across greater numbers of trials (20 trials for simple reaction time and 40 for four-choice reaction time). Cronbach's alpha was 0.62 for the 13 items on the UK Biobank reasoning test (Hagenaars et al., 2016), which is lower than values reported (range 0.87 to 0.98) for subtests of a 'gold standard' intelligence test, the Wechsler Adult Intelligence Scale Fourth Edition (Wechsler, 2010). Cronbach's alpha coefficients for reasoning and reaction time tests in cohort subgroups analysed in the present thesis are reported in the relevant chapters.

Test reliability has important implications for both bias and power in statistical analyses. Random measurement error in independent variables causes bias towards the null in regression slope estimates (Hutcheon, Chiolero, & Hanley, 2010). Random measurement error in the dependent variable does not bias the slope estimate, but does increase its standard error. This latter problem can be addressed by using larger study samples.

Increasing the sample size does not mitigate the problem of bias arising from unreliably-measured independent variables, instead simply producing slope estimates that are more precisely wrong (Hutcheon et al., 2010). In the studies reported in this thesis, the cognitive test scores were the dependent variables in all the analyses.

3.7 Genetic data

Participants provided a blood sample at the baseline assessment, and genotyping was carried out centrally by UK Biobank. An interim genotypic dataset covering approximately 150,000 participants was made available to researchers in mid-2015, and the full cohort dataset was released in mid-2017.

3.7.1 Genotyping procedure

Full details of the genotyping, imputation and quality control processes used by UK Biobank are publicly available at http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100314 and in Bycroft et al. (2017). Direct genotyping was performed using two custom Affymetrix arrays: approximately 50,000 participants were genotyped on the UK BiLEVE Axiom array, which was designed for the BiLEVE study of lung function (a partner study of UK Biobank), and the remainder were genotyped using the UK Biobank Axiom array. The two arrays are very similar, with over 95% common marker content. The arrays included more than 800,000 single nucleotide polymorphisms (SNP), chosen because of known or likely associations with a wide range of diseases and health-related phenotypes, as well as to provide good genome-wide coverage for imputation purposes in European populations across common (>5%) and low (1-5%) minor allele frequency (MAF) ranges. Each SNP marker encodes variation in the population at a single base pair locus in the DNA sequence, and the MAF value indicates the proportion of the population carrying the second most common of the possible alleles at that SNP location.

The directly genotyped data were imputed by UK Biobank to reference panels from the Haplotype Reference Consortium (http://www.haplotype-reference-consortium.org/site) and the UK10K project and the 1000 Genomes Project (Phase 3)[3]. Imputation allows unobserved genotypes to be inferred on the basis of known clusters of genotypes in the population: a set of SNPs inherited together is termed a haplotype, and observed data from

[3] Only the Haplotype Reference Consortium imputed data were available at the time the studies in this thesis were conducted, due to quality control problems with the UK10K and 1000 Genomes imputations.

a subset of that haplotype can be mapped to a population reference panel to reconstruct the genotype of the unobserved alleles.

A dataset including approximately 40 million markers was made available to researchers. The quality of imputation at each SNP was indicated in the form of an information score between 0 and 1, where 1 indicates that there is no uncertainty in the imputed genotype and 0 means that there is complete uncertainty. Data were provided regarding pair-wise relatedness of individuals within the cohort, mismatch between reported and genetic sex, and probable white British genetic ancestry. Principal components analysis was conducted by UK Biobank to identify ancestral population structure within the cohort (Price et al., 2006); spurious associations arising from population stratification may occur if study samples include groups of individuals who differ systematically in both genetic ancestry and the phenotype of interest (Turner et al., 2011). Data were also provided regarding the MAF of each SNP, and Hardy-Weinberg test results; significant deviations from Hardy-Weinberg equilibrium—the expected ratios of homozygous and heterozygous genotypes in the population—may indicate genotyping errors or population structure effects (Turner et al., 2011). These additional variables can be used to exclude SNPs or individual participants from analysis, or to adjust analyses for confounding due to population stratification.

3.7.2 Polygenic scores

For the studies in this thesis, the genotypic data were used to construct polygenic scores derived from previous genome-wide association studies (GWAS) of phenotypes of interest. A polygenic approach was used because psychological and cognitive traits are likely to be explained at the population level by multiple genetic variants, each exerting small effects. This approach is also well suited to the genotypic data available in UK Biobank, which comprise a very large number of SNPs across the genome, thus providing rich information for use in dimensional scores.

The polygenic scoring method was developed in recent years as it became clear from GWAS of complex traits that groups of markers, which individually did not reach a genome-wide statistical significance threshold, could collectively explain a meaningful proportion of phenotypic variation (Dudbridge, 2013). There are several ways of constructing polygenic scores. In their simplest form, genotypes across a small set of SNPs identified via previous research as being of interest can be summed to create an aggregate

score, which may be weighted by the strength of their association with the target phenotype (Plomin & Simpson, 2013). Increasingly, however, researchers are making use of GWAS results to construct genome-wide polygenic scores (GPS), which use information from all available SNPs, thresholded at various statistical significance levels. Using this method, summary statistics (directional effect sizes and *p* values) for every marker included in a GWAS of a given phenotype, regardless of whether they achieved genome-wide statistical significance, are used to create a weighted polygenic score in an independent sample (Dudbridge, 2013). The use of GPS is becoming increasingly common in mental health and cognitive research (Plomin & Deary, 2015; Wray et al., 2014), where the phenotypes of interest are assumed to have a polygenic architecture. Polygenic scoring has also been used to investigate pleiotropy, whereby a given polygenic profile score may be associated with multiple phenotypes (Hagenaars et al., 2016).

It has been pointed out that additive polygenicity is not the only possible explanation for the findings of GWAS and polygenic scoring analyses of complex phenotypes (K. J. Mitchell, 2012), and that the influence of diverse rare variants (which are not generally captured by SNP-based GWAS methods) with large effects is likely to be important at the extreme ends of the distribution of cognitive ability in the population (Le Hellard & Steen, 2014). Nevertheless, polygenic scores based on common SNPs remain a useful way to measure background population variation in large studies, and are employed in this thesis for that reason. GPS were calculated using a bespoke script for R (https://www.r-project.org/) and PLINK (https://www.cog-genomics.org/plink2) software. Details of the summary statistics and scoring specifications employed are provided in the relevant chapters.

3.8 Summary

UK Biobank is a rich resource containing genetic, environmental, sociodemographic, lifestyle, medical and cognitive data for over half a million adults in Britain. It provides valuable opportunities to investigate the prevalence of cognitive impairment and variation in cognitive performance in groups with a history of psychiatric or neurological conditions, taking into account potential confounders measured in a standard way. The assessment visit process necessitated by the very large scale of the cohort means that some baseline measures, including the cognitive assessments, were briefer and hence less reliable than those typically used in smaller clinic-based studies. Factors that may have influenced the likelihood of responding to the study invitation must be considered, especially where these

relate to the exposures and outcomes under study, and similar considerations are important when making assumptions about missing data mechanisms. These benefits and limitations of the UK Biobank data resource were taken into account when planning, analysing and interpreting the studies reported in subsequent chapters.

Chapter 4 Identifying psychiatric and neurological groups in UK Biobank

This chapter describes the process of identifying exposure groups of interest in UK Biobank. Multiple information sources were used to identify participants with a pre-baseline history of mood disorder, schizophrenia, MS or PD, and those with no psychiatric or neurological history. The characteristics of the groups identified using the different information sources were compared, in order to inform decisions regarding which exposure group definitions would be used in subsequent studies.

4.1 Information sources

Three information sources were available, each with different levels of coverage of the cohort: self-reported diagnoses, linked NHS records and a touchscreen mood questionnaire. The web-based mental health questionnaire administered in 2016 was not used to ascertain the exposure groups, because this post-dated the cognitive assessment which was the outcome under study in this thesis. The current depressive symptom score (from the PHQ items) and data regarding current medications did not contribute to any of the exposure group classifications.

4.1.1 Self-reported diagnoses

All participants were asked to report lifetime history of medical or psychiatric diagnoses made by a doctor, using touchscreen questions supplemented by a staff interview. Responses were coded into categories by UK Biobank analysts. This information source covered the whole cohort, and contributed to the ascertainment of all the exposures of interest.

4.1.2 Linked NHS records

Medical history was available from hospital records, retrieved by UK Biobank from NHS data providers across the UK and linked centrally to UK Biobank data using participant identifiers. Records covered hospital admissions (in-patient and day case only) starting in the mid-1990s (dates varied by country). The data analysed were ICD-10 diagnostic codes from any position in the record. Only data prior to the UK Biobank assessment date were used, to identify past diagnoses. Hospital records from England and Wales included both

acute and psychiatric hospitals, but records from Scotland covered acute hospitals only, at the time the present research was conducted. Participants in Scotland (7.1% of the whole cohort) were therefore coded as 'missing' when ascertaining the psychiatric exposures using this information source, but they contributed data when ascertaining the MS and PD exposures. Linked NHS records were sought by UK Biobank for the full cohort; participants with no linked hospital record are assumed to have had no in-patient or day case admission during the available period.

4.1.3 Touchscreen mood questionnaire

The touchscreen mood questionnaire[4] covering lifetime history of mood symptoms and medical help-seeking for mental health was described in Chapter 3. This was added to the assessment schedule in the last two years of baseline recruitment, and data are therefore only available for participants assessed in that period (approximately one-third of the cohort). The questions covered depression and mania symptoms only, and so these data contributed to the ascertainment of the mood disorder exposures but not schizophrenia, MS or PD.

4.2 Ascertaining exposures of interest

Table 4.1 shows which exposures of interest were potentially ascertainable from the three information sources.

Table 4.1 - Information sources for each exposure

	Mania/ bipolar	Major depression				Schizophrenia	Multiple sclerosis	Parkinson's disease
		Any depression	Severe recurrent	Mild-moderate recurrent	Single episode			
Self-reported diagnoses	✓ All	✓ All	×	×	×	✓ All	✓ All	✓ All
NHS hospital records	✓ England & Wales	✓ England & Wales	✓ England & Wales	✓ England & Wales	✓ England & Wales	✓ England & Wales	✓ All	✓ All
Touchscreen mood questionnaire	✓ Last 2 years	✓ Last 2 years	✓ Last 2 years	✓ Last 2 years	✓ Last 2 years	×	×	×

[4] For the remainder of this thesis, 'touchscreen mood questionnaire' refers to the questions about lifetime mood disorder symptoms only; the four PHQ questions are referred to separately as the 'current depressive symptom score'.

Each exposure group, and a comparison group with none of the exposures, were ascertained as described below. Participants who did not meet the criteria for any of the exposed or unexposed groups were not further considered.

4.2.1 Mood disorders

The mood disorder exposures (mania/BD and major depression) were classified hierarchically into mutually exclusive groups within each information source. Using the self-reported diagnosis data, the hierarchy order was: mania/BD ('mania/bipolar/manic depression'); any depression ('depression' or 'post-natal depression'). Using the hospital ICD-10 codes, the order was: mania/BD (F30x or F31x); severe recurrent major depression (F33.2 or F33.3); mild-moderate recurrent major depression (all other F33 codes); single episode of major depression (F32x). Using the touchscreen mood questionnaire data, the hierarchy was as described by D. J. Smith et al. (2013): mania/BD (BD-I and BD-II combined); severe recurrent major depression; mild-moderate recurrent major depression; single episode of major depression (see Appendix O for information about the item responses contributing to these classifications). To permit comparison between the self-reported diagnosis classification and the other two information sources, all categories of major depression derived from the ICD-10 codes and the touchscreen mood questionnaire were also combined into an overall 'any depression' category. Owing to the limited detail in the self-reported diagnosis data and the touchscreen mood questionnaire with regard to bipolar features, no distinction was made between single manic episode, bipolar disorder type I and bipolar disorder type II; the term 'mania/BD' is therefore used throughout.

4.2.2 Schizophrenia

Participants were classified by self-report of 'schizophrenia' and by hospital ICD-10 codes in the F20x category. The touchscreen mood questionnaire did not cover data relevant to this exposure.

4.2.3 Multiple sclerosis

Participants were classified by self-report of 'multiple sclerosis' and by hospital ICD-10 code G35.

4.2.4 Parkinson's disease

Participants were classified by self-report of 'Parkinson's disease' and by hospital ICD-10 code G20.

4.2.5 Unexposed comparison group

A single comparison group was constructed as a common reference for all of the exposure groups. Participants in this group were considered to be unexposed with regard to the conditions of interest, as well as any other psychiatric condition or any condition affecting brain function. This group met all of the following criteria: provided data on the touchscreen mood questionnaire and were not classified in the mania/BD or major depression groups; no hospital ICD-10 code of any psychiatric or brain condition (as listed in Appendix P); no self-reported diagnosis of any psychiatric or brain condition (as listed in Appendix Q). Two versions of this group were constructed, 'broad' and 'narrow'. Using the broad definition, sub-threshold mood symptoms were permitted on the touchscreen mood questionnaire; using the narrow definition, such symptoms were not permitted. Medical conditions other than those listed in the appendices were not excluded.

4.3 Comparison of group characteristics by each ascertainment method

Sociodemographic, lifestyle, medication and psychological measures were summarised and compared descriptively within each exposure group, by information source. Details of the medications coded in each class (psychotropics; MS disease modifying treatments; medications used in PD) are given in Appendix R. Exposed participants were also grouped into two further categories, representing 'broad' and 'narrow' exposure classifications. The broad classification required participants to have met the exposure criteria by at least one information source, and the narrow classification required them to have met the exposure criteria by at least two information sources.

The mania/BD and major depression exposure groups were mutually exclusive within each information source, but participant overlap was otherwise possible between the various exposed groups. Within each exposed group, no exclusions were applied regarding other psychiatric, neurological or medical comorbidities. No statistical tests were conducted for the data in the tables that follow, because of the potential participant overlap across information sources and definitions.

4.3.1 Mood disorders

4.3.1.1 Mania/BD

Table 4.2 shows the characteristics of the mania/BD exposure group as identified by each information source, and grouped by broad and narrow definitions. The most striking difference was in the use of psychotropic medication, which was much less common in the group identified via the touchscreen mood questionnaire.

4.3.1.2 Major depression

Table 4.3 shows the characteristics of the ‘any depression’ group, by ascertainment source and broad versus narrow definitions. This indicates that participants identified through hospital records were more likely to live in areas of greater deprivation, were less likely to have a degree, and were more likely to be current smokers and former alcohol drinkers. Participants identified through the touchscreen mood questionnaire were much less likely to be on psychotropic medication, and had lower neuroticism and current depressive symptom scores. Additional tables for the severe recurrent, mild-moderate recurrent and single episode depression groups are given in Appendix S.

Table 4.2 - Mania/bipolar disorder group characteristics

	Information source			Definition	
	Self-reported diagnosis	**NHS hospital records**	**Touchscreen mood questionnaire**	**Broad**	**Narrow**
n	1,412	688	1,615	3,020	607
Age, M (SD)	55.4 (7.9)	54.8 (8.0)	54.5 (8.1)	54.9 (8.0)	54.8 (8.0)
Female, *n* (%)[a]	788 (55.8)	412 (59.9)	790 (48.9)	1,589 (52.6)	345 (56.8)
Ethnic group, *n* (%)[a]					
White	1,333 (95.1)	619 (91.6)	1,422 (88.8)	2,727 (91.2)	563 (93.4)
Asian/Asian British	18 (1.3)	6 (0.9)	66 (4.1)	84 (2.8)	6 (1.0)
Black/black British	22 (1.6)	23 (3.4)	53 (3.3)	84 (2.8)	13 (2.2)
Other	29 (2.1)	28 (4.1)	61 (3.8)	94 (3.1)	21 (3.5)
Townsend quintile[b], *n* (%)[a]					
Qu1 (least deprived)	179 (12.7)	82 (11.9)	174 (10.8)	356 (11.8)	71 (11.7)
Qu2	189 (13.4)	90 (13.1)	209 (13.0)	405 (13.4)	78 (12.9)
Qu3	233 (16.5)	101 (14.7)	269 (16.7)	492 (16.3)	99 (16.3)
Qu4	300 (21.3)	128 (18.6)	402 (24.9)	676 (22.4)	133 (22.0)
Qu5 (most deprived)	509 (36.1)	286 (41.6)	560 (34.7)	1,088 (36.1)	225 (37.1)
Has a degree, *n* (%)[a]	585 (41.9)	237 (35.4)	578 (36.0)	1,117 (37.5)	250 (41.5)
Smoking status, *n* (%)[a]					
Never	612 (43.6)	287 (42.6)	683 (42.5)	1,281 (42.8)	262 (43.5)
Former	468 (33.4)	204 (30.3)	577 (35.9)	1,024 (34.2)	196 (32.6)
Current	323 (23.0)	183 (27.2)	347 (21.6)	689 (23.0)	144 (23.9)
Alcohol frequency, *n* (%)[a]					
Daily/almost daily	256 (18.2)	86 (12.8)	326 (20.3)	561 (18.7)	99 (16.4)
3-4 times per week	236 (16.8)	97 (14.4)	267 (16.6)	492 (16.4)	92 (15.3)
1-2 times per week	279 (19.9)	151 (22.4)	353 (21.9)	641 (21.4)	126 (20.9)
1-3 times per month	161 (11.5)	78 (11.6)	201 (12.5)	355 (11.8)	75 (12.4)
Special occasions only	235 (16.7)	113 (16.8)	244 (15.2)	473 (15.8)	99 (16.4)
Never (former drinker)	162 (11.5)	91 (13.5)	129 (8.0)	296 (9.9)	74 (12.3)
Never (not former drinker)	75 (5.3)	58 (8.6)	90 (5.6)	180 (6.0)	38 (6.3)
Any psychotropic medication, *n* (%)[a]	1,212 (86.3)	562 (82.9)	514 (32.3)	1,663 (55.8)	542 (89.7)
Lithium, *n* (%)[a]	458 (33.5)	190 (28.6)	86 (5.4)	501 (17.1)	209 (35.3)
Other mood stabiliser, *n* (%)[a]	451 (32.7)	266 (39.8)	130 (8.2)	531 (18.0)	267 (44.7)
SSRI antidepressant, *n* (%)[a]	322 (23.5)	98 (14.7)	203 (12.8)	508 (17.3)	102 (17.2)
Other antidepressant, *n* (%)[a]	270 (19.8)	146 (22.0)	165 (10.4)	428 (14.6)	128 (21.7)
Traditional antipsychotics, *n* (%)[a]	101 (7.4)	48 (7.3)	27 (1.7)	131 (4.5)	39 (6.6)
Second generation antipsychotics, *n* (%)[a]	387 (28.2)	225 (33.7)	100 (6.3)	470 (16.0)	205 (34.5)
Neuroticism score, M (SD)	7.3 (3.6)	6.8 (3.7)	6.6 (3.6)	6.8 (3.6)	7.1 (3.6)
Current depressive symptom score, M (SD)	3.6 (3.3)	3.4 (3.3)	3.6 (3.2)	3.6 (3.3)	3.4 (3.2)

M, mean; Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table 4.3 - 'Any depression' group characteristics

	Information source			Definition	
	Self-reported diagnosis	**NHS hospital records**	**Touchscreen mood questionnaire**	**Broad**	**Narrow**
n	28,366	4,279	31,844	56,425	7,583
Age, M (SD)	55.4 (7.8)	55.6 (8.0)	55.7 (8.0)	55.6 (7.9)	55.1 (7.9)
Female, *n* (%)[a]	18,744 (66.1)	2,520 (58.9)	20,510 (64.4)	36,574 (64.8)	4,923 (64.9)
Ethnic group, *n* (%)[a]					
White	27,125 (96.2)	4,021 (94.7)	29,981 (94.5)	53,515 (95.3)	7,171 (94.8)
Asian/Asian British	376 (1.3)	80 (1.9)	586 (1.9)	887 (1.6)	144 (1.9)
Black/black British	240 (0.9)	58 (1.4)	542 (1.7)	718 (1.3)	108 (1.4)
Other	467 (1.7)	87 (2.1)	624 (2.0)	1,026 (1.8)	140 (1.9)
Townsend quintile[b], *n* (%)[a]					
Qu1 (least deprived)	4,571 (16.1)	475 (11.1)	4,896 (15.4)	8,872 (15.8)	1,015 (13.4)
Qu2	4,942 (17.5)	596 (14.0)	5,911 (18.6)	10,071 (17.9)	1,304 (17.2)
Qu3	5,229 (18.5)	643 (15.1)	6,440 (20.3)	10,836 (19.2)	1,388 (18.4)
Qu4	5,882 (20.8)	910 (21.3)	7,357 (23.2)	12,380 (22.0)	1,663 (22.0)
Qu5 (most deprived)	7,692 (27.2)	1,648 (38.6)	7,175 (22.6)	14,165 (25.2)	2,194 (29.0)
Has a degree, *n* (%)[a]	8,431 (30.1)	928 (22.3)	11,308 (35.7)	18,188 (32.6)	2,344 (31.2)
Smoking status, *n* (%)[a]					
Never	13,652 (48.4)	1,811 (42.7)	15,975 (50.3)	27,577 (49.1)	3,611 (47.8)
Former	9,788 (34.7)	1,384 (32.6)	11,721 (36.9)	20,210 (36.0)	2,559 (33.8)
Current	4,780 (16.9)	1,047 (24.7)	4,074 (12.8)	8,402 (15.0)	1,392 (18.4)
Alcohol frequency, *n* (%)[a]					
Daily/almost daily	5,270 (18.7)	687 (16.2)	6,285 (19.8)	10,801 (19.2)	1,361 (18.0)
3-4 times per week	5,066 (17.9)	571 (13.5)	6,726 (21.1)	11,020 (19.6)	1,281 (17.0)
1-2 times per week	6,295 (22.3)	862 (20.4)	7,660 (24.1)	13,097 (23.3)	1,634 (21.6)
1-3 times per month	3,493 (12.4)	482 (11.4)	4,057 (12.8)	7,031 (12.5)	938 (12.4)
Special occasions only	4,459 (15.8)	773 (18.3)	4,220 (13.3)	8,153 (14.5)	1,207 (16.0)
Never (former drinker)	2,172 (7.7)	569 (13.4)	1,631 (5.1)	3,609 (6.4)	699 (9.3)
Never (not former drinker)	1,488 (5.3)	291 (6.9)	1,242 (3.9)	2,557 (4.5)	435 (5.8)
Any psychotropic medication, *n* (%)[a]	20,062 (72.0)	2,663 (63.4)	6,502 (20.7)	23,280 (42.0)	5,530 (73.5)
Lithium, *n* (%)[a]	290 (1.1)	116 (2.9)	100 (0.3)	373 (0.7)	117 (1.6)
Other mood stabiliser, *n* (%)[a]	460 (1.7)	234 (5.7)	305 (1.0)	768 (1.4)	196 (2.7)
SSRI antidepressant, *n* (%)[a]	14,211 (51.5)	1,423 (34.4)	3,919 (12.5)	15,529 (28.2)	3,792 (50.8)
Other antidepressant, *n* (%)[a]	5,879 (21.6)	1,098 (26.7)	2,177 (7.0)	7,239 (13.2)	1,737 (23.5)
Neuroticism score, M (SD)	7.5 (3.2)	7.5 (3.3)	5.7 (3.3)	6.4 (3.4)	7.4 (3.2)
Current depressive symptom score, M (SD)	4.0 (3.2)	4.5 (3.5)	2.4 (2.6)	3.1 (3.0)	4.0 (3.4)

M, mean; Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

4.3.2 Schizophrenia

Table 4.4 shows that the schizophrenia group characteristics were similar regardless of ascertainment source. Some differences were evident in the proportions taking psychotropic medication, which were slightly higher in the group ascertained via self-reported diagnosis. The proportion with black or black British ethnicity was higher in the group ascertained through hospital records.

4.3.3 Multiple sclerosis

Table 4.5 shows the characteristics of the MS groups. Differences were seen in deprivation and education measures only, with the group ascertained via hospital records being somewhat more likely to live in areas of greater deprivation and less likely to have a degree.

4.3.4 Parkinson's disease

Similar to the MS groups, Table 4.6 shows that participants with PD ascertained via hospital records were somewhat more likely to live in areas of greater deprivation and less likely to have a degree.

4.3.5 Unexposed comparison group

Table 4.7 indicates that the characteristics of the unexposed comparison group were virtually identical regardless of whether the broad (subthreshold mood symptoms permitted) or narrow (no subthreshold symptoms) definitions were used.

Table 4.4 - Schizophrenia group characteristics

	Information source		Definition	
	Self-reported diagnosis	**NHS hospital records**	**Broad**	**Narrow**
n	619	490	850	259
Age, M (SD)	53.8 (8.1)	53.4 (8.1)	53.9 (8.1)	52.6 (8.2)
Female, *n* (%)[a]	201 (32.5)	166 (33.9)	293 (34.5)	74 (28.6)
Ethnic group, *n* (%)[a]				
White	537 (88.3)	403 (84.7)	719 (86.7)	221 (86.7)
Asian/Asian British	16 (2.6)	16 (3.4)	24 (2.9)	8 (3.1)
Black/black British	29 (4.8)	37 (7.8)	49 (5.9)	17 (6.7)
Other	26 (4.3)	20 (4.2)	37 (4.5)	9 (3.5)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	24 (3.9)	17 (3.5)	33 (3.9)	8 (3.1)
Qu2	32 (5.2)	27 (5.5)	48 (5.7)	11 (4.3)
Qu3	49 (8.0)	45 (9.2)	75 (8.9)	19 (7.4)
Qu4	120 (19.5)	89 (18.2)	165 (19.5)	44 (17.1)
Qu5 (most deprived)	391 (63.5)	311 (63.6)	526 (62.1)	176 (68.2)
Has a degree, *n* (%)[a]	158 (26.4)	116 (24.7)	212 (25.9)	65 (25.3)
Smoking status, *n* (%)[a]				
Never	223 (36.4)	165 (34.7)	299 (35.9)	89 (34.9)
Former	173 (28.2)	138 (29.0)	237 (28.4)	74 (29.0)
Current	217 (35.4)	173 (36.3)	298 (35.7)	92 (36.1)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	96 (15.6)	63 (13.1)	125 (14.9)	34 (13.2)
3-4 times per week	63 (10.3)	45 (9.4)	84 (10.0)	24 (9.3)
1-2 times per week	120 (19.5)	94 (19.6)	161 (19.2)	53 (20.6)
1-3 times per month	54 (8.8)	47 (9.8)	79 (9.4)	22 (8.6)
Special occasions only	122 (19.9)	86 (17.9)	157 (18.8)	51 (19.8)
Never (former drinker)	103 (16.8)	94 (19.6)	151 (18.0)	46 (17.9)
Never (not former drinker)	56 (9.1)	51 (10.6)	80 (9.6)	27 (10.5)
Any psychotropic medication, *n* (%)[a]	528 (87.9)	391 (81.8)	688 (83.2)	231 (91.7)
Traditional antipsychotics, *n* (%)[a]	153 (26.4)	83 (18.2)	174 (21.9)	62 (25.6)
Second generation antipsychotics, *n* (%)[a]	346 (58.7)	262 (56.1)	439 (54.4)	169 (67.9)
Neuroticism score, M (SD)	6.8 (3.6)	6.9 (3.6)	6.8 (3.6)	6.9 (3.5)
Current depressive symptom score, M (SD)	3.5 (3.1)	3.9 (3.4)	3.7 (3.3)	3.9 (3.3)

M, mean; Qu, quintile; SD, standard deviation.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table 4.5 - Multiple sclerosis group characteristics

	Information source		Definition	
	Self-reported diagnosis	**NHS hospital records**	**Broad**	**Narrow**
n	1,777	1,059	1,905	931
Age, M (SD)	55.3 (7.5)	55.3 (7.5)	55.4 (7.5)	55.3 (7.5)
Female, *n* (%)[a]	1,300 (73.2)	765 (72.2)	1,400 (73.5)	665 (71.4)
Ethnic group, *n* (%)[a]				
White	1,734 (98.2)	1,024 (97.4)	1,859 (98.2)	899 (97.4)
Asian/Asian British	5 (0.3)	4 (0.4)	5 (0.3)	4 (0.4)
Black/black British	9 (0.5)	7 (0.7)	11 (0.6)	5 (0.5)
Other	18 (1.0)	16 (1.5)	19 (1.0)	15 (1.6)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	353 (19.9)	210 (19.9)	382 (20.1)	181 (19.5)
Qu2	366 (20.6)	189 (17.9)	382 (20.1)	173 (18.6)
Qu3	370 (20.9)	209 (19.8)	389 (20.5)	190 (20.5)
Qu4	349 (19.7)	211 (20.0)	373 (19.6)	187 (20.1)
Qu5 (most deprived)	335 (18.9)	238 (22.5)	375 (19.7)	198 (21.3)
Has a degree, *n* (%)[a]	602 (34.6)	318 (30.8)	642 (34.4)	278 (30.7)
Smoking status, *n* (%)[a]				
Never	828 (47.1)	488 (46.6)	897 (47.5)	419 (45.5)
Former	651 (37.0)	376 (35.9)	690 (36.6)	337 (36.6)
Current	281 (16.0)	184 (17.6)	301 (15.9)	164 (17.8)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	349 (19.7)	188 (17.8)	368 (19.4)	169 (18.2)
3-4 times per week	321 (18.1)	170 (16.1)	342 (18.0)	149 (16.1)
1-2 times per week	425 (24.0)	236 (22.4)	447 (23.5)	214 (23.1)
1-3 times per month	217 (12.2)	135 (12.8)	232 (12.2)	120 (12.9)
Special occasions only	271 (15.3)	182 (17.3)	295 (15.5)	158 (17.0)
Never (former drinker)	113 (6.4)	86 (8.2)	128 (6.7)	71 (7.7)
Never (not former drinker)	77 (4.3)	57 (5.4)	88 (4.6)	46 (5.0)
MS disease-modifying medication, *n* (%)[a]	171 (10.5)	124 (13.0)	175 (10.0)	120 (14.3)
Neuroticism score, M (SD)	4.5 (3.4)	4.6 (3.3)	4.6 (3.4)	4.5 (3.3)
Current depressive symptom score, M (SD)	2.6 (2.4)	2.9 (2.5)	2.6 (2.4)	2.9 (2.5)

M, mean; MS, multiple sclerosis; Qu, quintile; SD, standard deviation.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table 4.6 - Parkinson's disease group characteristics

	Information source		Definition	
	Self-reported diagnosis	**NHS hospital records**	**Broad**	**Narrow**
n	858	381	916	323
Age, M (SD)	62.3 (5.4)	62.3 (5.6)	62.2 (5.6)	62.5 (5.2)
Female, *n* (%)[a]	318 (37.1)	159 (41.7)	346 (37.8)	131 (40.6)
Ethnic group, *n* (%)[a]				
White	824 (96.6)	362 (95.5)	876 (96.3)	310 (96.3)
Asian/Asian British	12 (1.4)	9 (2.4)	15 (1.7)	6 (1.9)
Black/black British	7 (0.8)	2 (0.5)	8 (0.9)	1 (0.3)
Other	10 (1.2)	6 (1.6)	11 (1.2)	5 (1.6)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	214 (25.0)	88 (23.2)	222 (24.3)	80 (24.8)
Qu2	176 (20.5)	78 (20.6)	188 (20.6)	66 (20.5)
Qu3	164 (19.1)	72 (19.0)	177 (19.4)	59 (18.3)
Qu4	157 (18.3)	63 (16.6)	167 (18.3)	53 (16.5)
Qu5 (most deprived)	146 (17.0)	78 (20.6)	160 (17.5)	64 (19.9)
Has a degree, *n* (%)[a]	269 (32.0)	95 (25.5)	276 (30.8)	88 (27.8)
Smoking status, *n* (%)[a]				
Never	526 (61.8)	230 (61.7)	555 (61.3)	201 (63.2)
Former	273 (32.1)	118 (31.6)	293 (32.3)	98 (30.8)
Current	52 (6.1)	25 (6.7)	58 (6.4)	19 (6.0)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	170 (19.9)	70 (18.5)	178 (19.5)	62 (19.3)
3-4 times per week	158 (18.5)	59 (15.6)	167 (18.3)	50 (15.5)
1-2 times per week	209 (24.4)	87 (23.0)	221 (24.2)	75 (23.3)
1-3 times per month	79 (9.2)	39 (10.3)	86 (9.4)	32 (9.9)
Special occasions only	115 (13.5)	64 (16.9)	124 (13.6)	55 (17.1)
Never (former drinker)	73 (8.5)	33 (8.7)	79 (8.7)	27 (8.4)
Never (not former drinker)	51 (6.0)	27 (7.1)	57 (6.3)	21 (6.5)
PD medication, *n* (%)[a]	711 (88.3)	310 (84.9)	721 (84.0)	300 (96.2)
Neuroticism score, M (SD)	4.0 (3.3)	4.4 (3.4)	4.1 (3.3)	4.4 (3.4)
Current depressive symptom score, M (SD)	2.3 (2.4)	2.6 (2.5)	2.3 (2.4)	2.6 (2.5)

M, mean; PD, Parkinson's disease; Qu, quintile; SD, standard deviation.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table 4.7 - Unexposed comparison group characteristics

	Definition	
	Broad	**Narrow**
n	104,410	80,505
Age, M (SD)	57.0 (8.2)	57.3 (8.1)
Female, *n* (%)[a]	52,210 (50.0)	39,928 (49.6)
Ethnic group, *n* (%)[a]		
White	94,798 (91.1)	73,177 (91.3)
Asian/Asian British	3,643 (3.5)	2,832 (3.5)
Black/black British	3,121 (3.0)	2,369 (3.0)
Other	2,450 (2.4)	1,818 (2.3)
Townsend quintile[b], *n* (%)[a]		
Qu1 (least deprived)	18,130 (17.4)	14,303 (17.8)
Qu2	21,340 (20.5)	16,682 (20.8)
Qu3	21,782 (20.9)	16,879 (21.0)
Qu4	23,337 (22.4)	17,712 (22.0)
Qu5 (most deprived)	19,664 (18.9)	14,803 (18.4)
Has a degree, *n* (%)[a]	36,082 (34.9)	27,012 (33.9)
Smoking status, *n* (%)[a]		
Never	60,061 (57.7)	46,841 (58.4)
Former	35,126 (33.8)	26,913 (33.6)
Current	8,851 (8.5)	6,440 (8.0)
Alcohol frequency, *n* (%)[a]		
Daily/almost daily	22,026 (21.1)	17,070 (21.2)
3-4 times per week	24,920 (23.9)	19,356 (24.1)
1-2 times per week	26,928 (25.8)	20,728 (25.8)
1-3 times per month	11,266 (10.8)	8,428 (10.5)
Special occasions only	11,449 (11.0)	8,821 (11.0)
Never (former drinker)	2,783 (2.7)	2,118 (2.6)
Never (not former drinker)	4,963 (4.8)	3,924 (4.9)
Any psychotropic medication, *n* (%)[a]	1,424 (1.4)	1,072 (1.4)
Neuroticism score, M (SD)	3.3 (2.9)	3.2 (2.9)
Current depressive symptom score, M (SD)	1.1 (1.6)	1.1 (1.5)

M, mean; Qu, quintile; SD, standard deviation.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

4.4 Overlap between ascertainment methods

The extent of overlap between participants ascertained via each information source was examined, as depicted in the Venn diagrams below. For the mood disorder group calculations, only participants with complete data on the touchscreen mood questionnaire were included, in order that the base population would be comparable between methods.

Figure 4.1 - Overlap between mania/bipolar disorder ascertainment sources

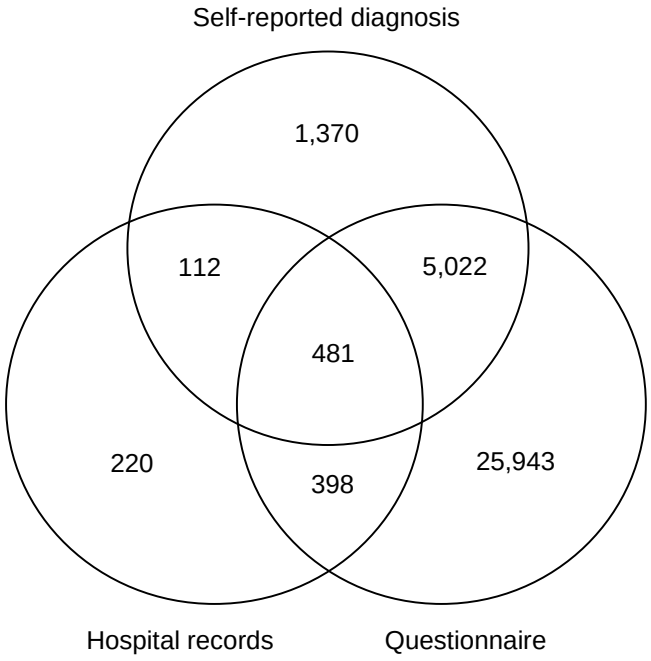


Figure 4.2 - Overlap between 'any depression' ascertainment sources

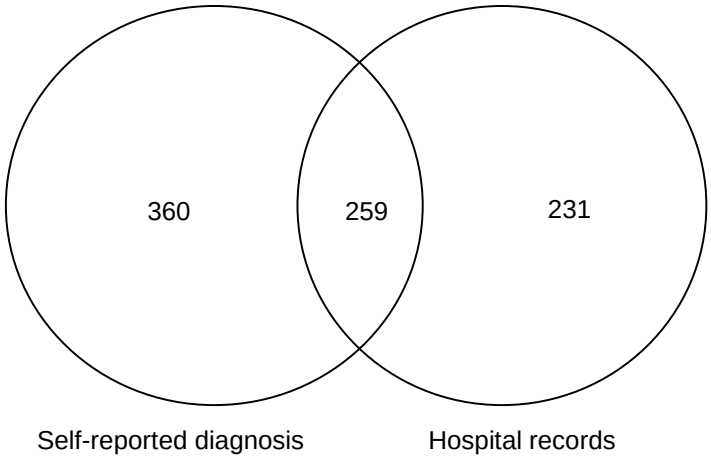


Figure 4.3 - Overlap between schizophrenia ascertainment sources

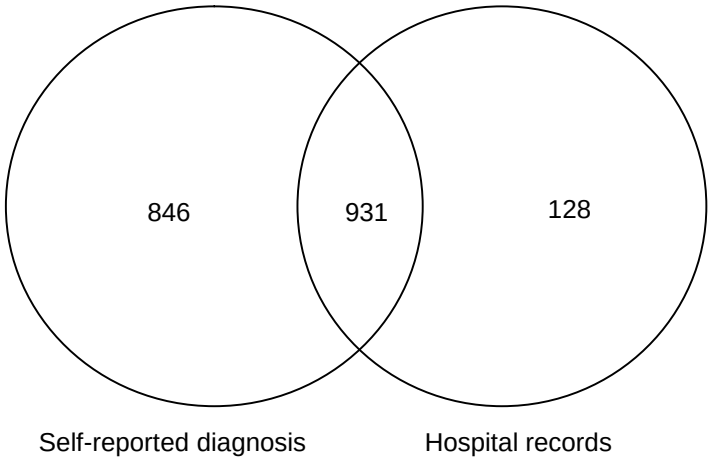


Figure 4.4 - Overlap between multiple sclerosis ascertainment sources

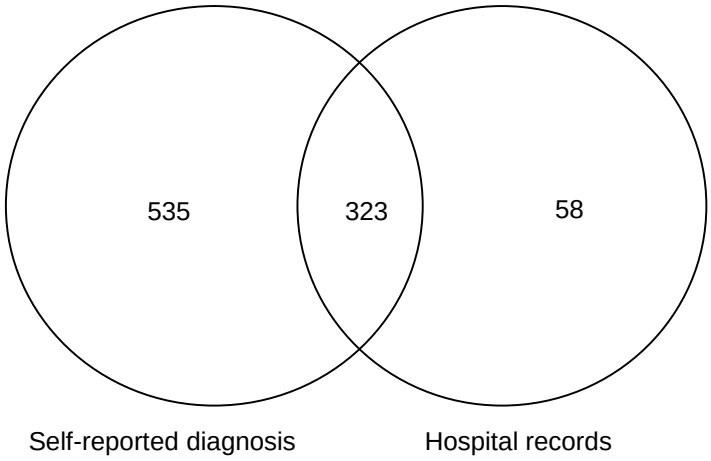


Figure 4.5 - Overlap between Parkinson's disease ascertainment sources

4.5 Discussion

There were marked differences in the numbers meeting the criteria for each exposure of interest, according to the three information sources available. For the mood disorder exposures, a much greater number were identified via the touchscreen mood questionnaire than via self-reported diagnoses or hospital records. This may indicate that the mood questionnaire was sensitive to milder illness history or that it misclassified people without true mood disorder, although it has previously been shown that the characteristics of the groups classified by this method are in line with what would be expected with regard to gender distribution, socioeconomic status, self-reported health rating, current depressive symptoms and smoking (D. J. Smith et al., 2013). Fewer participants reported having been diagnosed with mood disorder by a doctor, which may be unsurprising given that bipolar disorder and major depression diagnoses are under-recorded in primary care (Rait et al., 2009; D. J. Smith et al., 2011).

There were also disparities between the numbers of participants with self-reported and hospital records of each diagnosis, which may be largely explained by the fact that the records covered in-patient and day case admissions only, and did not extend further back than the mid-1990s. These disparities were smaller for the MS and PD groups, possibly because of the greater likelihood of a hospital admission for these disorders. It should be noted that a substantial minority of participants with a relevant diagnosis in the hospital records did not disclose this in the UK Biobank interview. This may have been because of participant preferences (for example, privacy concerns), or because the hospital record ICD-10 diagnosis did not reflect their understanding, perception or recall of their medical history at the time of the baseline assessment. The group classifications reported here accepted any instance of a relevant ICD-10 code at any time in the available records, and it is possible that revised diagnoses or clerical errors in the records may have contributed to discrepancies with the self-reported data (Davis, Sudlow, & Hotopf, 2016).

None of the three information sources can be said to be a ‘gold standard’: the self-reported diagnoses and the questionnaire responses are liable to under- and over-reporting, and the hospital records are time-limited, with ICD-10 codes that are subject to clinical revision or clerical error. A pragmatic approach is to use more than one classification method to define the exposures, as illustrated above with the broad and narrow definitions. The numbers in the broad and narrow groups for each exposure were again markedly different, for the same reasons as discussed above. There were also differences in their characteristics,

driven by the fact that many participants meeting the narrow definition will have had a hospital admission, possibly indicating a more severe illness course.

Despite the evident differences across each information source and the broad versus narrow definitions, certain characteristics support the validity of the distinctions between each exposure group, for example: women were over-represented in the depression and MS groups and under-represented in the schizophrenia and PD groups; the PD group was older than the other groups; the proportion with white ethnicity was lowest in the schizophrenia group and highest in the MS group; greater levels of deprivation were common in the mood disorder and schizophrenia groups but not those with MS or PD; smoking was most common in the mania/BD and schizophrenia groups and least common in PD. These patterns are in line with previous epidemiological evidence in these disorders (Andrade et al., 2003; de Leon & Diaz, 2005; Fajutrao et al., 2009; Gajwani, 2017; Kalia & Lang, 2015; Kamm et al., 2014; Kirkbride et al., 2012; MacDonald et al., 2000; D. F. MacKinnon, 2017; McGrath et al., 2008; Merikangas et al., 2011; Picchioni & Murray, 2007). Patterns of medication usage also supported the exposure group classifications.

4.6 Next steps

In consideration of the above findings, it was decided that the broad and narrow exposure group definitions would be used in the studies in this thesis. The broadly defined group was the primary group for analysis, and secondary analyses were conducted in the narrow group where relevant. The broadly and narrowly defined unexposed comparison groups were very similar, and so only the broadly defined group was analysed. Although subgroups with single episode, mild-moderate and severe recurrent depression were identifiable (Appendix S), the UK Biobank self-reported diagnosis data did not permit these distinctions, and the numbers in the narrowly defined versions of these subgroups were very small. For these reasons, only the ‘any depression’ exposure was used in the studies reported in this thesis. Reporting of all studies follows STROBE guidelines (von Elm et al., 2007).

Chapter 5 Prevalence of cognitive impairment in psychiatric and neurological groups in UK Biobank

This chapter reports on the prevalence of cognitive impairment in UK Biobank participants with a history of mania/BD, in comparison with those who have a history of major depression, schizophrenia, multiple sclerosis or Parkinson's disease. This is the first study to directly compare prevalence of cognitive impairment across these conditions, using consistent assessment methods and a single unexposed comparison group, and illustrates the power of using a large population cohort such as UK Biobank to investigate the 'cognitive footprint' of chronic psychiatric and neurological conditions.[5]

5.1 Background

As outlined in Chapter 1, cognitive impairment is a commonly reported feature of mood disorders, which persists between illness episodes and contributes substantially to ongoing disability. Estimates of cognitive impairment prevalence in euthymic adults with a history of mood disorder vary considerably, depending on the measures used and the definition of impairment. The systematic review in Chapter 2 showed that the prevalence in BD ranges from 5% to 58%, and in major depressive disorder it has been estimated at one-third to one-half (Rock et al., 2014). Cognitive impairment has been studied more extensively in schizophrenia; it is reportedly evident in the majority of patients—possibly as many as 80%—and has been proposed as a core diagnostic criterion (Kahn & Keefe, 2013; Keefe, 2008; Keefe & Fenton, 2007). The prevalence of cognitive impairment in mental health disorders may be compared with that in chronic neurological conditions: studies in MS have reported prevalence of 40-80% (Fischer et al., 2014; Patti et al., 2015; Rao, Leo, Bernardin, et al., 1991), and in PD the prevalence has been estimated at 50-55% (Svenningsson et al., 2012).

The comparative burden of cognitive impairment across disorders must be considered with reference to the relative prevalence of the disorders themselves. As noted in Chapter 1, lifetime prevalence in the UK per 100,000 population is estimated at up to 10,000 for

[5] The work described in this chapter has been published in Cullen, B., Smith, D. J., Deary, I. J., Evans, J. J., & Pell, J. P. (2017). The 'cognitive footprint' of psychiatric and neurological conditions: Cross-sectional study in the UK Biobank cohort. *Acta Psychiatrica Scandinavica, 135*, 593-605. doi:10.1111/acps.12733

major depression (NICE, 2011) and approximately 1,000 for BD (Fajutrao et al., 2009), compared with 400-1,000 for schizophrenia (McGrath et al., 2008), and 200 for both MS and PD (MacDonald et al., 2000). Despite the generally greater research focus on cognitive function in schizophrenia and neurological disorders, it is clear that the higher prevalence of mood disorders in the population means that the absolute number with associated cognitive impairment is likely to be substantial and of significant public health importance.

Previous studies of cognitive function in various psychiatric and neurological conditions are difficult to compare directly, because of variations in source populations, methods of recruitment, assessment tools, impairment definitions, composition of normative comparison groups, and adjustment for potential confounders. The UK Biobank cohort presents an opportunity to overcome these limitations. The aim of this cross-sectional analysis was to quantify the prevalence of cognitive impairment in adults with a history of mood disorder, schizophrenia, MS or PD, in the UK Biobank cohort.

5.2 Methods

5.2.1 Participants

Data from the full UK Biobank cohort at baseline (n = 502,642) were used.

5.2.2 Materials and procedure

Details of demographic, lifestyle and psychological assessment procedures were given in Chapter 3. The key sociodemographic variables for the present analyses were age, gender and educational attainment (dichotomised according to whether or not participants held a university/college degree), and additional measures were also reported to characterise the sample descriptively.

5.2.3 Exposed and unexposed groups

The exposures of interest were mania/BD, major depression (single or recurrent episodes), schizophrenia, MS and PD. Exposure status for each condition was classified using both broad and narrow definitions, as described in Chapter 4. A single unexposed comparison group was analysed, as described in Chapter 4 (broad definition). Participants who did not meet the criteria for any of the exposed or unexposed groups were not further analysed.

5.2.4 Cognitive impairment

Details of the cognitive assessment were given in Chapter 3. The measures analysed were: reasoning (total correct out of 13 items); reaction time (mean time in milliseconds to press a button in response to matching cards); numeric memory (longest numeric string recalled in reverse); visuospatial memory ('pairs matching' test: total errors when recalling the positions of matching cards); and prospective memory (successfully carrying out an instruction after a filled delay).

Performance on each of the five tests was classified as impaired or unimpaired. For all measures except prospective memory, the score distribution in the unexposed comparison group was converted into percentile ranks, and the raw score corresponding to the 5^{th} percentile (or 95^{th}, on tests where higher scores represented worse performance) was identified as the cut-off for impairment. If that raw score spanned more than one percentile rank, the nearest percentile rank with a raw score that uniquely divided the sample into impaired and unimpaired groups was used instead. Participants in the exposed groups were then classified as impaired according to that cut-off score, i.e. having a score that was equal to or worse than the lowest-performing 5% (or nearest feasible proportion) of the unexposed group. Since prospective memory was a categorical measure, impairment in all groups was defined as being incorrect on the first attempt.

5.2.5 Minimisation of bias

As described in Chapter 3, all assessments were administered according to a standard operating procedure, and administration and scoring of cognitive measures and questionnaires was automated. Bias in the ascertainment of the exposure groups was minimised by using the same sources of information (self-reported diagnoses and hospital records) for all exposures. Additional questionnaire-based data were, however, only available for the mania/BD and major depression exposures. Participants in Scotland were coded as missing for the mental health exposures, since their hospital records covered acute hospital admissions only; any mental health diagnoses in those records would be less likely to be in the primary position compared with participants in England and Wales, therefore possibly reflecting a different clinical presentation or comorbidity status in those participants. A single unexposed comparison group was used as the reference for all exposure groups, so that impairment prevalence ratios could be directly compared.

5.2.6 Data analysis

Statistical analyses were carried out using Stata software version 13 (StataCorp, 2013). Demographic, lifestyle and psychological measures were summarised descriptively to characterise the exposed and unexposed groups. Townsend deprivation index scores were categorised into quintiles based on frequency in the whole cohort. The reliability (internal consistency) of the cognitive tests in each group was examined using Cronbach's alpha where possible.

The prevalence of impairment on the five cognitive tests was calculated in each group, reported as a percentage with 95% CI based on the standard error calculated as $\sqrt{\frac{Prevalence \times (100 - Prevalence)}{n}}$. The ratio of prevalence in each exposed group versus the unexposed reference group was calculated, together with 95% CI, using the `epitab` functions in Stata v13. Crude results are reported, along with directly standardised results, which were computed using weights derived from the unexposed comparison group. In direct standardisation, sample sizes within each stratum of an external reference group (here, the unexposed comparison group) are used as weights to adjust the stratum-specific prevalence ratios in the exposed group: the prevalence ratio in each stratum of the exposed group is multiplied by the sample size of that stratum in the reference group, and the resulting products are then summed and divided by the total sample size of the reference group, to produce a standardised prevalence ratio for the whole exposed group. The purpose of standardisation was to control for demographic differences between the exposed and unexposed groups that might confound the crude results. Stratification was by age group (<60 versus ≥60 years; chosen to be close to the sample median of 58) and gender. Statistical interactions between exposure status and age group and between exposure status and gender were tested using robust Poisson regression models (Chen, Shi, Qian, & Azen, 2014) including a product term.

The population attributable prevalence of cognitive impairment (number of cases of cognitive impairment, per 100,000 total population, that are attributable to the exposure) was derived from 2x2 tables constructed for each exposure separately. The prevalence of each exposure within these tables was based on the lifetime estimates cited in section 5.1 above. The population attributable prevalence of cognitive impairment was calculated as

the total number of cases of cognitive impairment per 100,000 population multiplied by the population attributable fraction (PAF), where PAF was $\frac{Prevalence_{total} - Prevalence_{unexposed}}{Prevalence_{total}}$.

5.2.7 Sensitivity analyses

5.2.7.1 Comorbidity

Because the exposed groups may have had other comorbid psychiatric or neurological conditions in addition to the exposure of interest, which might increase the cognitive impairment prevalence relative to the unexposed comparison group, sensitivity analyses were conducted to examine the effect on the crude results of restricting the analyses to participants with no known comorbidities. Psychiatric and neurological comorbidities were identified from self-reported diagnoses and hospital records, as listed in Appendices P and Q.

5.2.7.2 Educational attainment

Sensitivity analyses were also conducted to examine the effect on the standardised results of accounting for educational attainment. Educational attainment is a potential confounder of cognitive performance that differed across the exposure groups in the present study. However, because the typical age of onset of each exposure of interest varies considerably (e.g. young adulthood in schizophrenia versus late middle age in PD), it may be that educational attainment is a consequence of exposure in the groups with younger onset, but a confounder in exposure groups with older onset. Stratifying on a variable that is a consequence of the exposure and outcome may cause 'collider bias', affecting the estimate of their association (Greenland et al., 1999). The age- and gender-stratified standardised analyses were therefore repeated with additional stratification on education (degree versus no degree), for comparison. Further to these analyses, missing data checks were also conducted: since there were some missing data on the education variable (but not on age or gender), a comparison was made between the crude results using all available data and the crude results in only those participants who had complete education data.

5.2.7.3 Information bias

It is possible that the additional information source (touchscreen mood questionnaire) by which participants could be classified as exposed for the mood disorders, in the absence of an equivalent information source for the other exposures of interest, might have led to

differential misclassification bias. Alternative versions of the mood disorder exposure groups were therefore constructed, only using data from the other two information sources (self-reported diagnoses and linked hospital records), and the analyses were repeated for comparison. Secondly, because hospital records data from Scotland were included when ascertaining the MS and PD groups but not the psychiatric exposures, the characteristics of the MS and PD groups identified via hospital records were examined including and excluding data from Scotland, for comparison. Lastly, because the proportion of missing data on the cognitive measures differed across the exposure groups, the characteristics of those with and without missing data were examined: age, gender, education and comorbidity status were compared within the broadly-defined exposure groups between participants who did and did not have missing data on each cognitive test.

5.3 Results

5.3.1 Characteristics of the exposed and unexposed groups

Table 5.1 shows the characteristics of each group on key demographic factors and the five cognitive measures. Additional group characteristics are provided in Appendix T. The total number in the cohort who did not meet the criteria for any of the exposed groups or the unexposed comparison group was 336,662; a large number were excluded from the unexposed group because they were recruited prior to the addition of the touchscreen mood questionnaire to the baseline assessment protocol. The proportion of missing data varied across measures, but tended to be higher in the exposed groups (particularly those with schizophrenia), compared with the unexposed comparison group. Cronbach's alpha is reported in Table 5.1 for the reasoning and reaction time tests in each group; it was not possible to calculate this for the other three cognitive tests, as they did not include multiple items. Alpha was higher for reaction time than for reasoning, but coefficients did not differ notably across groups.

Table 5.1 - Characteristics of the exposed and unexposed groups

	Unexposed comparison	Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow
n	104,410	3,020	607	56,425	7,583	850	259	1,905	931	916	323
Age (years)[a]											
Mean (SD)	57.0 (8.2)	54.9 (8.0)	54.8 (8.0)	55.6 (7.9)	55.1 (7.9)	53.9 (8.1)	52.6 (8.2)	55.4 (7.5)	55.3 (7.5)	62.2 (5.6)	62.5 (5.2)
Gender[a]											
n (%) female	52,210 (50.0)	1,589 (52.6)	345 (56.8)	36,574 (64.8)	4,923 (64.9)	293 (34.5)	74 (28.6)	1,400 (73.5)	665 (71.4)	346 (37.8)	131 (40.6)
Has a degree											
n (%) missing	1,033 (1.0)	40 (1.3)	4 (0.7)	582 (1.0)	59 (0.8)	31 (3.7)	10 (3.9)	39 (2.0)	26 (2.8)	20 (2.2)	6 (1.9)
n (%)[b]	36,082 (34.9)	1,117 (37.5)	250 (41.5)	18,188 (32.6)	2,344 (31.2)	212 (25.9)	65 (25.3)	642 (34.4)	278 (30.7)	276 (30.8)	88 (27.8)
No known comorbidity[a,c]											
n (%)	-	2,018 (66.8)	338 (55.7)	43,289 (76.7)	4,842 (63.9)	30 (3.5)	8 (3.1)	279 (14.7)	116 (12.5)	172 (18.8)	65 (20.1)
Reasoning score											
n (%) missing[d]	1,595 (1.6)	40 (2.1)	7 (2.2)	485 (1.4)	97 (1.6)	17 (5.8)	5 (5.5)	17 (2.9)	10 (3.6)	6 (2.1)	2 (1.7)
Mean (SD)	6.0 (2.2)	5.7 (2.2)	6.0 (2.4)	6.0 (2.1)	5.9 (2.2)	4.8 (2.1)	4.7 (2.0)	5.9 (2.0)	5.7 (2.0)	5.8 (2.2)	5.7 (2.2)
Cronbach's α	0.70	0.71	0.73	0.69	0.69	0.70	0.65	0.66	0.64	0.70	0.71
Reaction time (ms)											
n (%) missing	1,107 (1.1)	65 (2.2)	14 (2.3)	652 (1.2)	99 (1.3)	52 (6.1)	16 (6.2)	39 (2.1)	25 (2.7)	19 (2.1)	7 (2.2)
Mdn	543	558	571	543	551	601	583	590	594	571	578
(Q1, Q3)	(484, 621)	(492, 640)	(500, 664)	(484, 621)	(488, 633)	(520, 688)	(521, 668)	(512, 694)	(516, 716)	(512, 644)	(516, 653)
Cronbach's α	0.82	0.83	0.85	0.82	0.83	0.87	0.85	0.84	0.85	0.82	0.83
Numeric memory score											
n (%) missing[d]	749 (2.4)	22 (4.2)	2 (2.6)	301 (2.7)	59 (3.3)	9 (12.5)	5 (20.0)	9 (4.5)	8 (7.2)	3 (3.1)	2 (5.4)
Mean (SD)	6.7 (1.3)	6.5 (1.5)	6.5 (1.3)	6.7 (1.3)	6.6 (1.4)	5.9 (1.5)	5.9 (1.3)	6.6 (1.3)	6.4 (1.3)	6.5 (1.4)	6.4 (1.5)
Visuospatial memory errors											
n (%) missing	0 (0.0)	14 (0.5)	1 (0.2)	117 (0.2)	12 (0.2)	17 (2.0)	3 (1.2)	18 (0.9)	11 (1.2)	9 (1.0)	2 (0.6)
Mdn (Q1, Q3)	3 (2, 5)	4 (2, 6)	4 (2, 7)	3 (2, 6)	4 (2, 6)	4 (2, 7)	4 (2, 6)	3 (2, 6)	3 (2, 6)	4 (2, 6)	4 (2, 7)
Prospective memory[a,d]											
n (%)[a] correct	80,192 (77.2)	1,391 (71.0)	219 (65.4)	28,148 (77.7)	4,724 (75.4)	180 (55.4)	58 (56.3)	446 (73.4)	206 (70.8)	217 (70.9)	79 (63.7)

Mdn, median; Q, quartile; SD, standard deviation.
a. No missing data.

b. Missing excluded from denominator.
c. Not known to have any other psychiatric or neurological condition in addition to the exposure. By definition, no member of the unexposed comparison group had any primary or comorbid psychiatric or neurological condition.
d. Missing data refers only to the period when this measure was included in the battery.

5.3.2 Prevalence of cognitive impairment across groups

On the reaction time, numeric memory and visuospatial memory tests, the cognitive impairment threshold corresponded to the worst-performing 5% of the unexposed group. Owing to the restricted raw score range on the reasoning test, the same raw score spanned the 5th to 11th percentile range and so the 4th percentile score was instead used to divide the sample; the impairment threshold therefore corresponded to the worst-performing 4% of the unexposed group. Prospective memory was a pass/fail test; the proportion of the unexposed group with an incorrect score was 22.8%.

Table 5.2 shows the prevalence of impairment in each exposed group, along with prevalence ratios relative to the unexposed group. Standardised estimates are not reported for some groups, because of insufficient data in some strata.

The crude estimates (Table 5.2) indicated that cognitive impairment prevalence was higher in mania/BD than in the unexposed comparison group, with prevalence ratios in the broadly-defined group ranging from 1.27 (95% CI 1.18, 1.36) to 1.97 (95% CI 1.52, 2.56). Prevalence in major depression was closer to the comparison group level, with crude ratios in the broadly-defined group ranging from 0.98 (95% CI 0.96, 1.00) to 1.09 (95% CI 1.04, 1.14). Crude impairment prevalence was higher in MS and PD groups than in the comparison group for reaction time (2.87 [95% CI 2.56, 3.22] and 1.34 [95% CI 1.05, 1.72] respectively in the broadly-defined groups), visuospatial memory (1.42 [95% CI 1.19, 1.69] and 1.89 [95% CI 1.52, 2.35]) and prospective memory (1.17 [95% CI 1.02, 1.33] and 1.27 [95% CI 1.07, 1.52]). Crude impairment prevalence was highest in schizophrenia on all tests (prevalence ratios in the broadly-defined group ranged from 1.95 [95% CI 1.73, 2.21] to 3.95 [95% CI 2.43, 6.43]) except reaction time, for which prevalence was highest in MS.

Table 5.2 - Prevalence of cognitive impairment across groups

Impairment threshold		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**
Reasoning	*n*	1,866	318	35,211	6,052	274	86	572	272	283	115
≤ unexposed 4th	Crude P %	7.13	6.29	4.22	5.24	10.95	6.98	2.80	1.84	4.95	6.96
percentile score	95% CI	5.96, 8.30	3.62, 8.96	4.01, 4.43	4.68, 5.80	7.25, 14.65	1.59, 12.37	1.45, 4.15	0.24, 3.44	2.42, 7.48	2.31, 11.61
(Unexposed prevalence	Crude PR	1.71*	1.51	1.02	1.26*	2.63*	1.68	0.67	0.44	1.19	1.67
4.16%)	95% CI	1.45, 2.02	0.99, 2.31	0.96, 1.08	1.13, 1.41	1.88, 3.70	0.78, 3.63	0.41, 1.09	0.19, 1.05	0.71, 1.98	0.86, 3.27
	Standardised P %	7.20	6.49	4.33	5.41	10.61	-[b]	3.20	-[b]	5.91	9.24
	95% CI	6.02, 8.37	3.78, 9.20	4.11, 4.54	4.84, 5.98	6.96, 14.25	-	1.76, 4.65	-	3.16, 8.65	3.94, 14.53
	Standardised PR	1.73*	1.56*	1.04[a]	1.30*	2.55*	-	0.77	-	1.42	2.22*
	95% CI	1.45, 2.05	1.01, 2.42	0.98, 1.10	1.16, 1.47	1.75, 3.71	-	0.43, 1.37	-	0.70, 2.89	1.02, 4.83
Reaction time	*n*	2,955	593	55,773	7,484	798	243	1,866	906	897	316
> unexposed 95th	Crude P %	6.73	7.59	5.31	6.55	12.91	10.70	14.31	17.77	6.69	7.28
percentile score	95% CI	5.83, 7.63	5.46, 9.72	5.12, 5.50	5.99, 7.11	10.58, 15.24	6.81, 14.59	12.72, 15.90	15.28, 20.26	5.05, 8.33	4.42, 10.14
(Unexposed prevalence	Crude PR	1.35*	1.52*	1.07*	1.32*	2.59*	2.15*	2.87*	3.57*	1.34*	1.46
4.98%)	95% CI	1.18, 1.55	1.15, 2.02	1.02, 1.11	1.20, 1.44	2.16, 3.11	1.49, 3.09	2.56, 3.22	3.10, 4.12	1.05, 1.72	0.99, 2.17
	Standardised P %	7.27	8.12	5.58	7.02	13.70	12.10	15.19	18.82	6.27	5.68
	95% CI	6.33, 8.21	5.92, 10.32	5.39, 5.77	6.44, 7.60	11.31, 16.08	8.00, 16.20	13.56, 16.82	16.28, 21.37	4.69, 7.86	3.13, 8.23
	Standardised PR	1.46*	1.63*	1.12*[c]	1.41*[d]	2.75*	2.43*	3.05*[e]	3.78*[f]	1.26	1.14
	95% CI	1.27, 1.68	1.22, 2.19	1.07, 1.17	1.29, 1.55	2.25, 3.37	1.60, 3.69	2.68, 3.48	3.22, 4.43	0.92, 1.71	0.71, 1.83
Numeric memory	*n*	505	75	10,808	1,746	63	20	193	103	94	35
≤ unexposed 5th	Crude P %	10.30	5.33	5.65	7.67	20.63	15.00	6.74	8.74	8.51	14.29
percentile score	95% CI	7.65, 12.95	0.25, 10.41	5.21, 6.09	6.42, 8.92	10.64, 30.62	0.00, 30.65	3.20, 10.28	3.29, 14.19	2.87, 14.15	2.70, 25.88
(Unexposed prevalence	Crude PR	1.97*	1.02	1.08	1.47*	3.95*	2.87*	1.29	1.67	1.63	2.74*
5.22%)	95% CI	1.52, 2.56	0.39, 2.65	0.99, 1.19	1.24, 1.74	2.43, 6.43	1.01, 8.17	0.76, 2.19	0.90, 3.13	0.84, 3.17	1.21, 6.17
	Standardised P %	10.39	-[b]	5.74	7.78	20.46	-[b]	7.41	-[b]	12.68	18.69
	95% CI	7.73, 13.05	-	5.30, 6.18	6.52, 9.03	10.50, 30.42	-	3.72, 11.11	-	5.96, 19.41	5.77, 31.60
	Standardised PR	1.99*	-	1.10	1.49*	3.92*	-	1.42	-	2.43*[g]	3.58*[h]
	95% CI	1.51, 2.62	-	0.99, 1.21	1.24, 1.79	2.34, 6.57	-	0.72, 2.82	-	1.19, 4.97	1.60, 8.03
Visuospatial memory	*n*	3,006	606	56,308	7,571	833	256	1,887	920	907	321
> unexposed 95th	Crude P %	6.35	8.58	4.76	5.15	9.12	7.81	6.20	7.28	8.27	8.10
percentile score	95% CI	5.48, 7.22	6.35, 10.81	4.58, 4.94	4.65, 5.65	7.16, 11.08	4.52, 11.10	5.11, 7.29	5.60, 8.96	6.48, 10.06	5.12, 11.08
(Unexposed prevalence	Crude PR	1.45*	1.96*	1.09*	1.18*	2.08*	1.78*	1.42*	1.66*	1.89*	1.85*
4.38%)	95% CI	1.26, 1.67	1.51, 2.55	1.04, 1.14	1.06, 1.30	1.68, 2.59	1.17, 2.72	1.19, 1.69	1.32, 2.10	1.52, 2.35	1.28, 2.68

Impairment threshold		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**
	Standardised P %	6.57	8.98	5.12	5.65	9.68	8.58	6.57	8.02	6.66	7.36
	95% CI	5.68, 7.46	6.70, 11.26	4.94, 5.31	5.13, 6.17	7.67, 11.69	5.15, 12.02	5.45, 7.69	6.26, 9.77	5.04, 8.28	4.50, 10.21
	Standardised PR	1.50*	2.05*	1.17*	1.29*	2.21*	1.96*	1.50*	1.83*	1.52*	1.68*
	95% CI	1.30, 1.73	1.57, 2.69	1.12, 1.23	1.16, 1.44	1.75, 2.79	1.20, 3.18	1.22, 1.84	1.41, 2.37	1.16, 1.99	1.05, 2.68
Prospective memory	*n*	1,959	335	36,237	6,267	325	103	608	291	306	124
Incorrect score	Crude P %	28.99	34.63	22.32	24.62	44.62	43.69	26.64	29.21	29.08	36.29
(Unexposed prevalence	95% CI	26.98, 31.00	29.53, 39.73	21.89, 22.75	23.55, 25.69	39.22, 50.02	34.11, 53.27	23.13, 30.15	23.99, 34.43	23.99, 34.17	27.83, 44.75
22.82%)	Crude PR	1.27*	1.52*	0.98	1.08*	1.95*	1.91*	1.17*	1.28*	1.27*	1.59*
	95% CI	1.18, 1.36	1.31, 1.76	0.96, 1.00	1.03, 1.13	1.73, 2.21	1.54, 2.38	1.02, 1.33	1.07, 1.53	1.07, 1.52	1.26, 2.01
	Standardised P %	29.67	35.37	22.82	25.56	45.87	43.13	26.24	29.67	27.84	34.46
	95% CI	27.64, 31.69	30.25, 40.49	22.39, 23.25	24.48, 26.64	40.45, 51.29	33.57, 52.69	22.75, 29.74	24.42, 34.91	22.82, 32.86	26.09, 42.82
	Standardised PR	1.30*[i]	1.55*	1.00[j]	1.12*[k]	2.01*	1.89*	1.15	1.30*	1.22	1.51*
	95% CI	1.20, 1.39	1.34, 1.80	0.98, 1.02	1.07, 1.17	1.76, 2.28	1.47, 2.42	0.99, 1.34	1.07, 1.59	0.98, 1.54	1.13, 2.01

CI, confidence interval; P, prevalence; PR, prevalence ratio.

Standardised estimates are directly standardised by age and gender with reference to the unexposed comparison group.

* $p < 0.05$

a. Significant interaction with gender: women PR = 0.95 (CI 0.88, 1.02); men PR = 1.14 (CI 1.04, 1.25).

b. Estimates not reported because at least 1 of 4 strata contained no exposed participants with cognitive impairment.

c. Significant interaction with gender: women PR = 1.05 (CI 0.99, 1.11); men PR = 1.21 (CI 1.13, 1.30).

d. Significant interaction with gender: women PR = 1.29 (CI 1.15, 1.45); men PR = 1.56 (CI 1.34, 1.82).

e. Significant interaction with age and gender: <60 years PR = 4.31 (CI 3.65, 5.08); ≥60 years PR = 2.38 (CI 1.94, 2.90); women PR = 2.55 (CI 2.21, 2.93); men PR = 3.69 (CI 2.99, 4.55).

f. Significant interaction with age and gender: <60 years PR = 5.85 (CI 4.84, 7.08); ≥60 years PR = 2.66 (CI 2.04, 3.47); women PR = 3.00 (CI 2.50, 3.60); men PR = 4.76 (CI 3.72, 6.08).

g. Significant interaction with age: <60 years PR = 4.85 (CI 2.11, 11.17); ≥60 years PR = 0.72 (CI 0.24, 2.19).

h. Significant interaction with age: <60 years PR = 6.66 (CI 2.54, 17.50); ≥60 years PR = 1.40 (CI 0.36, 5.51).

i. Significant interaction with age: <60 years PR = 1.43 (CI 1.31, 1.56); ≥60 years PR = 1.19 (CI 1.06, 1.33).

j. Significant interaction with gender: women PR = 0.96 (CI 0.93, 0.99); men PR = 1.04 (CI 1.00, 1.08).

k. Significant interaction with gender: women PR = 1.06 (CI 0.99, 1.12); men PR = 1.19 (CI 1.10, 1.28).

Following direct standardisation with reference to the unexposed group weights for age group and gender, estimates increased in the mania/BD and major depression groups on all measures (Table 5.2). Some estimates increased and some decreased in the schizophrenia groups, although the general pattern was similar to the crude results. In the MS groups, all estimates but one increased, and in PD, reasoning and numeric memory estimates increased but other estimates decreased. Within the mania/BD, MS and PD groups, interaction tests indicated that impairment prevalence was significantly lower in the older age group on some measures. Significant interactions with gender were found in major depression and MS on some measures, showing lower impairment prevalence in women. Figure 5.1 shows the standardised absolute prevalence and Figure 5.2 shows the standardised prevalence ratios on all measures in the broadly-defined exposure groups.

When the highest standardised impairment prevalence estimates from Table 5.2 were applied to population prevalence estimates for each exposure (cited in section 5.1), the population attributable lifetime prevalence of cognitive impairment per 100,000 population was approximately 256 (95% CI 130, 381) for major depression, 151 (95% CI 52, 251) for schizophrenia, 45 (95% CI 23, 68) for mania/BD, 27 (95% CI 22, 32) for MS and 26 (95% CI 1, 52) for PD (Appendix U).

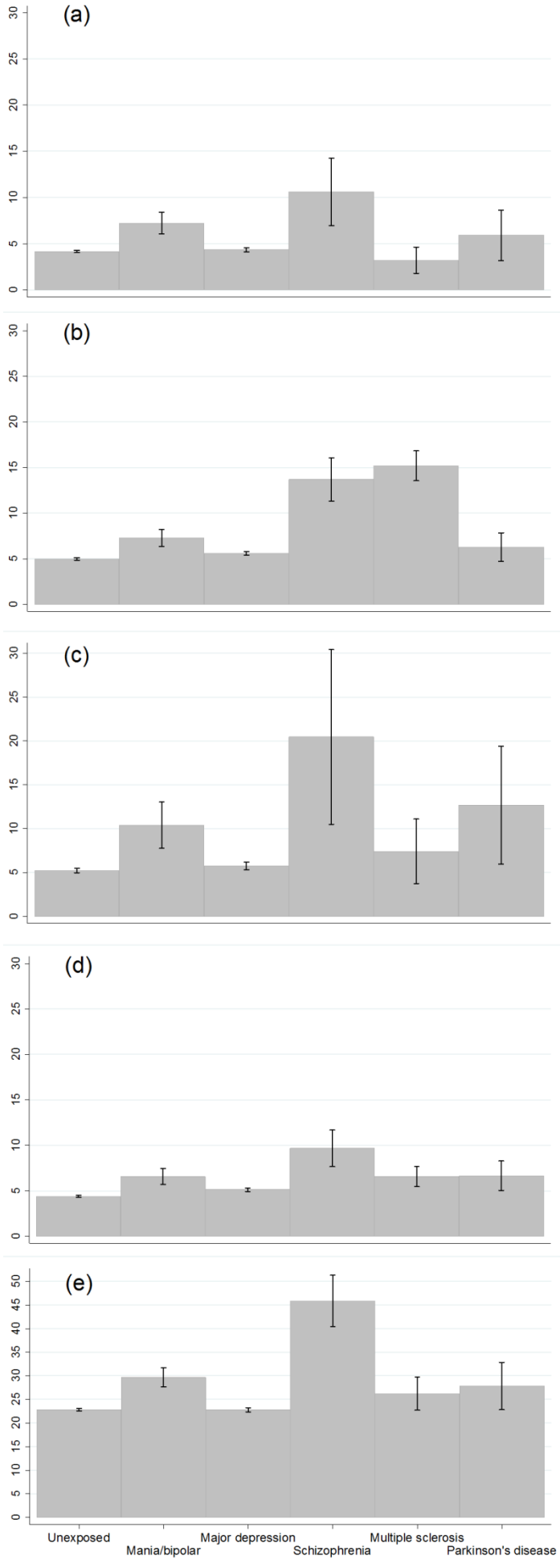

(a)
(b)
(c)
(d)
(e)
Unexposed
Mania/bipolar
Major depression
Schizophrenia
Multiple sclerosis
Parkinson's disease

Figure 5.1 - Standardised prevalence estimates for cognitive impairment

Estimates are prevalence (%), directly standardised by age and gender with reference to the unexposed comparison group. Point estimates and 95% confidence intervals are shown. Exposure groups are broadly defined (classified as exposed by at least one ascertainment method). Panels show: (a) Reasoning; (b) Reaction time; (c) Numeric memory; (d) Visuospatial memory; (e) Prospective memory.

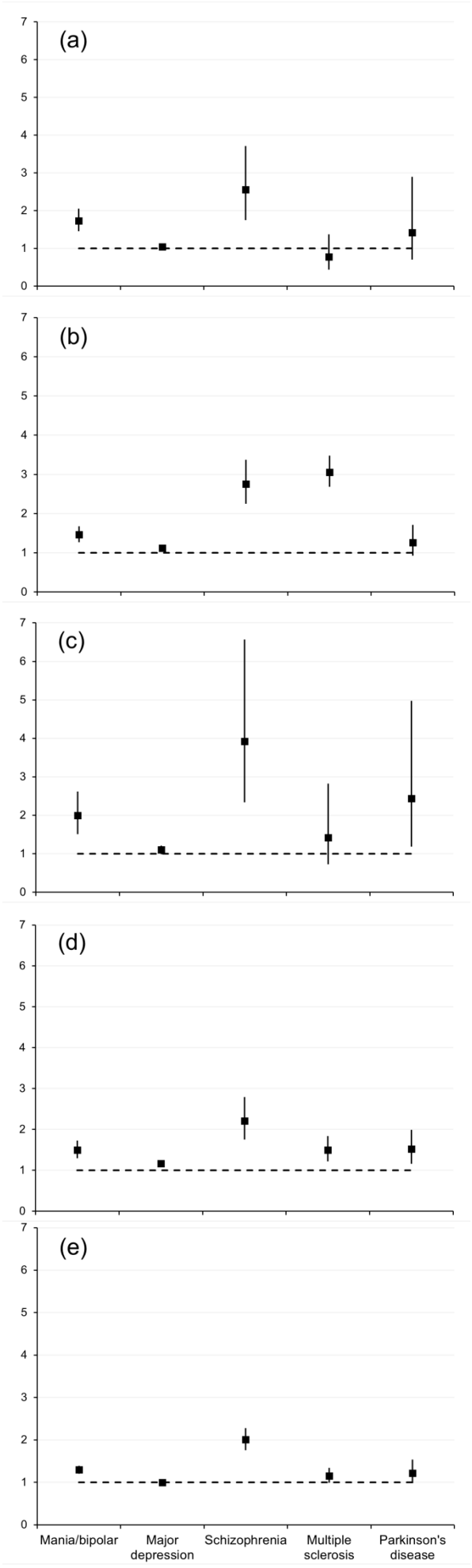

(a)
(b)
(c)
(d)
(e)
0
1
2
3
4
5
6
7
Mania/bipolar
Major depression
Schizophrenia
Multiple sclerosis
Parkinson's disease

Figure 5.2 - Standardised prevalence ratios for cognitive impairment

Estimates are prevalence ratios compared with the unexposed group, directly standardised by age and gender. Point estimates and 95% confidence intervals are shown. Dashed line represents the null (no difference relative to unexposed group). Exposure groups are broadly defined (classified as exposed by at least one ascertainment method). Panels show: (a) Reasoning; (b) Reaction time; (c) Numeric memory; (d) Visuospatial memory; (e) Prospective memory.

5.3.3 Sensitivity analyses

5.3.3.1 Comorbidity

When the crude estimates were re-calculated in participants with no known comorbidities, the prevalence ratios attenuated towards the null (see Table V.1 in Appendix V). In all the groups except major depression, the estimates generally remained higher compared with the unexposed group, although very small sample sizes in some groups (e.g. schizophrenia and PD) reduced the statistical power or prevented valid analysis. A somewhat different pattern was seen in the major depression group: some estimates in the broadly-defined group reversed such that impairment prevalence became significantly lower compared with the unexposed group, and most estimates in the narrowly-defined group were no longer significantly different from the unexposed group.

5.3.3.2 Educational attainment

Table V.2 in Appendix V shows standardised results taking account of educational attainment as well as age group and gender. Results are not reported for some groups owing to sparse data across strata. Compared with results standardised for age and gender only (Table 5.2), some estimates were higher and some were lower in all exposed groups. The magnitude of difference in either direction was generally small, and the overall pattern of results remained similar. Educational attainment was lower in the schizophrenia and PD groups compared with the unexposed group, but there was no clear indication of collider bias in the former: inclusion of education in the standardised analyses reduced rather than inflated the prevalence estimates on most measures. Comparison of crude results between the full sample (Table 5.2) and those with complete education data (Table V.3 in Appendix V) did not indicate an effect of missing data.

5.3.3.3 Information bias

The analyses were also repeated using alternative versions of the mania/BD and major depression groups, formed without reference to the touchscreen mood disorders questionnaire data. The results showed that almost all estimates were higher in these alternative groups, and all were significantly higher than in the unexposed group. The crude and standardised results for these groups are presented in Table V.4 in Appendix V.

The characteristics of the MS and PD groups identified via hospital records were very similar regardless of whether data from Scotland were included (Table V.5 in Appendix V).

Comparison of participant characteristics between those with and without missing data on the cognitive measures showed that participants with missing data across each exposed group were generally older, less likely to have a degree, and more likely to have comorbidities (Tables V.6a to V.6d in Appendix V; no participants had missing data on the prospective memory test). Unexposed comparison participants with missing cognitive data were also older and less likely to have a degree. Men were over-represented in the missing data groups for major depression, schizophrenia and MS, but were under-represented in the missing data group for mania/BD. The missing data mechanism was likely to be 'not-at-random', assuming that participants with worse cognitive function would be more likely to discontinue the tests. The impairment prevalence estimates reported in this chapter are therefore likely to be biased downward, compared with true prevalence in the UK Biobank cohort and in the general population.

5.4 Discussion

In this community-based population in middle to early old age, standardised prevalence of cognitive impairment was higher in people with a history of psychiatric or neurological conditions than in those with no such history. Across the exposure groups studied, standardised prevalence was highest on most measures in participants with schizophrenia and lowest in those with major depression. Mania/BD was the second most impaired exposure group on three of the five measures (reasoning, numeric memory and prospective memory), and sensitivity analyses showed that impairment in this group was higher still when exposure information sources were strictly equivalent with the other groups. Reaction time impairment was most prevalent in the MS group, which is in line with previous research showing particular problems with processing speed in MS and other white matter disorders (Prins et al., 2005; Rao, Leo, Bernardin, et al., 1991; Rao et al., 2014). Although the increased prevalence (compared with the unexposed group) of cognitive impairment in major depression was relatively small, lifetime prevalence of major depression is approximately ten times that of BD and schizophrenia and fifty times that of MS and PD, which meant that the population attributable prevalence of cognitive impairment was highest overall for this group. Sensitivity analyses suggested that

comorbidities may be contributing to the increased likelihood of cognitive impairment in major depression.

This is the first study to directly compare prevalence of cognitive impairment across these conditions, using consistent assessment methods and a single unexposed comparison group. Multiple sources of information were used to classify exposure status, and impairment status was defined with reference to a very large normative group. Direct standardisation permitted like-for-like comparisons across the exposure groups, and sensitivity analyses were conducted to examine possible sources of bias and confounding. The overall pattern of findings was consistent regardless of exposure group definitions (broad or narrow) and adjustment for key demographic characteristics.

The prevalence of cognitive impairment in all groups was lower than expected. This may indicate that people living with psychiatric and neurological conditions in the general population are less impaired than the patient population represented in clinical studies. Alternatively, the low prevalence of impairment may reflect selection bias in UK Biobank, such that invitees may have been more motivated to join a medical research study if they had prior experience of health problems, and may have been more willing or able to take part if they had better cognitive function and/or less severe disorder. This represents collider bias (Greenland et al., 1999), whereby exposed status and (lower) probability of impaired outcome together influenced study participation. This would introduce bias, by attenuating or reversing any true positive association of exposure with impaired outcome. Similarly, the greater proportion of missing cognitive data in the exposed groups, and the finding that missingness was itself associated with older age, lower educational attainment and comorbidities, suggests that selection into the analysed sample was biased towards more cognitively able participants.

The prevalence of impairment was higher when the mood disorder exposure groups were based on self-reported doctor diagnosis and/or hospital records, without reference to the touchscreen questionnaire data on lifetime mood disorder experiences (Table V.4); the results in Table 5.2 can therefore be taken as a lower bound of true prevalence. It may be that the questionnaire data misclassified participants as exposed (differentially so for less cognitively impaired participants), or, alternatively, there might be a large number of people living with undiagnosed mood disorders in the general population, whose true prevalence of cognitive impairment is lower than previous studies have suggested.

With regard to demographic factors, the finding that cognitive impairment was less common in women compared with men in the MS group on some measures may reflect generally less severe disease course in women (Bergamaschi, 2007). A similar finding for women in the major depression group was unexpected in light of a recent meta-analysis, which found no such association (Bora et al., 2013). Impairment was also less common in older compared with younger participants within the mania/BD, MS and PD groups. Given the consistent association between older age and greater cognitive impairment in the general population, this finding is likely to indicate survivor bias in these exposed groups.

A number of study limitations need to be considered. Unlike previous studies in these conditions, clinician-confirmed diagnoses were not available, and linked health records covered in-patient and day case admissions only. Psychiatric hospital data were unavailable for participants in Scotland, although these comprised only 7% of the whole cohort. Information regarding exposure status relied substantially on self-reported diagnoses or responses to questionnaire items. Nevertheless, descriptive data regarding sociodemographic factors, psychological measures and medication use supported the distinctions between the groups. The cross-sectional nature of the study and the limitations of the available clinical information also meant that the onset times and durations of exposures and outcomes were not known; exposure-outcome associations may not be causal, and it is possible that cognitive impairment may precede clinical onset of some disorders (e.g. schizophrenia). The UK Biobank cohort is not representative of the UK population in many respects, and the exposure groups that were identified within it are likely to differ from psychiatric and neurological samples in other studies and in clinical practice, with regard to sociodemographic characteristics, illness severity and motivational factors. It is not known whether the degree of non-representativeness differed across the exposure groups, however. If it did not, then between-group comparisons remain valid. Within the mood disorder groups, it was not possible to distinguish reliably between subtypes of bipolar presentations and single versus recurrent depression, thus limiting the comparability of the findings to previous clinical studies. Current depressive symptoms were reported, but status with regard to clinical euthymia was not known. The population attributable prevalence calculations were based on the standardised results presented in Table 5.2, in which comorbidities were permitted in the exposed groups but not in the unexposed comparison group; this does not reflect the composition of the general population, and the attributable estimates would only be considered valid if the results reflected a causal effect of the exposures on cognitive outcome.

The limitations of the cognitive tests were discussed in Chapter 3. It should also be noted that the impairment threshold in the present analyses differed slightly across the five tests, because of variation in the reference score distributions. It is possible that the relatively low impairment prevalence on these cognitive tests compared with previous clinical studies reflects insensitivity of the brief measures, or differential reliability across exposure groups, but variation in performance across groups was detectable whilst internal consistency was similar. Despite the very large size of the cohort, sample sizes were modest on some cognitive measures and data were sparse across strata in the narrowly-defined exposure groups, which limited the standardised analyses. No information was available regarding the impact of cognitive impairment on instrumental functioning: it is possible that the impact of impairment on disability and participation is not well captured by the 5th percentile impairment threshold used here, given that the range of severity below this threshold may differ across disorders, and other people will experience instrumental dysfunction even when measured cognitive performance remains above the 5th percentile threshold.

5.5 Conclusion

In directly comparative analyses in this large general population cohort, cognitive impairment was most common in participants with schizophrenia and least common in those with major depression, although the much higher population prevalence of major depression means that the overall burden of cognitive impairment attributable to this disorder is likely to be considerable. Cognitive impairment in mania/BD has previously received less research and clinical attention than impairment in schizophrenia and neurological disorders, but direct comparisons in the present study indicated that impairment prevalence in mania/BD was similar to that in MS and PD, both of which are much less common in the population. Study limitations included self-reported data and selection bias.

5.6 Next steps

The analyses presented in this chapter aimed to quantify the prevalence of cognitive impairment in order to permit direct comparisons across disorders of interest. Although potential confounding influences of age, gender and education were taken into account, more complex analyses are required to investigate the multitude of other social, clinical, lifestyle and genetic factors that are likely to contribute to cognitive risk and resilience in

these disorders. Understanding such variation is the second key aim of this thesis. Chapter 6 describes graphical and statistical methods for causal inference in the context of observational studies. Chapters 7 and 8 report the results of multivariate models to explain the effect of psychiatric exposures on cognitive impairment, and the role of intermediate risk factors.

Chapter 6 Causal inference with observational data

The axiom that correlation does not imply causation is ingrained in scientific research. Most prudent researchers in the biomedical and social sciences are wary of making causal claims, especially in observational studies. Although this caution is well-founded, it illustrates the mismatch between a key rationale or motivation for conducting research—to explain phenomena—and the reality of the reported findings, which are often more descriptive than explanatory. Dissatisfaction with this mismatch, along with advances in statistical methods and computing power, have stimulated a different approach to data analysis that makes causal inference an explicit goal. Proponents of this approach encourage researchers to stop "retreating into the associational haven" and instead "to take the causal bull by the horns" (Hernán, 2005, p. 620). This chapter summarises the causal definitions on which this framework rests, the reasons why statistical associations arise in data, and the circumstances under which such associations may justify a causal interpretation. Methods for identifying and estimating total causal effects are outlined, together with approaches for effect decomposition within the causal framework, and threats to internal and external validity are considered. Although the causal inference framework has been applied to a variety of study designs including randomised controlled trials (e.g. Emsley, Dunn, & White, 2010; Kaufman, Kaufman, & Poole, 2003) and longitudinal repeated measures designs (e.g. Daniel, Cousens, De Stavola, Kenward, & Sterne, 2013), this chapter focuses primarily on causal inference in cross-sectional observational research.

6.1 Conceptualisations of causality

Many different conceptualisations and definitions of causality have been proposed by thinkers and researchers in a diverse range of fields, from philosophy to artificial intelligence. The 'sufficient-component' model (Rothman, 1976) has gained widespread acceptance in epidemiology, and underpins other key epidemiological concepts such as interaction and attributable risk. Under this model, a sufficient cause is defined as "a minimal set of conditions and events that are sufficient for the outcome to occur" (Rothman, Greenland, Poole, & Lash, 2008, p. 6). This set of conditions or events comprises multiple components, some of which are unknown, and the outcome will not occur unless all the components are present. Conversely, blocking or removing one or

more components will prevent the sufficient cause from acting to produce the outcome. For any given outcome, many different sufficient causes can be conceptualised, any one of which will produce the outcome in a particular individual. If there is a component that is part of all such sufficient causes, that component is considered a (universally) necessary cause.

A key characteristic of all causal components is that they are contrasted with a specified alternative or referent. Tobacco smoking is considered a component cause of lung cancer, for example. Smoking might be specified as smoking 20 cigarettes per day for 40 years, and the specified referent would differ from this with regard to the number of cigarettes (or other tobacco product), and/or the duration of smoking. This type of contrast is intrinsic to the definition of causality that underpins the sufficient-component model: "a cause of a disease occurrence is an event, condition, or characteristic that preceded the disease onset and that, *had the event, condition, or characteristic been different in a specified way*, the disease either *would not have* occurred at all or *would not have* occurred until some later time" (Rothman et al., 2008, p. 6, italics added). The italicised phrases show that this definition of causality rests on counterfactuals, i.e. conditional statements containing antecedents (if-clauses) that are contrary to fact. The counterfactual view of causality has a long history in experimental research, statistics and economics, and within epidemiology it is primarily identified with the 'potential outcomes' framework of Rubin (1974), influenced by earlier research by Neyman (1990[1923]). Researchers working in this framework express causality in terms of potential outcomes under contrasting conditions. Using potential outcomes notation, where X is a binary exposure (or treatment) and Y is an outcome, $Y(1)$ denotes the outcome if X were set (possibly contrary to fact) to 1, and $Y(0)$ denotes the outcome if X were set (possibly contrary to fact) to 0. If the expected values ($E[Y]$) of these potential outcomes in the population are not equal, i.e.

$$E[Y(1)] \neq E[Y(0)]$$

then X is said to have a causal effect on Y.

An alternative notation system to express causal relationships is through graphs. This approach is primarily identified with Pearl (Pearl, 2009; Pearl et al., 2016). Pearl and colleagues defined causality informally as follows: "A variable X is a cause of a variable Y if Y in any way relies on X for its value…[We] think of causation as a form of listening; X

is a cause of *Y* if *Y* listens to *X* and decides its value in response to what it hears" (Pearl et al., 2016, p. 5). This simple situation is depicted graphically in Figure 6.1 below.

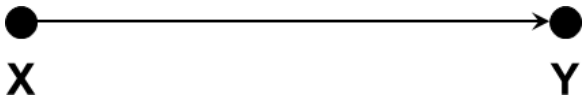


Figure 6.1 - Directed acyclic graph showing a causal effect of X on Y

Figure 6.1 is a directed acyclic graph (DAG), which encodes a structural causal model (SCM). Box 6.1 (overleaf) briefly summarises rules for the construction and interpretation of DAGs and SCMs. Despite the notational differences between graphical/SCM and potential outcomes representations of causal effects, these two frameworks have been shown to be logically equivalent (Pearl, 2009, Ch. 7).

6.2 Sources of statistical association in data

Aside from the play of chance, there are three key reasons why associations are observed between variables: causation, confounding, and collider conditioning (Elwert, 2013). These can be shown graphically in DAGs, thus demonstrating the correspondence between the causal relations depicted in the DAG and the probabilistic relations observed in the data.

6.2.1 Causation

As noted above, causation means that one variable relies for its value on another. This will induce a statistical association between measurements of those variables. Interpreting the direction of causation from an observed association requires background knowledge, for example regarding temporal order. Causal relations between variables are depicted graphically as chains (Figure 6.2 below). Variables at either end of a causal chain are likely to be unconditionally associated[6] with each other, assuming (for example) that positive and negative causal effects at intermediate points along the chain do not exactly cancel each other out. If *X* and *Y* are two variables in a causal chain, the association between them can be broken by conditioning on an intermediate variable or set of variables, *M*. This means that *X* and *Y* are conditionally independent given *M* (provided that there is no other open path between *X* and *Y* except through *M*).

[6] i.e. marginally associated, in a crude analysis

Box 6.1 - Directed acyclic graphs

Directed acyclic graphs (DAG) are graphical tools that are employed by researchers to visually represent qualitative causal assumptions (Elwert, 2013). The structural nature of these assumptions permits the detection of implied patterns of dependency and independency among variables, which can then be tested with data. The structural relations encoded by the DAG are also used to detect whether causal effects of interest (e.g. the total causal effect of one variable on another) are identified, with or without conditioning on other variables.

Each node in a DAG represents a variable, and each single-headed arrow represents a putative causal effect flowing from the node that emits the arrow (the parent) to the node that receives it (the child). The presence of an arrow represents the weak assumption of a causal relationship for at least one member of the population; the absence of an arrow between two nodes represents the strong assumption of no causal relationship for any member of the population. A node that only emits arrows is exogenous, and a node that receives any arrow is endogenous. A sequence of nodes connected by arrows (regardless of the direction of the arrowheads) is a path. If the sequence can be traced in one direction along the arrowheads, then the path is a directed path. Cyclical or reciprocal paths are not permitted in a DAG, but the same variable at different points in time can be shown as separate nodes.

For a DAG to be considered a causal DAG, every common parent of every pair of nodes must be depicted, whether they are measured in the dataset or not. Unmeasured variables can be depicted as hollow or light-coloured nodes, as shown in Figure B.1 below.

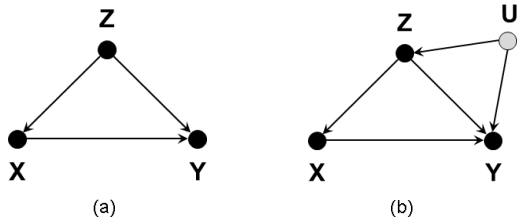


Figure B.1 Panel (a) shows a DAG with one exogenous variable (Z) and two endogenous variables (X and Y). X is caused by Z, and Y is caused by X and Z. Panel (b) adds an exogenous node U, which is unmeasured. An alternative notation is to replace node U and its emitted arrows with a dashed bi-directed arc ↖ ↗ between Z and Y.

DAGs visually encode structural causal models (SCM), which provide a general framework for describing causal relations among variables. An SCM comprises a set of exogenous and endogenous variables, and a set of functions *f* that assigns each endogenous variable a value based on the values of the other variables in the model (Pearl et al., 2016). In this way, causality under the SCM is defined as follows: "A variable *X* is a direct cause of a variable *Y* if *X* appears in the function that assigns *Y*'s value. *X* is a cause of *Y* if it is a direct cause of *Y*, or any cause of *Y*" (Pearl et al., 2016, p. 26). Figure B.2 below shows a DAG whose corresponding functions can be expressed in the most general case as:

$$z = f_Z(u_Z) \qquad x = f_X(z, u_X) \qquad y = f_Y(x, z, u_Y)$$

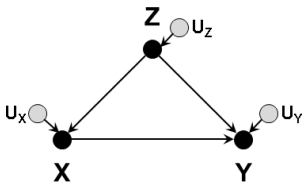


Figure B.2 This DAG is shown 'under magnification' to include independent unmeasured nodes, representing unknown or random contributors to the observed variables. These unmeasured factors are assumed to be independent of each other and of all other nodes in the graph, and so are usually not shown for reasons of clarity. Only factors that influence two or more nodes must be depicted in a causal DAG.

SCMs are non-parametric, which means that no assumptions are made about the distribution of the variables (e.g. continuous, discrete) or the functional form of the relationships between them (linear or non-linear). The possibility of interaction or effect modification is therefore accommodated within DAGs as a matter of course, without requiring additional graphical notation.

Researchers who are accustomed to linear structural equation models (SEM) will note the visual similarity between DAGs and path diagrams used in SEM; DAGs can in fact be interpreted as non-parametric structural equation models (NPSEM; Elwert, 2013), of which linear SEM is a special case.

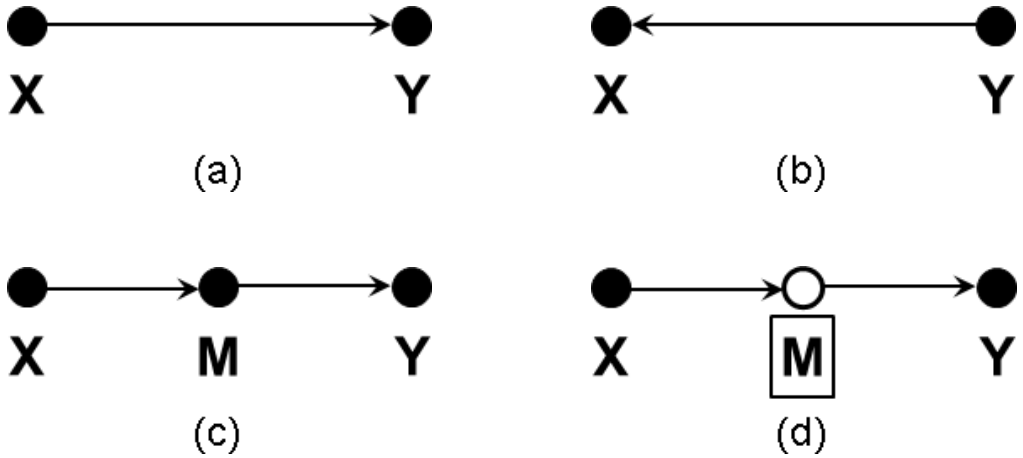


Figure 6.2 - Causal chains

In DAGs (a) to (c), X and Y are likely to be unconditionally associated. In DAG (d), X and Y are independent, conditional on M.

6.2.2 Confounding

Two variables may be unconditionally associated in a dataset if they share a common cause. This spurious association is the result of confounding bias, shown graphically by forks (Figure 6.3 below). Conditioning on the common cause (via statistical adjustment, stratification or sample restriction) will block the portion of the observed association that is spurious. If there is no other path between two variables *X* and *Y* apart from via common cause *C*, then *X* and *Y* are independent conditional on *C*.

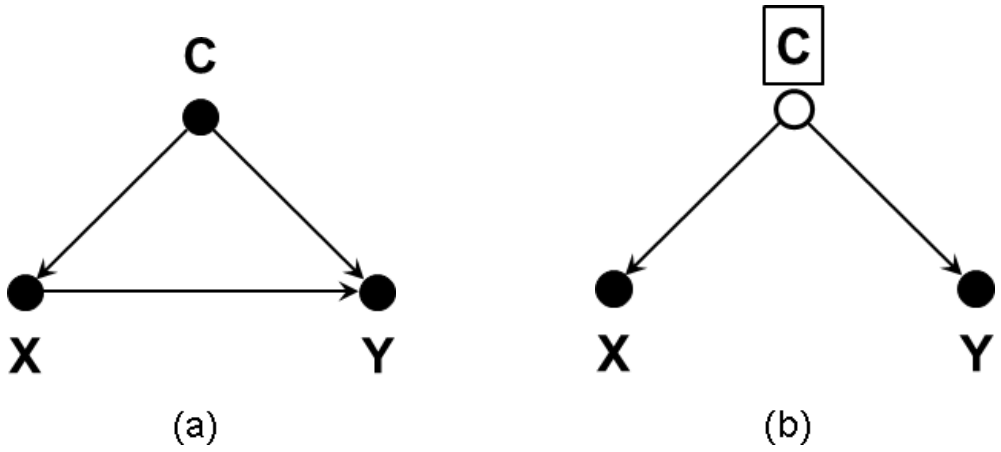


Figure 6.3 - Confounding

DAG (a) shows a causal effect of X on Y, and a confounding effect of their common cause, C. In DAG (b), X and Y are independent conditional on C, because the only path between them is the 'back-door' path via C, which is blocked.

The criteria typically used to identify confounding of an exposure-outcome association (McNamee, 2003) are that the putative confounding variable: must be a cause of the outcome (or a surrogate measure of a cause) in unexposed people; must be correlated with the exposure in the study population; and must not be on the causal pathway between the exposure and the outcome (or more generally, must not be affected by the exposure). The use of DAGs offers a structural method to detect confounding. The advantage of this approach is that it allows multiple possible confounders to be considered jointly, enabling the identification of crucial variables that must be taken into account when estimating the true effect of an exposure on an outcome. This method can be used when planning data collection or data analysis, to detect those variables that must be measured and those that need not be. This is achieved by applying the 'd-separation' criterion (see Box 6.2 below) to the DAG, either by hand or (for larger graphs) using an automated algorithm such as DAGitty (Textor, van der Zander, Gilthorpe, Liskiewicz, & Ellison, 2016). Importantly, examining variables jointly in terms of d-separation demonstrates that it is not always necessary to include all common causes in an analysis, because the potential confounding influence of one variable may already be blocked by another. This knowledge is particularly helpful when some common causes are unmeasured, or when measurement is difficult or expensive. Box 6.2 provides further information about d-separation and conditional independence in DAGs. The use of graphical methods to determine if causal effects are identified is described in section 6.4 below.

6.2.3 Collider conditioning

A shared outcome of two variables is referred to as a collider. This is shown graphically as a node where two arrowheads meet (Figure 6.4 below), i.e. an inverted fork (Elwert, 2013). Two variables, *X* and *Y*, are unconditionally independent given their shared outcome, *S* (assuming there are no other open paths between them), because the structural position of *S* as a collider prevents any association from flowing between *X* and *Y*. Importantly, however, conditioning on *S*, or on a descendant of *S*, unblocks the path between *X* and *Y*, thus inducing a spurious association within at least one stratum of *S*. This is collider conditioning (or collider stratification) bias, also known as Berkson's bias or selection bias[7] (Westreich, 2012). It is often illustrated with reference to restricted sampling situations, e.g.: sporting and musical talent may be independent in the population, but data

[7] The term 'selection bias' is sometimes used in the literature to refer to non-random selection into treatment groups (i.e. confounding). In this thesis, 'selection bias' is used to denote non-random selection into a sample or an analysis, rather than selection into treatment/exposure groups.

gathered from scholarship students in a school that selects for these characteristics will indicate a spurious negative association. In the general population, knowing that someone lacks sporting prowess would not give any clues about their musical ability, but having additional information about scholarship status induces an inverse association such that lack of sporting talent increases the likelihood of musical excellence.

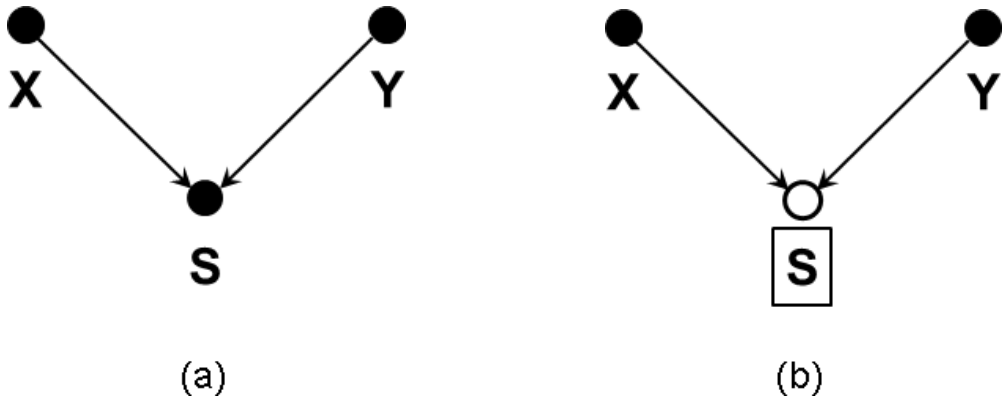


Figure 6.4 - Collider conditioning

In DAG (a), X and Y are unconditionally independent, because the path between them is blocked by their shared outcome S, which is a collider node. In DAG (b), conditioning on S (e.g. through sample selection) induces a spurious association between X and Y in at least one stratum of S.

Collider conditioning bias is recognised as a problem in a variety of research situations, not just with regard to sample selection at the study outset but also when there are missing data issues more generally (Westreich, 2012). For example, if loss to follow-up or patterns of missing assessment data are a result of the exposure and outcome of interest in the study, collider conditioning bias may affect the measured associations in the analysis sample. The potential impact on genetic association studies has also been highlighted (Munafò et al., 2017; Yaghootkar et al., 2017). Box 6.2 describes how the DAG d-separation criterion seeks to avoid collider conditioning bias whilst blocking confounding bias.

Box 6.2 - D-separation and conditional independence in DAGs

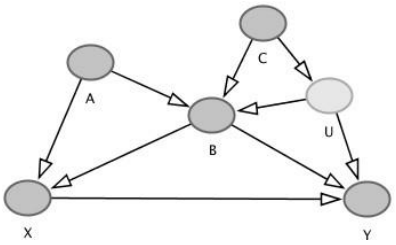
Elwert (2013, p. 252) notes that "All associations between variables in a DAG are transmitted along paths. Not all paths, however, transmit association". If two nodes *X* and *Y* are linked by any open path then they are d-connected (d stands for directional), and if all paths between them are blocked then they are d-separated. A path can be blocked by a single node; it is not necessary to block all the nodes on the path. In this way, a variable that meets the standard definition of a confounder (McNamee, 2003) will not necessarily exert a confounding influence in an analysis, provided all biasing paths are blocked by other variables.

Pearl's d-separation criterion (Pearl, 1986; Pearl, Glymour, & Jewell, 2016) determines whether each pair of nodes in a DAG is d-separated or d-connected along every possible path, regardless of the size or complexity of the DAG. This is determined entirely from the structure of the graph, rather than from measured data. The criterion for deciding whether *X* is d-separated from *Y*, given a covariate set *Z*, can be expressed as follows (Pearl et al., 2016, Ch.2):

A path is blocked by a set of nodes *Z* if and only if

1. The path contains a chain or a fork such that the middle node is in *Z* (i.e. it is conditioned on), or
2. The path contains a collider (inverted fork) such that the collider node is not in *Z*, and no descendant of the collider node is in *Z* (i.e. no conditioning on colliders or their descendants).

By applying this criterion, it is possible to generate a list of unconditional and conditional independencies implied by a DAG. For example, the DAGitty web algorithm (Textor, Hardt, & Knuppel, 2011) indicates that the following independencies are implied by the DAG below (where ⊥ denotes 'independent of' and | denotes 'conditional on'):



A ⊥ C
C ⊥ X | A, B

Figure B.3 DAG created in the DAGitty web tool, along with its testable implied independencies.

The DAGitty algorithm only generates lists of testable independencies, ignoring any unmeasured nodes (e.g. *U*, shown in light grey above). The predicted independencies can be tested empirically using measured data, as an indicator of local model fit. Observed associations that depart from the null may indicate misspecification in certain parts of the model, or other issues such as measurement error. The predicted conditional independencies derived from a DAG are equivalent to vanishing partial correlations in linear path diagrams (Pearl, 1998).

6.3 Assumptions necessary for valid causal inference

Within the potential outcomes framework, a number of assumptions have been set out as prerequisites for valid causal inference. Further assumptions are required when working with DAGs. These core assumptions are summarised here. Where additional assumptions apply to the identifiability of particular causal effects (e.g. indirect effects), these are noted in sections 6.4 and 6.5 below.

6.3.1 Conditional exchangeability

At the heart of the potential outcomes framework is the idea of comparing outcomes under mutually exclusive conditions: being exposed (or treated) versus being unexposed (or untreated). The 'fundamental problem' of causal inference (Holland, 1986) is that these outcomes cannot be observed simultaneously in the same individual; they necessitate imagining alternative possible worlds. The closest practicable way of realising and comparing potential outcomes is through randomised controlled trials (RCT), in which the treated and untreated participants experience the counterfactual conditions. The validity of RCTs rests on the comparability of the treatment and control groups at the point of randomisation: their potential outcomes are taken to be equivalent prior to treatment, because the distribution of confounding factors is the same in both groups (assuming adequate sample size and an effective randomisation scheme). In this way, the two groups are considered exchangeable by design at the point of assignment to treatment ($X = 1$) or control ($X = 0$), and their potential outcomes Y (prior to treatment) are unconditionally independent of their treatment assignment. This is expressed in counterfactual notation as

$$[Y(1), Y(0)] \perp X$$

When working with observational data in the causal inference framework, the goal is to achieve conditional exchangeability[8] between the exposed (treated) and unexposed (untreated) groups, such that their potential outcomes are independent of their exposure status, conditional on a set of observed pre-exposure covariates, Z:

$$[Y(1), Y(0)] \perp X \mid Z$$

Furthermore, an assumption of positivity must hold, which means that there must be some exposed and some unexposed individuals in each stratum of Z in the population (Westreich & Cole, 2010).

The conditional exchangeability assumption is not empirically verifiable, because the counterfactual outcomes for each individual are unobservable, and many pre-exposure characteristics will be unmeasured. Data analysis methods aimed at justifying this assumption are described in section 6.4 below.

[8] Conditional exchangeability is known by other terms such as ignorable treatment assignment, selection on observables, exogeneity, and no unmeasured confounding given the measured variables (Guo & Fraser, 2010, Ch.2; Hernán & Robins, 2006).

6.3.2 Stable unit treatment value assumption (SUTVA)

This assumption was set out by Rubin (1986), and is necessary for valid causal inference in both RCTs and observational studies (Hernán & Robins, 2006). SUTVA encompasses two elements. The first element is known as consistency: for individuals who actually receive exposure level x, actual outcome Y is the same as potential outcome $Y(x)$. In studies where there may be more than one version of the same exposure/treatment (i.e. $X = 1$ differs in some uncontrolled way across individuals, for example in a therapy trial with multiple therapists or an unstandardised treatment programme), the consistency assumption may not hold. In practice, the consistency assumption necessitates a causal contrast between well-defined exposure/treatment conditions (Hernán & Taubman, 2008). The second element is known as no interference: one individual's outcome is not influenced by the exposure status of another individual. Together, the elements of SUTVA assume that each individual has only one potential outcome under each exposure condition (Schwartz, Gatto, & Campbell, 2012). If SUTVA is violated, the causal effect cannot be predicted consistently.

6.3.3 Graphical model assumptions

Two additional assumptions apply when working with DAGs, to enable the structure of the DAG to be linked with probability statements about the variables (Glymour, 2006).

6.3.3.1 Causal Markov assumption

The causal Markov assumption underpins the d-separation criterion stated in Box 6.2 above. DAGs are Markovian or semi-Markovian (also known as recursive). In such models, each variable is independent of all its non-descendants in the graph, conditional on its parents (Pearl, 1998). This property means that all conditional independencies implied by the graph will be detected using the d-separation criterion.

6.3.3.2 Faithfulness

Under faithfulness, it is assumed that positive and negative causal effects never exactly cancel each other out (Glymour, 2006). If a positive causal effect of X on Y along one path were exactly offset by a negative causal effect of X on Y along a different path, the net statistical association would be zero; this would be unfaithful to the causal relation. It is assumed in practice that this will not occur.

6.4 Total causal effects

The total causal effect of X on Y, under the assumptions outlined above (conditional exchangeability and SUTVA), can be given as

$$E[Y(1)] - E[Y(0)]$$

which denotes the average causal effect[9] expressed as a marginal mean difference, for binary X and continuous Y data. This effect is of interest when one wishes to quantify an overall causal relationship, through all intermediate causal pathways, unbiased by confounding or selection issues. The aim is to establish whether a causal effect is likely to exist. Further research can then be undertaken to understand the mechanism by which the causal effect is transmitted; this is considered in section 6.5 below.

6.4.1 Identification

The decision as to whether the total causal effect is identified (i.e. estimable with the available data) in an observational study rests on the structural causal model developed by the researcher. DAGs are especially useful here, because they require the researcher to explicitly show his or her assumptions regarding the exposure-outcome relationship in the context of multiple covariates, in light of background knowledge in the field. D-separation-based methods are then used to determine an appropriate set of covariates to justify the conditional exchangeability assumption.

The core method by which total causal effects are identified from DAGs is called the 'back-door' criterion. It takes this name because its purpose is to block non-causal paths between the exposure (X) and the outcome (Y), and these paths begin with an arrowhead pointing to the exposure ($X\leftarrow$). The back-door criterion can be expressed as follows (Elwert, 2013; Pearl et al., 2016, Ch. 3): a set of observed variables Z (which may be empty) satisfies the back-door criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X. Therefore, conditioning on the set Z will block all spurious paths between X and Y, while leaving all directed (causal) paths open, and without creating any new spurious paths (Pearl et al., 2016, Ch. 3).

[9] This is known in much of the literature as the average treatment effect (ATE), which represents the average effect, at the population level, of moving everyone from untreated to treated status (Austin, 2011).

If there is a set *Z* that satisfies this criterion, then the total causal effect of *X* on *Y* is non-parametrically identified. The smallest possible set *Z* that satisfies the back-door criterion is referred to as a minimally sufficient set (Elwert, 2013). There may be more than one minimally sufficient set; the estimate of the total causal effect should be the same regardless of which minimally sufficient set is conditioned on, if the model is correctly specified.

An alternative approach called the 'front-door' criterion (Pearl et al., 2016, Ch. 3) may allow a total causal effect to be identified in models where the back-door criterion is not met. This approach entails "piecing together a total causal effect from its constituent parts through repeated application of the back-door criterion" (Elwert, 2013, p. 261). This would be possible in a situation where *X* and *Y* were d-connected along a back-door path by an unmeasured confounder, but all causal paths from *X* to *Y* were intercepted by a set of measured intermediate variables *Z*, and the causal paths from *X* to *Z* and from *Z* to *Y* are identified.

Instrumental variables offer an additional method for identifying total causal effects. A variable *Z* is an instrumental variable ('instrument' or IV) for the total causal effect[10] of *X* on *Y* if: *Z* affects *X*; *Z* affects *Y* only through *X*; and *Z* is independent of the *X*-*Y* confounders (Glymour & Greenland, 2008). Although used for many decades in other fields such as econometrics, IV approaches have become well-known only more recently in epidemiology. This is partly through the popularity of Mendelian randomisation, which uses genotypic data as instruments for health-related exposures (Lawlor, Harbord, Sterne, Timpson, & Smith, 2008). Instrumental variables can be readily detected from the structure of a DAG, and this is automated in the DAGitty program (Textor et al., 2016).

6.4.2 Estimation

Once a causal effect has been structurally identified, there are many different methods for estimating this using the data. The most common approaches will be outlined here.

[10] The total causal effect identified in this way is known as the local average treatment effect (LATE) or the complier average causal effect (CACE), because it is interpreted as the causal effect among those participants for whom the instrument *Z* actually affects the value of the exposure/treatment *X* (Angrist, Imbens, & Rubin, 1996).

6.4.2.1 Multiple regression

Multiple regression can be used to estimate total causal effects using observational data, adjusting for all necessary covariates determined by the identification process. Most statistical analysis packages accommodate a wide range of linear and generalised regression models, thus permitting flexibility in modelling different types of outcome variable distributions and non-linear functions. Provided the assumptions necessary for causal inference can be justified, the coefficient of the exposure variable is interpreted as a causal effect, conditional on the covariates. The coefficients of the covariates do not have a causal interpretation; erroneous interpretation of these coefficients has been termed the 'Table 2 fallacy' (Westreich & Greenland, 2013). It should also be noted that the use of multiple regression to estimate a causal effect is distinct from its use for prediction purposes: a prediction score can be developed from multiple regression model coefficients, reflecting the predictive power of certain risk factors, but this does not require, or imply, a causal interpretation (Hernán & Robins, 2017, Ch. 3).

6.4.2.2 Propensity score-based methods

These methods aim to use a measure of exposure propensity to achieve conditional exchangeability (predicated on the assumptions already discussed above). Various estimators are readily available in commonly used statistics packages, including Stata and R.

Rather than using the exposure and covariate data to model the outcome directly, these estimators first model the probability (propensity) of belonging to the exposed group (Austin, 2011; Garrido et al., 2014). Exposure status is entered as the dependent variable in a logit or probit regression model, with the observed covariates entered as the regressors, including interaction terms as required. The fitted probability values from this model form a propensity score, which is then used to match or weight the exposure groups when modelling the outcome. The aim is to achieve conditional balance in the distribution of the observed covariates between the two groups; this can be checked using balance diagnostic procedures, although of course the influence of unmeasured confounding factors cannot be tested empirically. It has been suggested that all measured pre-exposure variables can be entered into this model (Austin, 2011). Somewhat counter-intuitively, simulation studies have indicated that including covariates that are thought to affect the outcome but not the exposure can improve the propensity model fit, possibly because such variables can show chance associations with the exposure in a measured dataset (Brookhart et al., 2006). The

chosen set of variables should also satisfy the back-door criterion for the total causal effect of the exposure on the outcome (Pearl et al., 2016, Ch. 3). Finding the best specification of the propensity model can be aided by data-driven or machine learning methods such as generalised boosted modelling (McCaffrey, Ridgeway, & Morral, 2004), although it should be noted that machine learning classifiers that aim to maximise exposure classification accuracy will not necessarily produce optimal covariate balance in a particular dataset.

A benefit of the propensity score is that it is a single measure based jointly on all the covariates, thus avoiding the problem of dimensionality that arises when attempting to stratify or match directly on multiple covariates (Guo & Fraser, 2010, Ch. 5). A further advantage is that imbalance can be directly 'diagnosed' in the propensity model, whereas misspecification of a multiple regression model for the outcome is harder to detect. It has also been noted that the process of modelling exposure propensity separately before modelling the outcome is analogous to the stages of an RCT, where design factors are dealt with before the outcome is analysed (Austin, 2011), thus providing a principled way of approaching the analysis.

Propensity score-based estimators typically employ either matching or weighting procedures. Propensity score matching (PSM) involves deriving matched sets of exposed and unexposed participants whose propensity scores are similar. This can be done on a one-to-one or one-to-many basis, with or without replacement, and the researcher can define the acceptable limits of similarity of the propensity scores (Austin, 2011). A potential drawback is that this entails excluding participants who cannot be adequately matched, thus reducing the sample size and restricting the region of common support in the analysed data (Guo & Fraser, 2010, Ch. 5). An alternative approach is inverse probability weighting (IPW): a participant's analysis weight is calculated as the reciprocal of the probability of the exposure status that was actually observed, which equates to $\frac{1}{propensity\ score}$ for the exposed participants and $\frac{1}{(1-propensity\ score)}$ for those in the unexposed group, and this weight is then applied to a regression model of the outcome on exposure status. This method allows the whole sample to be analysed, but results may be sensitive to extreme weights (Austin, 2011).

PSM and IPW both focus on specifying a model for exposure status rather than a model for the outcome. Related methods that include separate models for both the exposure and the outcome are also available. These are known as doubly robust estimators, because the

result will be unbiased and consistent as long as either the exposure model or the outcome model is correctly specified; it is not necessary for both to be correctly specified (Bang & Robins, 2005). A propensity model for the exposure is specified as described above, and a model for the outcome is separately specified, including covariates that overlap with the propensity model regressors if appropriate. Estimators of this type include IPW with regression adjustment and augmented IPW (Cattaneo, Drukker, & Holland, 2013). Targeted maximum likelihood estimation (TMLE) can also be used, with the option of incorporating machine learning methods to optimise specification (Schuler & Rose, 2017).

6.4.2.3 IV estimation

If the total causal effect has been identified using the instrumental variables approach, this can be estimated using IV models such as two-stage least squares (Wooldridge, 2016, Ch. 15); this essentially entails regressing the exposure variable on the IV(s), obtaining the fitted values, and then regressing the outcome variable on the fitted values. This can be implemented in one step, for example using `ivregress` in Stata. Additional IV estimation methods have been developed specifically for Mendelian randomisation studies, where there may be problems with weak instruments or pleiotropy (Bowden, Smith, & Burgess, 2015; Bowden, Smith, Haycock, & Burgess, 2016; Burgess, Small, & Thompson, 2015).

6.5 Decomposition into direct and indirect effects

The practice of partitioning total effects into direct and indirect effects (Figure 6.5) has a long history in psychology and the social sciences. Many researchers in these fields are familiar with regression-based techniques for quantifying mediation, and extensions into larger path analyses, usually within the framework of the so-called Baron-Kenny approach (Baron & Kenny, 1986) and linear structural equation modelling (SEM; e.g. Kline, 2016).

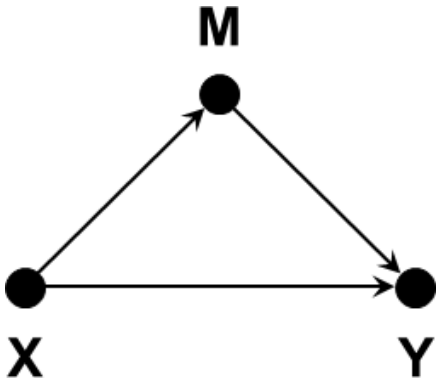


Figure 6.5 - Simple mediation model

DAG showing a direct causal effect of X on Y, and an indirect causal effect through a mediator, M. The total effect of X on Y is the effect through all paths together.

Effect decomposition has historically received less attention in epidemiology, although recent years have seen an increasing interest in the goal of opening the 'black box' of causality to understand the pathways by which causes lead to outcomes (Hafeman & Schwartz, 2009); it is argued that this can help to strengthen confidence in the total effect estimates, as well as highlighting intermediate factors that are relevant for intervention planning.

This increase in interest has sparked greater attention to methods for causal mediation analysis within the potential outcomes framework. In particular, the assumptions required for estimates in a mediation model to have a causal interpretation have come under detailed scrutiny. A key issue that has been highlighted is that even if exchangeability can be assumed with respect to exposure status, the association between the mediator and the outcome may still be confounded. This is true even in an RCT, because although the exposure (treatment) is randomised, the mediator usually is not (D. P. MacKinnon & Pirlott, 2015; VanderWeele, 2016b). There is reportedly less awareness of the problem of mediator-outcome confounding among practitioners working in the Baron-Kenny or SEM frameworks (Pearl, 2014), and it has been suggested that this is partly because this issue was not discussed in the classic paper of Baron and Kenny (1986), despite having been clearly highlighted in an earlier paper by Judd and Kenny (1981). This has been a key focus for causal mediation researchers within epidemiology, and will be discussed further below in the context of effect identification. Conversely, the opportunity to incorporate measurement error through latent variables in causal mediation models has been

highlighted by experts working in the SEM tradition (Muthén & Asparouhov, 2015), and the importance of this for valid causal inference warrants greater attention.

The intrinsic links between the graphical and counterfactual frameworks and linear SEM path models have been demonstrated in detail (De Stavola, Daniel, Ploubidis, & Micali, 2015; Pearl, 2012, 2014). One difference of emphasis, however, concerns the conceptualisation of the direct and indirect effects. In the Baron-Kenny approach, these are stated in terms of products or differences of linear regression coefficients (outlined in the context of estimation in section 6.5.2 below). Within the counterfactual approach, non-parametric definitions have been put forward for two different types of direct effects, and one indirect effect (Pearl, 2014). These definitions rest on the comparison of two hypothetical worlds, in which the values of the exposure and/or the mediator differ. In the examples below, X is a binary exposure, Y is a continuous outcome, and M is a binary or continuous mediator. Note that with respect to one particular mediator M, the term 'direct effect' refers to the portion of the total effect that does not go through M (although it may go through other mediators), and the 'indirect effect' refers to the portion of the total effect that goes only through M.

The controlled direct effect (CDE) of X on Y when M is controlled at m, expressed as a marginal mean difference, is:

$$\text{CDE}(m) = E[Y(1, m)] - E[Y(0, m)]$$

Here, X is set to 1 in one hypothetical world, and set to 0 in the other, but M is set to the same fixed value m in both worlds. By fixing M at m, the CDE represents the direct effect of X only, unmediated by M. The complementary concept of a controlled indirect effect is undefined, however, because the CDE will change according to the value that M is set to, yet the total effect stays constant (i.e. it is not possible to subtract the CDE from the total effect to obtain a controlled indirect effect). Instead, a natural direct effect (NDE) has been defined, which does have a complementary natural indirect effect, together summing to the total effect. The NDE compares one world in which X is set to 1, and another world in which X is set to 0. But in both worlds, M is set to $M(0)$, i.e. the value that M would naturally take had X been set to 0:

$$\text{NDE} = E[Y(1, M(0))] - \text{E}[Y(0, M(0))]$$

For the natural indirect effect (NIE), *X* is set to 1 in both worlds. In one world, *M* is set to the value it would naturally take when *X* is 1, but in the other world, *M* is set to the value it would have naturally taken had *X* been 0:

$$\text{NIE} = E[Y(1, M(1))] - E[Y(1, M(0))]$$

If the indirect effect is of primary interest in a research study, specific identification assumptions must be met, and these are outlined below.

6.5.1 Identification

The following extended assumptions apply to the identification of natural direct and indirect effects (De Stavola et al., 2015; Pearl, 2014): the no interference assumption extends to the effect of *X* on *M* and of *M* on *Y*; the consistency assumption is generalised to all potential outcomes for *M* and *Y*; and the conditional exchangeability assumption requires all back-door paths between *X* and *M* and between *M* and *Y* to be blocked, as well as those between *X* and *Y*.

If the structural model indicates that the mediator of interest is itself affected by another mediator, this is termed intermediate confounding (De Stavola et al., 2015), as shown in Figure 6.6 below.

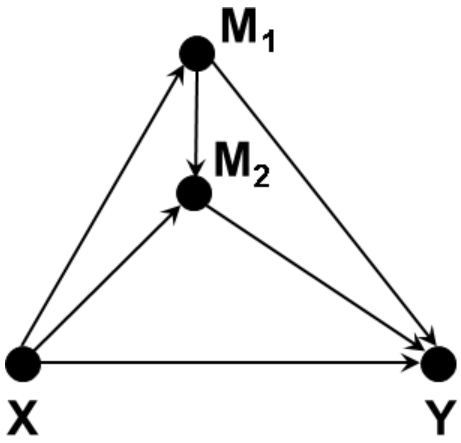


Figure 6.6 - Intermediate confounding

In this DAG, M_1 and M_2 are both mediators of the causal effect of X on Y. M_1 also affects M_2, which means that M_1 is an intermediate confounder: M_1 lies on an open back-door path from M_2 to Y.

This situation imposes additional identification challenges for the natural direct and indirect effects. If M_2 is the mediator of interest in Figure 6.6, then M_1 must be conditioned on, because it is a confounder of the causal relationship between M_2 and Y. M_1 is also affected by X, however, which normally suggests it should not be conditioned on. In this situation, it is possible to condition on the intermediate confounder M_1 if a further parametric assumption is made that there is no interaction between X and M_2 at an individual level (De Stavola et al., 2015; Robins & Greenland, 1992). Alternatively, an assumption can be made that, conditional on other relevant confounders, the CDE does not vary according to the value of M_2 when $X = 0$ (Petersen, Sinisi, & van der Laan, 2006). It has been suggested that these estimands be considered 'randomised interventional analogs' of the NDE and the NIE (VanderWeele, Vansteelandt, & Robins, 2014): rather than setting the values of the mediator to what they would have been for each individual under a particular level of the exposure, the effect is instead defined with the mediator set to a random value from the distribution of mediator values among all those at a particular level of exposure.

6.5.2 Estimation

6.5.2.1 General methods

As with total causal effects, several options are available for estimating direct and indirect effects. The most general method is based on G-computation (Robins, 1986), a technique equivalent to standardisation to the total population, which was originally introduced as a way of estimating the effects of time-varying exposures. As implemented in the Stata command `gformula` (Daniel, De Stavola, & Cousens, 2011), this is extended to generate estimates of the CDE, NDE, NIE and total effects, in flexible models that accommodate continuous and binary outcome variables. The `gformula` command also estimates models with intermediate confounding. The command `paramed` (Emsley & Liu, 2013) is also available in Stata, although it does not accommodate models with intermediate confounding. Other estimators mentioned earlier, such as IPW and doubly robust approaches, can be used in mediation analysis (De Stavola et al., 2015; Linden & Karlson, 2013).

6.5.2.2 Linear regression methods

SEM and Baron-Kenny framework mediation models are based on linear regression estimation. The core feature of this approach to mediation, as developed by Baron and

Kenny (1986), is that the relationship between total, direct and indirect effects is expressed in terms of products or differences of linear regression coefficients. Figure 6.7 below shows a simple mediation model with structural parameters labelled α, β and γ.

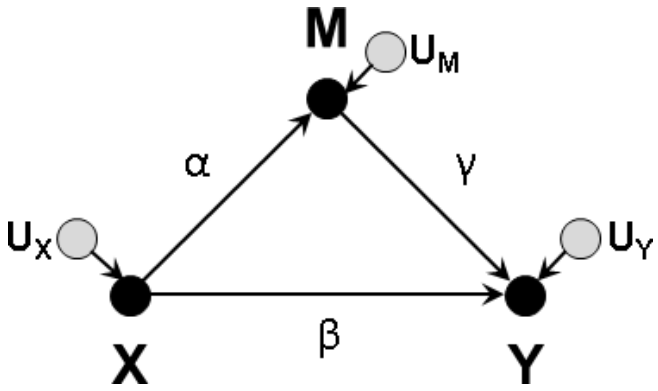


Figure 6.7 - Linear mediation model

This is the linear case of a basic DAG for mediation. The structural parameters labelled on each path correspond to linear regression estimates, assuming independent error terms and no interaction.

Provided that the aforementioned causal inference assumptions hold, in addition to assuming linearity, independent errors and no interaction, the causal effects are identified. The equations for the three variables are:

$$x = u_X \qquad m = \alpha x + u_M \qquad y = \beta x + \gamma m + u_Y$$

The total causal effect τ is:

$$\tau = \beta + \alpha\gamma$$

This decomposes into the direct effect, β, and the indirect effect, αγ. The total effect τ can be estimated by regressing *Y* on *X*. Then, *Y* is regressed on *X* and *M* to obtain the coefficient of *X* conditional on *M* (the direct effect, β), and this is subtracted from the total effect to quantify the indirect effect mediated through *M* (αγ). Alternatively, *M* can be regressed on *X* to obtain coefficient α, and *Y* can be regressed on *M* conditional on *X* to give γ. These can be multiplied to give the indirect effect, then subtracted from τ to give the direct effect β.

If all the assumptions hold, τ corresponds to the counterfactually-defined total causal effect, β corresponds to both the NDE and the CDE(m), and $\alpha\gamma$ corresponds to the NIE (Pearl, 2014).

Linear regression-based mediation models can be estimated with specific packages such as PROCESS for SPSS and SAS software (Hayes, 2013). More extensive path diagrams or structural models with latent variables can be estimated using SEM programs, including Stata's `sem` command or standalone programs such as Mplus (Muthén, Muthén, & Asparouhov, 2016). These allow a system of equations to be estimated simultaneously, following which specific directed path coefficients within the model can be multiplied and summed to derive any desired indirect effect estimates. Strong assumptions are necessary for the results of such a procedure to have a causal interpretation, because the no-confounding requirement applies to every path in the model. An advantage of the SEM estimation method is that, assuming all assumptions hold, the issue of multiple mediators and intermediate confounding is easily accommodated in the simultaneous estimation process (De Stavola et al., 2015).

6.6 Validity and generalisability

As has been evident in this chapter so far, valid causal inference is heavily dependent on a range of testable and untestable assumptions. Care must be taken to consider these at all stages of study planning and analysis, and sensitivity analyses for key threats to validity should be undertaken where possible. This section summarises some of these considerations.

6.6.1 Model misspecification and residual confounding

Uncertainty regarding temporal order is a particular problem in cross-sectional studies, and in many longitudinal studies, unless the study cohort was recruited long before the relevant exposure period. Despite the availability of data-driven estimation methods and balance 'diagnostics', misspecification is not generally amenable to empirical evaluation, primarily because the role of unmeasured covariates is unknown. It is important that researchers spend time developing comprehensive graphical models, including plausible alternative specifications. The final analysed models can also be examined for structural equivalence with other models that are compatible with the data. A range of fixed and probabilistic methods for quantitative bias analysis are now available, allowing researchers to explore

the magnitude of residual bias that would negate the observed results (Lash, Fox, & Fink, 2009; Rosenbaum, 2002).

6.6.2 Selection bias and missing data

The non-random nature of recruitment into most observational studies means that collider conditioning bias is likely to be widespread, particularly with regard to the psychiatric and cognitive characteristics that are the subject of this thesis. Patterns of missing data among the participants will also be subject to this problem, leading to biased selection into the analysis sample. Methods are available to determine the probable direction of bias in light of the presumed missingness mechanism (Daniel, Kenward, Cousens, & De Stavola, 2012), and quantitative bias analysis approaches can also be applied (Lash et al., 2009), although these may require additional data (e.g. regarding invitees who declined or participants who dropped out). Weighting and/or imputation methods may be appropriate, if missingness mechanism assumptions are met, and some statistical programs offer optimised estimation methods such as full information maximum likelihood as an alternative to complete case analysis.

6.6.3 Measurement error, misclassification and power

The potential effect of measurement error and misclassification depends on whether it arises randomly or systematically, and whether it is differential or non-differential with respect to the outcome. VanderWeele and Hernán (2012) have developed ‘signed DAGs’ to facilitate a qualitative appraisal of the likely presence and direction of true causal effects in the context of different error scenarios. Lash et al. (2009) again provide quantitative options to estimate the effect of different levels of classification accuracy on the results estimates. The biasing impact of measurement error in a mediator can be appraised with regard to the direct and indirect effect estimates (VanderWeele, Valeri, & Ogburn, 2012), and measurement error can be incorporated directly in SEM analysis through the use of latent variables. The implications of random measurement error for study power should also be considered.

6.6.4 Generalisability

Related to the issue of sample selection bias are questions regarding the extent to which study results are generalisable (or transportable) to the source population or to populations elsewhere. A range of weighting and imputation-based methods can be used to estimate the

magnitude of bias in study estimates compared with the reference population, provided that data are available for comparison (Gorman et al., 2017; Stuart, Cole, Bradshaw, & Leaf, 2011; Westreich, Edwards, Lesko, Stuart, & Cole, 2017). It is also important to gather information about the population prevalence of the various component causes that act in concert with identified exposures, as this will influence the extent to which the exposure expresses its effects in a given population (Keyes & Galea, 2017).

6.7 Criticisms of the causal inference approach

The causal inference approach is relatively new in epidemiology, and it has generated some strong criticism (Broadbent, Vandenbroucke, & Pearce, 2016; Krieger & Smith, 2016a, 2016b; Vandenbroucke, Broadbent, & Pearce, 2016), together with an equally robust defence from its proponents (Blakely, Lynch, & Bentley, 2016; Daniel, De Stavola, & Vansteelandt, 2016; VanderWeele, 2016a; VanderWeele, Hernán, Tchetgen, & Robins, 2016).

One of the criticisms expressed is that causal inference methods require unfeasibly strong assumptions. Conversely, it has been pointed out that many of the same assumptions simply remain hidden in other types of analysis, and that causal inference methodologists helpfully direct much of their efforts to determining when such assumptions can be relaxed and what kinds of sensitivity analyses should be done (Daniel et al., 2016). In the words of Pearl and Bareinboim (2014, p. 580), "…assumptions are self-destructive in their honesty. The more explicit the assumption, the more criticism it invites, for it tends to trigger a richer space of alternative scenarios in which the assumption may fail. Researchers prefer therefore to … make assumptions in private". The approach taken in this thesis is to aim for transparency regarding the assumptions that are made.

The assumption that has sparked the most debate in this regard is that exposures or treatments should be well-defined (Hernán, 2016; Kaufman, 2016; Schwartz, Gatto, & Campbell, 2016; Vandenbroucke et al., 2016). In its strongest form, this entails taking the position that there is "no causation without manipulation" (Holland, 1986) in epidemiological research, i.e. that causal effects cannot be attributed to any factor that is not amenable to intervention. The position of current practitioners appears to be more nuanced than this (VanderWeele, 2016a), with a distinction being drawn between a broad definition of a cause, and a somewhat narrower focus on what sorts of counterfactuals can reasonably be given a numerical estimate in an analysis. Nevertheless, key figures in the

field such as Hernán have argued cogently for the importance of well-defined exposures/interventions for underpinning the core assumptions of causal inference methods (Hernán, 2005, 2016; Hernán & Taubman, 2008).

A second point of debate has been the relative merits of DAGs to aid causal inference (Krieger & Smith, 2016b). As already noted above, it is important that time is taken to develop comprehensive DAGs that appropriately reflect expert knowledge in the field, and to acknowledge plausible alternative specifications. Concerns centre on the possibility that users of DAGs will become blinkered by the diagram, and hence will make reactive model alterations that are data-driven rather than guided by theory or knowledge. This concern appears to be partly based on misunderstandings about what automated tools such as DAGitty actually do; the identification algorithms are entirely driven by the graph structure (i.e. the researcher's beliefs), without involving any measured data at all, but it is not clear if this is fully appreciated by some critics[11] (Krieger & Smith, 2016a).

A broader issue that has been raised is that causal inference methods are but one of many complementary strategies for getting closer to the truth of causation. Critics have reminded users that causal inference in a broad sense relies on triangulation across different research methods with their own particular strengths and unique sources of bias (Lawlor, Tilling, & Smith, 2016). The benefits of 'inference to the best explanation' have also been highlighted, as a guide for moving away from overly deductive thinking and towards a more open-ended process of evaluating a causal explanation in the context of other plausible explanations (Krieger & Smith, 2016b). Thus the results of any one study—within the causal inference framework or not—must be interpreted cautiously without losing sight of the wider context.

6.8 Overview of analyses in UK Biobank

Chapters 7 and 8 describe the application of causal inference methods to understand the effects of psychiatric conditions on cognitive function, and the role of intermediate risk factors in explaining any such effect. These methods were applied to cross-sectional observational data in the UK Biobank cohort, and the benefits and limitations of this framework were considered in light of the available data and the plausibility of the

[11] The cited paper states critically that DAGitty offers automation for "selection of instrumental variables from the data provided", and that "The use of instruments that are data-derived, rather than from subject-specific knowledge, is likely to lead to highly misleading findings, given the impact of measurement characteristics on such selection".

assumptions required for valid causal inference. Graphical models were developed, which informed methods to estimate total causal effects, and direct and indirect effects where these were identified. Multiple estimators were used for comparison, and sensitivity analyses were conducted where possible with respect to key threats to validity.

Chapter 7 Explaining variation in cognitive function in bipolar disorder

The study presented in this chapter aimed to estimate the total causal effect of bipolar disorder on cognitive function, and to decompose any such effect into direct and indirect effects through potentially modifiable intermediate factors. A graphical model was first developed, based on previous literature and assumptions regarding confounding influences and intermediate causal pathways. This was then applied to baseline data from UK Biobank. Following initial evaluation of the fit of the graphical model to these data, the total effect was identified and estimated using a range of regression- and matching-based methods. Decomposition of the total effect into direct and indirect effects was then conducted, where possible. The results were interpreted in light of the assumptions outlined in Chapter 6.

7.1 Methods

7.1.1 Participants

Cross-sectional data from the full cohort at baseline (n = 502,618) were used. Of these, participants were retained for analysis if they had sufficient data to classify their exposure status (see below) and had data on at least one cognitive outcome measure. Adjusted analyses were restricted to participants of white British genetic ancestry, as explained in more detail below; the unadjusted analyses were not restricted by ethnic group. The sample size available for analysis is detailed in section 7.2.1.

7.1.2 Exposure status

The illness exposure of interest was mania or bipolar disorder (mania/BD), classified using the broad definition given in Chapter 4 (i.e. met criteria for mania or bipolar disorder according to self-reported diagnosis, or touchscreen mood questionnaire algorithm, or pre-baseline hospital records; participants were classified as exposed if they met the criteria in at least one information source, even if other information sources were missing). The unexposed comparison group comprised participants who had complete self-reported and touchscreen mood questionnaire data indicating that they did not meet criteria for mania/BD or major depression, and whose pre-baseline hospital records had no primary or secondary diagnosis of mania/BD or major depression. Major depression was excluded in

addition to mania/BD here because the latter superseded the former in the hierarchical exposure definition, and it was therefore considered that major depression was already excluded, in effect, from the exposed group. Furthermore, because misclassification is common between mania/BD and schizophrenia spectrum disorders (Bromet et al., 2011; Laursen, Agerbo, & Pedersen, 2009), participants with a self-reported diagnosis or hospital record of schizophrenia (as defined in Chapter 4) were excluded from the exposed and unexposed groups. Participants who did not meet the above criteria for either the exposed or unexposed groups were not further analysed.

7.1.3 Cognitive outcome measures

The cognitive measures analysed were reasoning, reaction time, numeric memory, visuospatial memory and prospective memory, as described in Chapter 3. These data were provided by UK Biobank as raw scores, and for the purposes of the present analysis were standardised within five-year age strata, using all available data in the cohort at baseline. Five-year bands were deemed appropriate in light of the typical rate of age-related change in cognitive performance in middle to older adulthood (Strauss et al., 2006, Ch. 2), and this is similar to the strata sizes used across this age range in 'gold standard' cognitive batteries such as the Wechsler scales (Wechsler, 2010, 2011). To address skew in the raw data distributions, the scores were first transformed into percentiles and then into z-scores (mean 0 and standard deviation 1). The scores for analysis therefore represent the number of standard deviations above or below the mean score within each five-year age stratum. The scores for reaction time and visuospatial memory were reflected so that higher scores represent better performance, in line with the other tests. It was not possible to standardise the prospective memory data in this way because responses were dichotomous (correct response at the first attempt or not), and so the raw data were used in the analyses involving this test.

7.1.4 Covariates

A range of background measures and potential intermediate factors were considered in the analyses, and their putative relationships with mania/BD, cognitive performance and each other were depicted in a directed acyclic graph (see sections 7.1.5.1 and 7.2.2 below).

7.1.4.1 Sociodemographic measures

Age was truncated to whole years, and was centred at 55 (approximating the cohort mean age at baseline) in the analyses. Gender was self-reported as male or female. Ethnic background was self-reported as white, Asian/Asian British, black/black British, Chinese, mixed or other. Participants who had self-reported a white British background were further grouped by similarity of genetic ancestry based on a principal components analysis of the genotypic data (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22006). Participants self-reported their birth country, and these were grouped according to whether or not English was an official/first language (UK, Isle of Man, Channel Islands, Gibraltar, Ireland, Australia, New Zealand, USA, Canada, Anguilla, Antigua & Barbuda, Aruba, Bahamas, Barbados, Bermuda). Self-reported data regarding participants' highest educational qualification were dichotomised as university/college degree or not. Neighbourhood deprivation level was recorded by UK Biobank prior to baseline using the Townsend Index (Townsend, 1987), and this was converted into quintiles in the whole cohort.

7.1.4.2 Local environment

The population density of each area of residence was classified categorically by UK Biobank, by combining participants' residential postcodes with data generated from the 2001 census, using the GeoConvert tool provided by the UK Data Service Census Support (http://geoconvert.mimas.ac.uk/). Proximity to the nearest major road (traffic intensity >5,000 motor vehicles per 24 hours) was calculated by UK Biobank as the inverse distance (1/m) from the baseline address, using data for the year 2008 provided by the Department for Transport, and was converted to quintiles in the whole cohort. Neighbourhood air pollution data from a land use regression model and satellite-derived estimates (Vienneau et al., 2013) were linked by UK Biobank to participants' baseline addresses; particulate matter of up to 10μm diameter (PM_{10}) and nitrogen dioxide (NO_2) were measured as annual average values in $\mu g/m^3$ (for the years 2007 and 2005 respectively) and were converted to quintiles in the whole cohort. Other air pollution data were also available but these were measured in later years, thus post-dating the cognitive assessment date for most participants, and so were not analysed.

7.1.4.3 Lifestyle and physical measures

Tobacco smoking status (current, former or never) was classified by UK Biobank using self-reported data. Self-reported frequency of alcohol consumption was categorised as

daily/almost daily, 3-4 times per week, 1-2 times per week, 1-3 times per month, special occasions only, former drinker, or never drinker. Sleeplessness/insomnia was self-reported as never/rarely, sometimes, or usually; if participants were unsure how to respond to this item, they were prompted to answer in relation to the past four weeks. Physical activity (walking, moderate and vigorous) in a typical week was recorded using self-reported items from the International Physical Activity Questionnaire short form (Booth, 2000), from which a single measure of total physical activity in metabolic equivalent of task (MET) hours per week was derived; this was converted into quintiles in the whole cohort. Body mass index (BMI; kg/m^2) was calculated from measures of height and weight taken by UK Biobank staff, and was categorised as underweight (<18.5), normal (18.5 to 24.9), overweight (25.0 to 29.9), obese class I (30.0 to 34.9), obese class II (35.0 to 39.9), and obese class III (≥40.0).

7.1.4.4 Medical and family history

Participants were asked to self-report any illnesses previously diagnosed by a doctor, during the touchscreen questionnaire and verbal interview. These data were also combined with hospital records by UK Biobank analysts to generate 'adjudicated' classifications of myocardial infarction (http://biobank.ctsu.ox.ac.uk/crystal/docs/alg_outcome_mi.pdf) or stroke (http://biobank.ctsu.ox.ac.uk/crystal/docs/alg_outcome_stroke.pdf). For the present analysis, a dichotomous indicator was created for history of any cardiometabolic disease (self-reported diagnosis of angina, hypertension or non-gestational diabetes, or adjudicated diagnosis of myocardial infarction or stroke, pre-dating the baseline assessment). A dichotomous indicator was also created for history of any neurological or psychiatric condition (apart from mood disorder or schizophrenia) in the self-reported or hospital records data; the conditions included were as described in Chapter 4. Family history of certain illnesses in biological parents and siblings was queried during the touchscreen questionnaire, and for the present analysis dichotomous indicators were generated for history of psychiatric or neurological conditions (dementia, PD or severe depression, coded separately) in any parent or sibling. Participants also self-reported whether their mothers had smoked regularly around the time of their birth.

7.1.4.5 Mental health and psychotropic medication

In addition to the self-reported and hospital records data regarding psychiatric diagnoses, participants also provided self-reported information at baseline about depressive symptoms in the past two weeks (scored 0 to 12) and current psychotropic medications (dichotomous

indicator for any mood stabiliser, antidepressant, antipsychotic, sedative or hypnotic), as detailed in Chapters 3 and 4. Additional self-reported information was elicited using the web-based mental health questionnaire, which was administered in 2016. Information about the number of episodes of depressed mood or anhedonia experienced across the lifetime was collected both at baseline assessment and in the web-based questionnaire; for the present analysis this was coded ordinally (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, >10 and 'too many to count') using the baseline data if available, or the web-based data if the baseline data were missing. These data were available for participants regardless of their mood disorder exposure status (i.e. participants may have reported one or more such episodes without necessarily meeting the criteria for mood disorder). It was not possible to distinguish how many depression episodes preceded the baseline assessment date, in participants who only had web-based data. Participants were also asked five questions from the brief Childhood Trauma Questionnaire (Bernstein et al., 2003) within the web-based mental health questionnaire, representing examples of abuse (physical, emotional, sexual) and neglect (physical, emotional). Ordinal responses on each of the five questions were dichotomised using thresholds extrapolated from previous research (Walker et al., 1999), and an overall dichotomous indicator was created to represent above-threshold responses on one or more of the five items.

7.1.4.6 Genome-wide polygenic scores

Genome-wide polygenic scores (GPS) were generated from published genome-wide association study (GWAS) results, as described in Chapter 3. A GPS was generated from summary statistics from a large GWAS of years of education (Okbay et al., 2016); to minimise sample overlap with UK Biobank participants, the GWAS authors provided summary statistics from analyses that did not include UK cohorts (participants from the 23andMe data resource were also omitted due to data-sharing restrictions). Years of education was used here as a proxy for general cognitive ability because, at the time the present analysis was conducted, results were unavailable from any similarly-sized GWAS of general cognitive ability that did not involve UK Biobank participants. Education and cognitive ability have a genetic correlation of ~0.8 (Hill et al., 2018), and current evidence suggests that a GPS based on the very large available GWAS of education has greater predictive power for observed cognitive ability than does a GPS based on smaller GWAS of cognitive ability itself (Plomin & von Stumm, 2018). A GPS was also generated for bipolar disorder, using summary statistics from the Psychiatric Genomics Consortium

(Sklar et al., 2011); this GWAS included UK cohorts and participant overlap is therefore possible with UK Biobank.

Both GPS were generated from all available SNPs in the UK Biobank cohort, applying the following quality control criteria: information score >0.8; Hardy-Weinberg equilibrium test $p > 1x10^{-6}$; minor allele frequency >0.01; linkage disequilibrium clumping $R^2 < 0.1$ using a 250kb window. GPS were created at various thresholds based on the *p* values in the source GWAS ($5x10^{-8}$ to 0.9), and the optimum GPS was chosen based on the magnitude of the variance explained (R^2) in the relevant phenotype measures in the UK Biobank data. These analyses were conducted in unrelated UK Biobank participants of white British genetic ancestry, after standard quality control exclusions for sex mismatch, sex chromosome aneuploidy, and outlying values of heterozygosity and missingness (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22027). Each GPS was first regressed on variables indicating the genotyping array and batch, UK Biobank assessment centre, and the first 20 genetic principal components. The residuals from these models were then used as the independent variable in models to predict the relevant UK Biobank phenotype.

Appendix W shows the R^2 at each *p* threshold, and the regression coefficient or odds ratio for the association between deciles of the optimum GPS and the relevant phenotypes. In each case, the optimum GPS had a *p* threshold of 0.5. The R^2 for the association between the optimum education/cognition GPS and having a degree in the UK Biobank cohort was 0.018, and was 0.013 for the raw reasoning test score. These results are similar to those previously reported in independent samples, e.g. $R^2 = 0.02$ for both educational attainment (de Zeeuw et al., 2014; Krapohl & Plomin, 2016) and general cognitive ability (Krapohl et al., 2016). They are notably lower, however, than more recent analyses which made use of the full Okbay et al. GWAS results including UK participants (Selzam et al., 2017), in which the R^2 for educational attainment at age 16 was 0.091 and for general cognitive ability was 0.036. The R^2 for the association between the optimum bipolar GPS and the broadly-defined mania/BD phenotype in the UK Biobank cohort was 0.01. This is lower than the variance explained in other independent samples, e.g. $R^2 = 0.024$ in Aminoff et al. (2015), although results may not be directly comparable due to different methods of calculating pseudo R^2 in logistic regression models.

7.1.5 Data analysis

7.1.5.1 Graphical model

A DAG was constructed to represent causal assumptions about the relationship between lifetime history of mania/BD and cognitive performance, in the context of possible confounding factors and intermediate pathways. This was done before any data were analysed. The nodes in the DAG and the assumed directional relationships between them were determined from previous systematic reviews of cognitive function in BD (Arts et al., 2008; Bora, Yucel, & Pantelis, 2009; Bora, Yucel, Pantelis, & Berk, 2011; Bourne et al., 2013; Camelo, Velasques, Ribeiro, Netto, & Cheniaux, 2013; Lee et al., 2014; Mann-Wrobel et al., 2011; Robinson et al., 2006; Samamé, Martino, & Strejilevich, 2015) and the systematic review reported in Chapter 2, as well as general background knowledge and assumptions regarding other shared causes that were necessary to depict in order for the DAG to have a causal interpretation (as explained in Chapter 6). This included nodes depicting constructs that were not measured in the UK Biobank dataset (e.g. premorbid intellectual ability). The structure of the DAG also took into account the cross-sectional nature of the planned data analysis, in that nodes were labelled to represent past states where necessary. The DAG was reviewed by the thesis supervisors, to reach consensus about the likely structure and direction of relationships. If the putative direction of a relationship was unclear, multiple versions of the DAG were generated prior to data analysis, to represent alternative plausible specifications.

The DAG was drawn using DAGitty software (Textor et al., 2016), which automatically generated a list of all the testable independencies implied by its structure (ignoring any nodes that were tagged as being unmeasured). These were then tested in the dataset, by calculating partial correlation coefficients between each pair of nodes that were predicted to be independent, adjusting for other covariates if this was specified in the prediction. For example, a predicted conditional independency generated by DAGitty such as

$$\text{age} \perp \text{physical\_environment} \mid \text{deprivation educational\_attainment}$$

would be tested by calculating the partial correlation coefficient between age and a measure of the local physical environment, adjusted for the Townsend deprivation score and having a degree. These calculations were done using correlation or regression models, depending on the need to adjust for covariates (see section 7.2.2 below, and Appendix X). For simplicity, only continuous or dichotomous measures were used in the initial

calculations. Where a node had more than one relevant available measure (e.g. physical_environment measured by population density or road proximity), the measure with the largest sample size was used, in order to minimise missing data bias. Where the results indicated lack of independence (i.e. partial correlation coefficient >|0.1|; Kline, 2016, p. 240), follow-up regression models were conducted to obtain further detail. Modifications to the structure of the DAG were then considered, if appropriate, via discussion with the thesis supervisors.

7.1.5.2 Total causal effects

The total causal effect of lifetime history of mania/BD on cognitive performance was firstly identified in the DAG using the back-door criterion, as implemented automatically in DAGitty. If the algorithm was able to find at least one minimum sufficient set of measured nodes in the DAG that met the back-door criterion, this information was then used to plan regression- and matching-based analyses to estimate the effect in the UK Biobank dataset. This was estimated separately for each of the five cognitive outcome measures, to allow for the possibility of task-specific variation in the results. For the purpose of comparison, estimation was conducted in several ways:

- Unadjusted regression model in all available participants;
- Unadjusted regression model only in participants who had complete data on all covariates that were to be used in the adjusted models;
- Multiple regression model adjusted for the minimum sufficient covariate set identified by DAGitty;
- Multiple regression model adjusted for the minimum sufficient set plus all other measured common antecedents of exposure and outcome;
- Multiple regression model adjusted for a propensity score created by regressing the mania/BD exposure on background covariates;
- Matched analyses (1:1 and 1:3) using the propensity score to form matched participant sets;
- Weighted regression model using inverse probability weights (IPW) derived from the propensity score;
- Doubly robust models (IPW-weighted regression with additional covariate adjustment or augmented weighting).

Where models included age as a covariate, age squared was also entered, to account for possible curvilinear relationships. The propensity score model was specified in three ways,

and the score that resulted in the best covariate balance (evaluated by comparing descriptive statistics for each covariate between the propensity score-matched samples) was taken forward into the total effects analyses listed above. This decision was based solely on covariate balance, without reference to the cognitive outcome data. The first propensity score model regressed the mania/BD exposure status variable on all ancestors of the exposure and all ancestors of the outcome that were not descended from the exposure (Brookhart et al., 2006). The second propensity score model used the same predictor variables as the first, but also included all pairwise interaction terms. The third approach used boosted regression modelling (a machine learning method) to find the optimum prediction specification (Friedman, 2001), again using the same predictor variables as the other two models. The propensity score was also converted to an inverse probability weight, which was rescaled to sum to 1 (Garrido et al., 2014).

All analyses were conducted using Stata v15 (StataCorp, 2017). The propensity scores were estimated using `psmatch2` and `boost`, and covariate balance was checked using `pstest` and `pbalchk`. The total effects models were estimated using `regress` or `logistic`, `psmatch2` and `teffects`, and results were reported as beta coefficients or average treatment effects (ATE) with 95% confidence intervals (CI) calculated from robust standard errors. For the logistic regression models (prospective memory outcome measure), `adjrr` was used to convert the OR estimates into risk differences, which are comparable to the ATE estimates for the other cognitive measures. The matched models were performed with replacement and used a caliper set at 0.2 SD of the logit of the propensity score (Austin, 2011).

7.1.5.3 Direct and indirect effects

The DAG was used to assess whether indirect effects via various intermediates of interest could be identified. As outlined in Chapter 6, this required all back-door paths between the exposure and the intermediate and between the intermediate and the outcome to be blocked, as well as those between the exposure and the outcome. Where this requirement was satisfied (i.e. covariate adjustment sets could be found), G-computation was used to estimate the natural direct and indirect effects. This was implemented using the Stata package `gformula`, because `gformula` permits effect decomposition in the presence of intermediate confounding. The covariate adjustment sets in these models were the minimum sufficient adjustment sets to block all back-door paths, as determined by

DAGitty. All outcome and intermediate confounder variables were entered in continuous or binary form, as required by `gformula`.

7.1.6 Sensitivity analyses

7.1.6.1 Conditional exchangeability

The assumption of conditional exchangeability implies that, within a matched pair of participants, each participant had equal odds of being exposed ('treated') and unexposed. Covariate balance checks allow this to be verified with respect to measured background factors, but cannot confirm that matched pairs are balanced (i.e. exchangeable) for unmeasured or unknown background variables. Sensitivity of treatment effect estimates or their *p* values to different potential magnitudes of departure from exchangeability can be evaluated quantitatively using 'bounds' methods developed by Rosenbaum (2002). Potential deviations from exchangeability are summarised in a parameter referred to as gamma (where $\Gamma = 1$ represents equal odds), and the value of gamma at which the effect estimate or *p* value crosses the null is ascertained using permutation methods. This was conducted following the propensity score matched models, using the Stata package `rbounds`[12].

7.1.6.2 Missing data

Because the estimation methods used here involved adjustment and/or propensity score estimation for a large number of covariates, results were potentially sensitive to selection bias or reduced power, arising from missing data. Multiple imputation with chained equations was implemented using the `ice` package in Stata, and the regression models for total effects were repeated on the imputed datasets (25 imputations) using the `mi estimate` function. The cognitive outcome variables were included in the imputation model specification (White, Royston, & Wood, 2011), but their original (unimputed) values were analysed in the outcome models. A chained equations imputation option was also implemented in `gformula`, to allow a comparison of the results of the effect decomposition models using raw versus imputed mediator and covariate data. These methods assume missingness at random (MAR), which will be discussed further below.

[12] The `rbounds` package conducts sensitivity analyses on the 'average treatment effect in the treated' (ATT), rather than on the overall ATE. The ATT represents the average effect of treatment/exposure on outcome within the group that actually received the treatment/exposure.

7.1.6.3 Exposure misclassification

The effect of different hypothetical levels of mania/BD exposure misclassification on cognitive outcome was assessed using the Stata package `episens`. The outcome was dichotomised as impaired (z-score ≤ -1.645, i.e. 5th percentile) or not. The range of assumed sensitivity and specificity values for correct mania/BD classification was entered as a trapezoidal function, by specifying minimum and maximum values around a narrower range of equally probable values (e.g. minimum 0.7 and maximum 1.0, around a peak interval of 0.8 to 0.9). Differential misclassification was assumed, on the grounds that cognitively impaired participants would be more likely to be misclassified on the self-reported mania/BD exposure data.

7.1.6.4 Equivalent models

The final DAG that was used as the basis for the total effects and mediation models was analysed structurally in the `dagitty` R package, to determine the number of alternative ways it could be drawn while retaining the same predicted conditional independencies. This ascertains whether there are other specifications of the model that would be statistically indistinguishable from the version that was analysed, i.e. an 'equivalence class' (Textor et al., 2016).

7.2 Results

7.2.1 Characteristics of the sample

Figure 7.1 shows a flowchart of exclusions leading to the final analysis sample, which comprised 2,709 participants with mania/BD and 105,284 comparison participants. A large number of participants were excluded due to missing data in at least one exposure information source, which meant they could not be classified in the comparison group. Where genotyping data indicated relatedness (third degree or closer), one member of each related set was chosen at random for analysis. Ethnic ancestry exclusions were applied only in the adjusted models (see below).

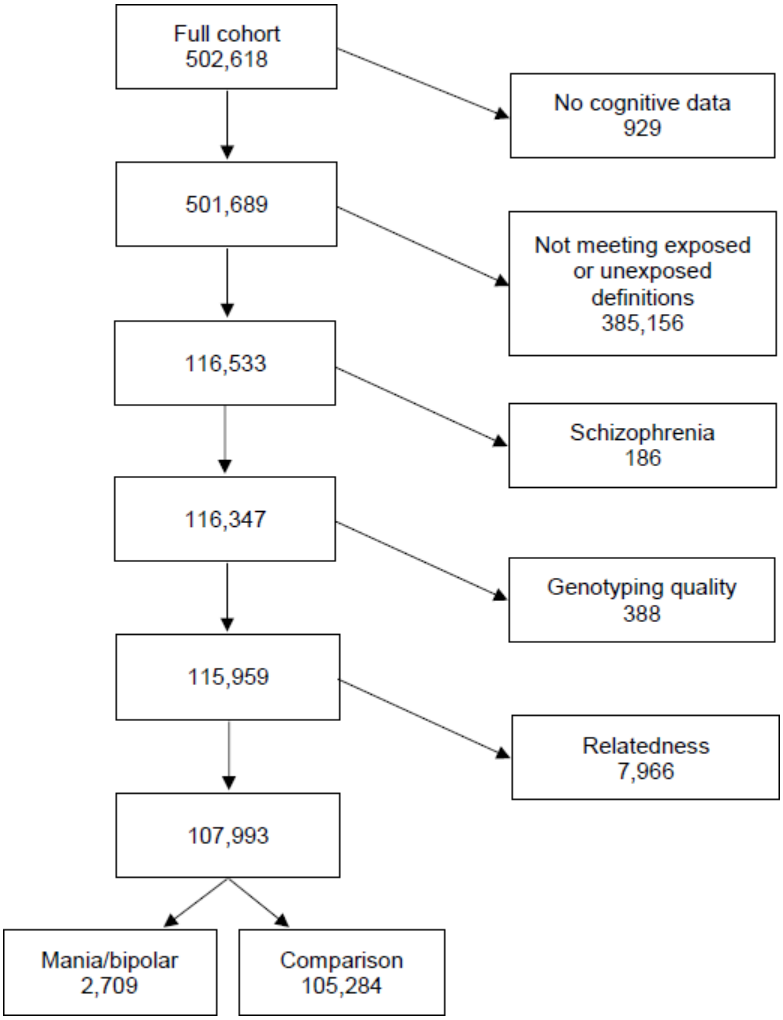


Figure 7.1 - Mania/bipolar disorder analysis sample flowchart

Table 7.1 summarises the cognitive outcome data and Table 7.2 summarises the covariate data, for the mania/BD and comparison groups separately. Descriptive results are also provided separately for the subset of participants who had complete data on the covariates that were used in the maximally-adjusted total effects models.

Table 7.1 indicates that all the cognitive test scores appeared lower on average in the mania/BD group. Scores appeared higher on average in both groups among the participants with complete covariate data. Missing cognitive data was most common on the visuospatial memory test (5.8% and 2.8% in the mania/BD and comparison groups, respectively), and was more common in the mania/BD group on all tests. Reliability (Cronbach's alpha) was higher for reaction time than for reasoning, although both were in the acceptable range. Reliability was similar between the mania/BD and comparison groups, apart from on the reaction time task in the complete covariate data sub-sample, where alpha was lower in the mania/BD group (0.72) than in the comparison group (0.82).

Table 7.1 - Summary of cognitive outcome measures in the mania/bipolar and comparison groups

	All available data		Complete covariate data[a]	
	Mania/BD	**Comparison**	**Mania/BD**	**Comparison**
n	2,709	105,284	504	26,997
Reasoning z-score				
n (%) missing[b]	31 (1.8)	1,627 (1.6)	1 (0.3)	89 (0.3)
Mean (SD)	-0.35 (1.01)	-0.20 (0.97)	0.05 (0.96)	0.12 (0.92)
Cronbach's α	0.71	0.70	0.70	0.68
Reaction time z-score				
n (%) missing	51 (1.9)	1,167 (1.1)	1 (0.2)	72 (0.3)
Mean (SD)	-0.19 (1.01)	-0.03 (0.98)	0.03 (0.94)	0.07 (0.96)
Cronbach's α	0.82	0.82	0.72	0.82
Numeric memory z-score				
n (%) missing[b]	18 (3.9)	755 (2.4)	0 (0.0)	51 (0.7)
Mean (SD)	-0.52 (1.00)	-0.35 (0.94)	-0.23 (1.08)	-0.16 (0.93)
Visuospatial memory z-score				
n (%) missing	157 (5.8)	2,896 (2.8)	10 (1.2)	212 (0.8)
Mean (SD)	0.07 (1.07)	0.26 (1.05)	0.20 (1.09)	0.37 (1.03)
Prospective memory[b,c]				
n (%) correct	1,264 (72.1)	80,502 (76.9)	286 (80.3)	23,112 (85.7)

BD, bipolar disorder; GPS, genome-wide polygenic score; SD, standard deviation.
a. Participants with complete data on all the covariates that were entered into the maximally-adjusted total effects models (age, gender, white British genetic ancestry, English-speaking country of birth, degree, comorbid neurological/psychiatric condition, family history of dementia, family history of Parkinson's disease, family history of severe depression, maternal smoking around birth, childhood trauma, education/cognition GPS, bipolar disorder GPS).
b. Missing data refers only to the period when this measure was included in the battery.
c. No missing data.

Table 7.2 shows a similar pattern with regard to missing data, with the exception of the number of depressed episodes, on which relatively fewer mania/BD than comparison participants had missing responses. Owing to the low response rate on the web-based mental health questionnaire (*n* = 157,366; 31.3% of the whole cohort), the proportion of missing data was highest for the childhood trauma variable. Missingness was also common on the family medical history, maternal smoking, current depressive symptoms and physical activity variables. Table 7.2 indicates that the mania/BD group was younger on average than the comparison group, and had a higher proportion of women and of degree-holders. The participants with mania/BD also appeared to be more likely to live in urban and more deprived areas, and to be current smokers and former drinkers. The proportions with frequent sleeplessness, obesity, cardiometabolic disease, comorbid neurological/psychiatric conditions, family history of severe depression, current psychotropic medication, and history of childhood trauma were higher in the mania/BD group, and this group also reported more depressed episodes and a higher current depressive symptom score on average. The distribution of the education/cognition GPS score appeared to be somewhat different between the mania/BD and comparison groups, with both low (decile 1) and high (decile 10) GPS values being slightly over-represented in the mania/BD group. The distribution of the bipolar GPS score was skewed towards higher values in the mania/BD group. The subset of participants with complete covariate data appeared different from the full analysis sample, being on average younger, more highly educated and from less deprived areas, for example.

Table 7.2 - Summary of covariates in the mania/bipolar and comparison groups

	All available data		Complete covariate data[a]	
	Mania/BD	**Comparison**	**Mania/BD**	**Comparison**
n	2,709	105,284	504	26,997
Sociodemographic				
Age (years)[b]				
Mean (SD)	55.0 (8.1)	57.0 (8.2)	54.3 (7.5)	56.3 (7.9)
Gender[b]				
n (%) female	1,437 (53.1)	52,730 (50.1)	277 (55.0)	14,414 (53.4)
Ethnic group				
n (%) missing	18 (0.7)	411 (0.4)	0 (0.0)	73 (0.3)
White, *n* (%)[c]	2,457 (91.3)	95,463 (91.0)	479 (95.0)	25,708 (95.5)
Asian/Asian British	74 (2.8)	3,697 (3.5)	16 (3.2)	459 (1.7)
Black/Black British	78 (2.9)	3,182 (3.0)	2 (0.4)	322 (1.2)
Chinese	4 (0.2)	474 (0.5)	1 (0.2)	99 (0.4)
Mixed & other background	78 (2.9)	2,057 (2.0)	6 (1.2)	336 (1.3)
White British genetic ancestry				
n (%) missing	104 (3.8)	3,419 (3.3)	0 (0.0)	0 (0.0)
n (%)[c]	1,968 (75.6)	81,183 (79.7)	394 (78.2)	22,606 (83.7)
English-speaking country of birth				
n (%) missing	6 (0.2)	149 (0.1)	0 (0.0)	0 (0.0)
n (%)[c]	2,432 (90.0)	93,802 (89.2)	462 (91.7)	24,993 (92.6)
Has a degree				
n (%) missing	29 (1.1)	1,045 (1.0)	0 (0.0)	0 (0.0)
n (%)[c]	1,034 (38.6)	36,797 (35.3)	265 (52.6)	13,210 (48.9)
Townsend quintile[d]				
n (%) missing	3 (0.1)	162 (0.2)	0 (0.0)	34 (0.1)
Qu1 (least deprived), *n* (%)[c]	334 (12.3)	18,097 (17.2)	78 (15.5)	5,162 (19.1)
Qu2	361 (13.3)	21,344 (20.3)	70 (13.9)	5,846 (21.7)
Qu3	461 (17.0)	21,794 (20.7)	93 (18.5)	5,924 (22.0)
Qu4	607 (22.4)	23,685 (22.5)	124 (24.6)	6,094 (22.6)
Qu5 (most deprived)	943 (34.9)	20,202 (19.2)	139 (27.6)	3,937 (14.6)
Local environment				
Home area population density[e]				
n (%) missing	38 (1.4)	968 (0.9)	5 (1.0)	271 (1.0)
England/Wales urban, *n* (%)[c]	2,311 (86.5)	90,610 (86.9)	433 (86.8)	22,766 (85.2)
England/Wales town	129 (4.8)	6,603 (6.3)	25 (5.0)	1,776 (6.7)
England/Wales village	88 (3.3)	4,967 (4.8)	20 (4.0)	1,533 (5.7)
England/Wales hamlet/isolated	45 (1.7)	2,136 (2.1)	10 (2.0)	651 (2.4)
Scotland large urban	80 (3.0)	0 (0.0)	9 (1.8)	0 (0.0)
Scotland other urban	11 (0.4)	0 (0.0)	1 (0.2)	0 (0.0)
Scotland small town	5 (0.2)	0 (0.0)	1 (0.2)	0 (0.0)
Scotland rural	2 (0.1)	0 (0.0)	0 (0.00	0 (0.0)
Proximity to major road (1/m)				
n (%) missing	65 (2.4)	1,299 (1.2)	9 (1.8)	348 (1.3)
Mean (SD)	0.006 (0.021)	0.006 (0.013)	0.006 (0.010)	0.005 (0.011)
Particulate matter ≤10μm (μg/m^3)				
n (%) missing	72 (2.7)	1,679 (1.6)	11 (2.2)	437 (1.6)
Mean (SD)	22.9 (3.1)	22.8 (3.0)	23.2 (3.2)	22.8 (3.2)
Nitrogen dioxide (μg/m^3)				
n (%) missing	65 (2.4)	1,299 (1.2)	9 (1.8)	348 (1.3)
Mean (SD)	32.9 (11.0)	31.9 (10.6)	33.1 (11.7)	31.8 (11.0)
Lifestyle and physical				
Smoking status				
n (%) missing	15 (0.6)	380 (0.4)	1 (0.2)	47 (0.2)
Never, *n* (%)[c]	1,185 (44.0)	60,305 (57.5)	251 (49.9)	16,428 (61.0)
Former	922 (34.2)	35,436 (33.8)	175 (34.8)	8,885 (33.0)
Current	587 (21.8)	9,163 (8.7)	77 (15.3)	1,637 (6.1)

	All available data		Complete covariate data[a]	
	Mania/BD	**Comparison**	**Mania/BD**	**Comparison**
Alcohol frequency				
n (%) missing	12 (0.4)	80 (0.1)	1 (0.2)	4 (0.01)
Daily/almost daily, *n* (%)[c]	497 (18.4)	22,179 (21.1)	112 (22.3)	6,514 (24.1)
3-4 times per week	451 (16.7)	24,718 (23.5)	98 (19.5)	7,176 (26.6)
1-2 times per week	583 (21.6)	26,774 (25.5)	114 (22.7)	6,627 (24.6)
1-3 times per month	312 (11.6)	11,397 (10.8)	63 (12.5)	2,872 (10.6)
Special occasions only	432 (16.0)	11,824 (11.2)	66 (13.1)	2,387 (8.8)
Never (former drinker)	261 (9.7)	3,186 (3.0)	36 (7.2)	620 (2.3)
Never (not former drinker)	161 (6.0)	5,126 (4.9)	14 (2.8)	797 (3.0)
Sleeplessness				
n (%) missing	2 (0.1)	79 (0.1)	0 (0.0)	14 (0.1)
Never/rarely, *n* (%)[c]	535 (19.8)	29,205 (27.8)	113 (22.4)	8, 019 (29.7)
Sometimes	1,199 (44.3)	50,293 (47.8)	211 (41.9)	12,802 (47.4)
Usually	973 (35.9)	25,707 (24.4)	180 (35.7)	6,162 (22.8)
Physical activity (MET h/week)				
n (%) missing	256 (9.5)	6,861 (6.5)	31 (6.2)	1,054 (3.9)
Median (Q1, Q3)	25.6 (11.6, 56.7)	29.8 (13.7, 60.1)	27.5 (12.3, 54.9)	29.1 (14.2, 55.7)
Body mass index				
n (%) missing	27 (1.0)	741 (0.7)	1 (0.2)	90 (0.3)
Underweight, *n* (%)[c]	16 (0.6)	514 (0.5)	3 (0.6)	156 (0.6)
Normal	719 (26.8)	34,893 (33.4)	174 (34.6)	10,751 (40.0)
Overweight	1,043 (38.9)	44,979 (43.0)	210 (41.8)	11,104 (41.3)
Obese class I	610 (22.7)	17,804 (17.0)	87 (17.3)	3,750 (13.9)
Obese class II	205 (7.6)	4,729 (4.5)	20 (4.0)	860 (3.2)
Obese class III	89 (3.3)	1,624 (1.6)	9 (1.8)	286 (1.1)
Medical and family history				
Cardiometabolic disease				
n (%) missing	11 (0.4)	178 (0.2)	0 (0.0)	22 (0.1)
n (%)[c]	986 (36.6)	32,120 (30.6)	140 (27.8)	6,310 (23.4)
Comorbid neurological or psychiatric condition[f]				
n (%)	814 (30.1)	9,430 (9.0)	120 (23.8)	2,093 (7.8)
Family history of dementia				
n (%) missing	473 (17.5)	14,799 (14.1)	0 (0.0)	0 (0.0)
n (%)[c]	384 (17.2)	15,755 (17.4)	83 (16.5)	4,661 (17.3)
Family history of Parkinson's disease				
n (%) missing	589 (21.8)	16,406 (15.6)	0 (0.0)	0 (0.0)
n (%)[c]	107 (5.1)	4,250 (4.8)	18 (3.6)	1,276 (4.7)
Family history of severe depression				
n (%) missing	466 (17.2)	15,655 (14.9)	0 (0.0)	0 (0.0)
n (%)[c]	874 (39.0)	10,503 (11.7)	170 (33.7)	3,086 (11.4)
Maternal smoking around birth				
n (%) missing	390 (14.4)	13,420 (12.7)	0 (0.0)	0 (0.0)
n (%)[c]	716 (30.9)	24,807 (27.0)	151 (30.0)	7,065 (26.2)
Mental health				
Current depressive symptoms				
n (%) missing	265 (9.8)	8,758 (8.3)	27 (5.4)	1,223 (4.5)
Mean (SD)	3.5 (3.2)	1.2 (1.7)	2.8 (2.9)	1.0 (1.4)
Any psychotropic medication				
n (%) missing	38 (1.4)	1,247 (1.2)	5 (1.0)	286 (1.1)
n (%)[c]	1,457 (54.6)	2,647 (2.5)	221 (44.3)	503 (1.9)
Number of depressed episodes				
n (%) missing	127 (4.7)	6,573 (6.2)	23 (4.6)	1,448 (5.4)
Median (Q1, Q3)	1 (0, 6)	0 (0, 1)	4 (2, 11)	0 (0, 1)
Any childhood trauma[g]				
n (%) missing	1,976 (72.9)	68,715 (65.3)	0 (0.0)	0 (0.0)
n (%)[c]	476 (64.9)	16,015 (43.8)	317 (62.9)	11,240 (41.6)

	All available data		Complete covariate data[a]	
	Mania/BD	**Comparison**	**Mania/BD**	**Comparison**
Genome-wide polygenic scores				
Education/cognition GPS decile[d]				
n (%) missing	104 (3.8)	3,419 (3.3)	0 (0.0)	0 (0.0)
D1 (lowest), *n* (%)[c]	291 (11.2)	10,156 (10.0)	36 (7.14)	2,262 (8.4)
D2	259 (9.9)	10,188 (10.0)	57 (11.3)	2,439 (9.0)
D3	281 (10.8)	10,166 (10.0)	51 (10.1)	2,497 (9.3)
D4	234 (9.0)	10,213 (10.0)	55 (10.9)	2,619 (9.7)
D5	245 (9.4)	10,202 (10.0)	51 (10.1)	2,599 (9.6)
D6	244 (9.4)	10,203 (10.0)	48 (9.5)	2,666 (9.9)
D7	251 (9.6)	10,196 (10.0)	42 (8.3)	2,822 (10.5)
D8	257 (9.9)	10,190 (10.0)	45 (8.9)	2,878 (10.7)
D9	246 (9.4)	10,201 (10.0)	54 (10.7)	3,014 (11.2)
D10 (highest)	297 (11.4)	10,150 (10.0)	65 (12.9)	3,201 (11.9)
Bipolar disorder GPS decile[d]				
n (%) missing	104 (3.8)	3,419 (3.3)	0 (0.0)	0 (0.0)
D1 (lowest), *n* (%)[c]	199 (7.6)	10,248 (10.1)	39 (7.7)	2,773 (10.3)
D2	225 (8.6)	10,222 (10.0)	45 (8.9)	2,698 (10.0)
D3	232 (8.9)	10,215 (10.0)	47 (9.3)	2,696 (10.0)
D4	228 (8.8)	10,219 (10.0)	47 (9.3)	2,654 (9.8)
D5	228 (8.8)	10,219 (10.0)	36 (7.1)	2,707 (10.0)
D6	264 (10.1)	10,183 (10.0)	46 (9.1)	2,638 (9.8)
D7	274 (10.5)	10,173 (10.0)	49 (9.7)	2,707 (10.0)
D8	262 (10.1)	10,185 (10.0)	48 (9.5)	2,706 (10.0)
D9	306 (11.8)	10,141 (10.0)	63 (12.5)	2,717 (10.1)
D10 (highest)	387 (14.9)	10,060 (9.9)	84 (16.7)	2,701 (10.0)

BD, bipolar disorder; D, decile; GPS, genome-wide polygenic score; MET, metabolic equivalent of task; Q, quartile; Qu, quintile; SD, standard deviation.

a. Participants with complete data on all the covariates that were entered into the maximally-adjusted total effects models (age, gender, white British genetic ancestry, English-speaking country of birth, degree, comorbid neurological/psychiatric condition, family history of dementia, family history of Parkinson's disease, family history of severe depression, maternal smoking around birth, childhood trauma, education/cognition GPS, bipolar disorder GPS).

b. No missing data.

c. Missing excluded from denominator.

d. Based on data distribution in the whole UK Biobank cohort.

e. Scottish psychiatric hospital records were unavailable, which meant no Scotland-based participants could be classified in the comparison group; therefore all locations for comparison participants are in England/Wales.

f. Apart from mood disorder or schizophrenia; not possible to distinguish between missing data and self-report of no condition, therefore both classified as No.

g. From the web-based questionnaire, which was completed by 157,366 (31.3%) of the cohort.

7.2.2 Evaluation of the graphical model

The original DAG is shown in Figure 7.2, and Appendix Y explains the correspondence between each node and the available measures in UK Biobank, along with the rationale for the key assumptions made about the relationships between the nodes. This DAG postulated that educational attainment was a causal antecedent of mania/BD status, but a plausible alternative specification would show the arrow in reverse such that mania/BD causally influences educational attainment (for example, if illness onset occurs at a young age). The different predicted independencies implied by both specifications were tested. The results indicated poorer fit in the second specification, with a greater proportion of the partial correlation coefficients being above |0.1|. The first specification also showed poor fit involving certain nodes, particularly current psychotropic medication use. The reasons for poor fit (e.g. model misspecification, measurement error, selection bias) cannot be discerned from the data alone, but consideration was given to whether the graph should be modified by adding new nodes or paths.

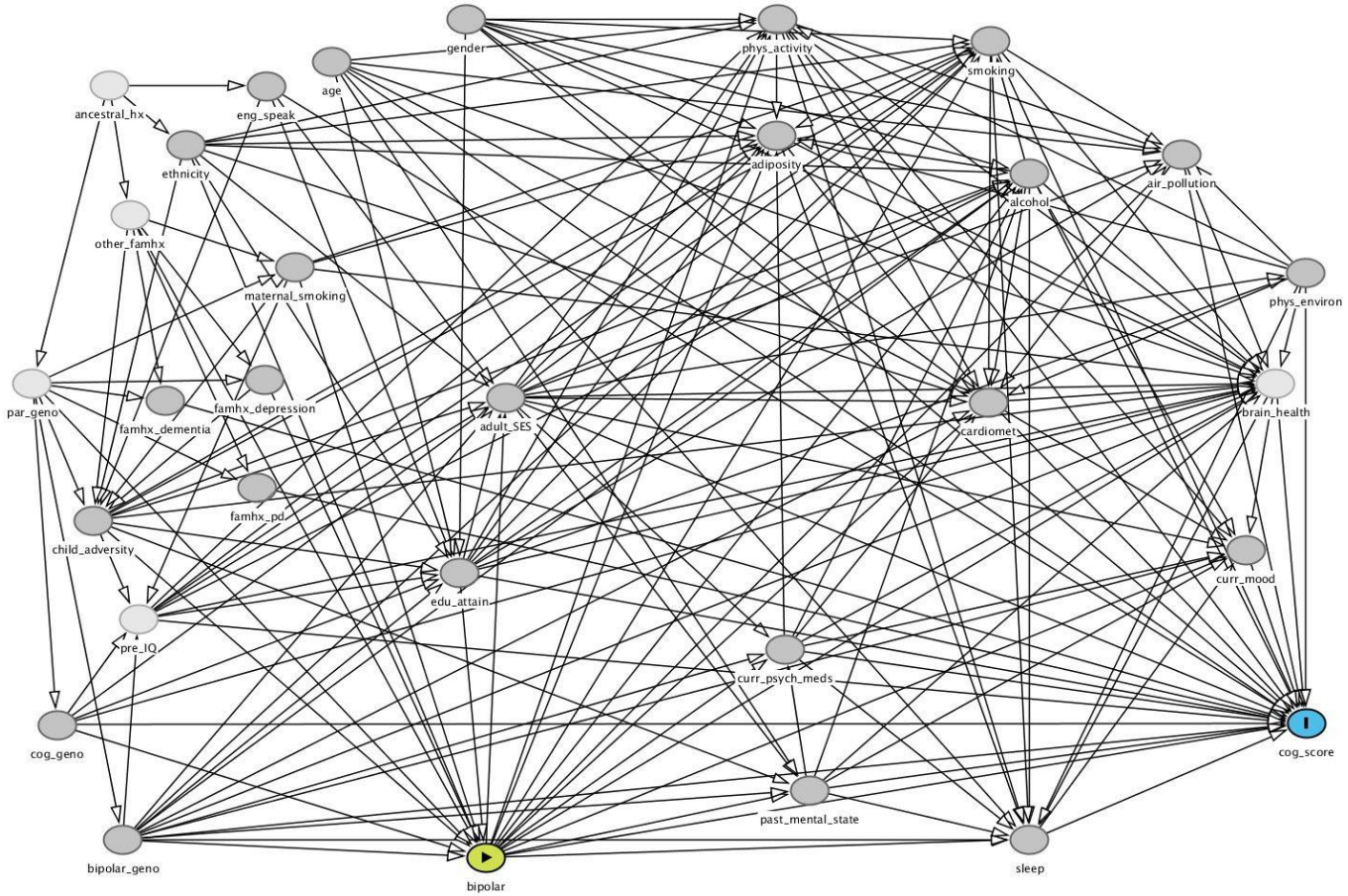


Figure 7.2 - Original directed acyclic graph for the effect of mania/bipolar disorder on cognitive outcome

Cardiomet, cardiometabolic disease; cog, cognitive; curr, current; edu, educational; eng_speak, English speaking birth country; famhx, family history; geno, genotype; hx, history; par, parental; PD, Parkinson's disease; phys, physical; pre_IQ, premorbid intelligence; psych_meds, psychotropic medications; SES, socioeconomic status. Node labels and meanings are explained in detail in Appendix Y. Green node is the exposure and blue node is the outcome. Light nodes represent unmeasured constructs and darker nodes represent measured constructs.

It was deemed plausible that paths should be added between educational attainment and current psychotropic medication, and between gender and current psychotropic medication (reflecting possible influences of, for example, knowledge or attitudes on likelihood of seeking or accepting treatment). A new node was also added, representing other psychiatric or neurological conditions (apart from mood disorder or schizophrenia); this was in line with the results reported in Chapter 5, which had indicated different patterns of cognitive impairment in participants with and without psychiatric or neurological comorbidities. The implied independencies of this modified DAG were then tested, firstly with educational attainment specified as a causal antecedent of both mania/BD and other psychiatric/neurological conditions (Figure 7.3), and then with educational attainment specified as a causal consequence of both.

The best fit was observed in the first version of the modified DAG, in which educational attainment was specified as a causal antecedent of both mania/BD and other psychiatric/neurological conditions (see Appendix X). Twenty percent (28 of 138) of the partial correlation coefficients remained above |0.1|, but most of these were below |0.2| and the largest coefficient was |0.43|. The largest coefficients again involved the current psychotropic medication node. When evaluated further, however, regression models indicated wide variance in the estimated associations between the pairs of nodes, including the null in most instances (Appendix X). No further modifications were made to the DAG; it was decided that the diagram broadly reflected the evidence-based assumptions drawn from the literature and expert knowledge, and additional data-driven modifications may have been misleading if they were in fact influenced by issues such as measurement error rather than model misspecification. The model depicted in Figure 7.3 was used to plan the causal effect models reported below.

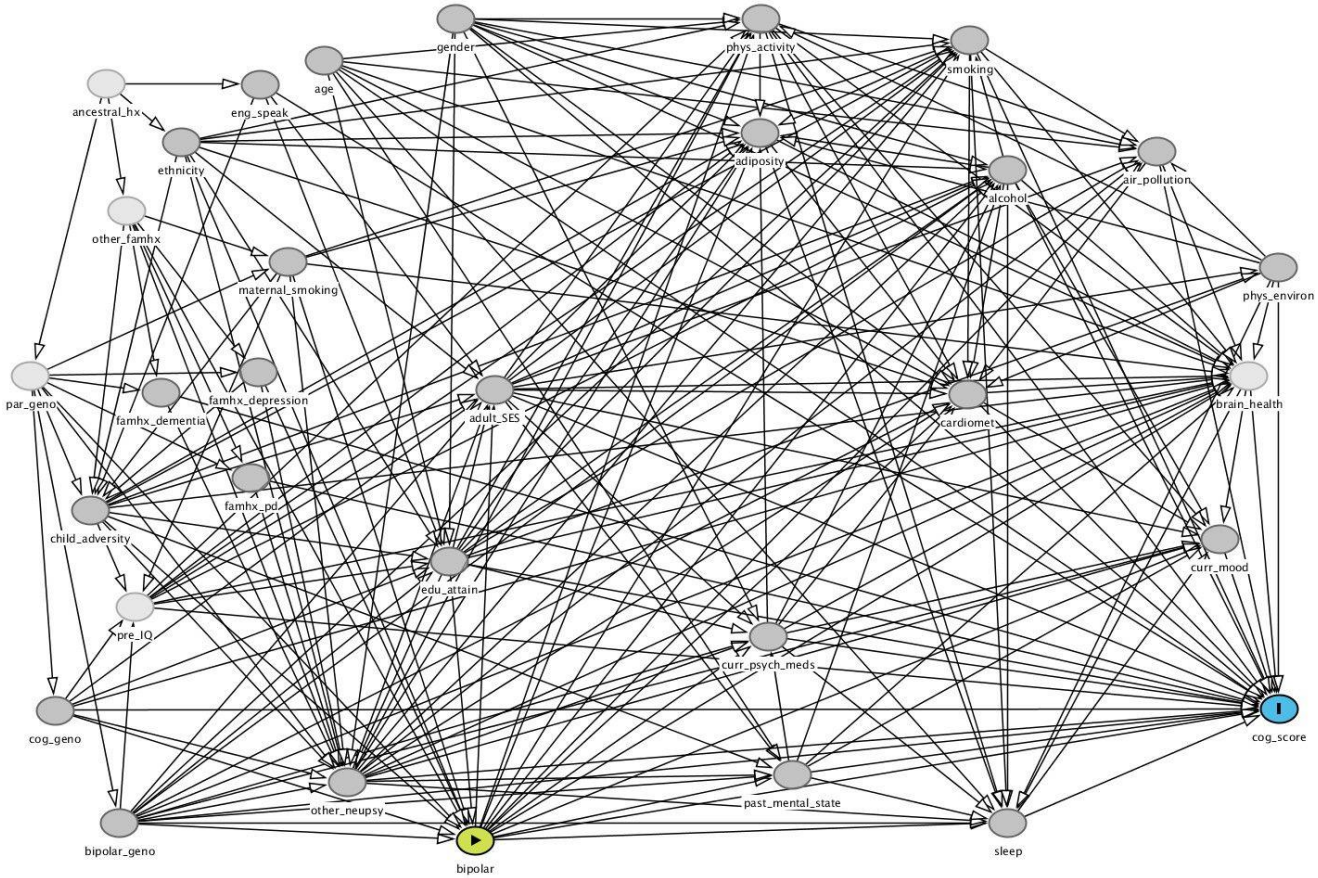


Figure 7.3 - Modified directed acyclic graph for the effect of mania/bipolar disorder on cognitive outcome

Cardiomet, cardiometabolic disease; cog, cognitive; curr, current; edu, educational; eng_speak, English speaking birth country; famhx, family history; geno, genotype; hx, history; other_neupsy, other neurological/psychiatric condition; par, parental; PD, Parkinson's disease; phys, physical; pre_IQ, premorbid intelligence; psych_meds, psychotropic medications; SES, socioeconomic status. Node labels and meanings are explained in detail in Appendix Y. Green node is the exposure and blue node is the outcome. Light nodes represent unmeasured constructs and darker nodes represent measured constructs. This graph and the underlying code are publicly accessible online at dagitty.net/mSTG_SM

7.2.3 Total causal effects

Applying d-separation criteria to the DAG indicated that the minimum sufficient adjustment set for the total effect of mania/BD on cognitive performance comprised gender, educational attainment, English-speaking birth country, ethnicity, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, and other psychiatric/neurological conditions. The extended adjustment set encompassed additional measured nodes that were antecedents of the exposure, namely age and family history of depression. This extended adjustment set was also used as the predictor set for the propensity score model. Ethnicity was accounted for in all the multivariable analyses and in the propensity score estimation by restricting these to participants of white British genetic ancestry. The GPS scores were residualised as described in section 7.1.4.6 above, and were entered as deciles, based on the distribution in the full analysis sample.

The best propensity score model, in terms of covariate balance, was the first model with no interaction terms. This was used in all the outcome models that involved propensity score adjustment or matching, and was used as the basis for the inverse probability weights. Table 7.3 shows the degree of covariate balance, as illustrated within the matched samples used in the reaction time analyses reported below.

Figures 7.4 to 7.8 show the results of all the total effects models for each of the five cognitive scores. Only the visuospatial memory test (and, more equivocally, the prospective memory test) indicated a detrimental effect of mania/BD that remained evident in the multivariable models. The effect sizes were small: the mania/BD group scored approximately 0.2 SD lower than the unexposed comparison group on the visuospatial memory test, and the proportion of the mania/BD group succeeding on the prospective memory task was lower by approximately 5 percentage points (approximately 82% in the mania/BD group versus 87% in the unexposed group). The visuospatial and prospective memory estimates showed little change between the unadjusted and adjusted/matched models, whereas the estimates for the other three cognitive measures generally attenuated towards the null.

Table 7.3 - Summary of covariates in matched mania/bipolar and comparison groups

	Mania/BD	Comparison
	Mean (SD)	
Age (years)	54.9 (7.4)	55.1 (7.4)
	%	
Female gender	52.9	52.0
English-speaking country of birth	97.7	98.3
Has a degree	52.4	55.2
Comorbid neurological or psychiatric condition	25.2	25.1
Family history of dementia	18.6	17.5
Family history of Parkinson's disease	3.6	4.3
Family history of severe depression	34.6	36.2
Maternal smoking around birth	33.1	34.5
Any childhood trauma	60.8	60.2
Education/cognition GPS		
D1 (lowest)	7.5	7.1
D2	9.9	8.6
D3	10.2	8.0
D4	10.7	12.3
D5	10.4	9.9
D6	9.2	10.3
D7	8.7	7.8
D8	8.7	10.0
D9	11.5	11.1
D10 (highest)	13.2	14.9
Bipolar disorder GPS		
D1 (lowest)	7.8	6.6
D2	9.7	10.3
D3	7.6	8.5
D4	9.7	9.3
D5	7.1	7.8
D6	8.9	9.9
D7	10.2	10.2
D8	9.4	9.5
D9	11.5	11.4
D10 (highest)	18.1	16.5

BD, bipolar disorder; D, decile; GPS, genome-wide polygenic score; SD, standard deviation.
These results are from the propensity-score matched samples used in the 1:3 matched model for the total effect of mania/BD on reaction time (mania/BD *n* = 392; comparison *n* = 1,081).

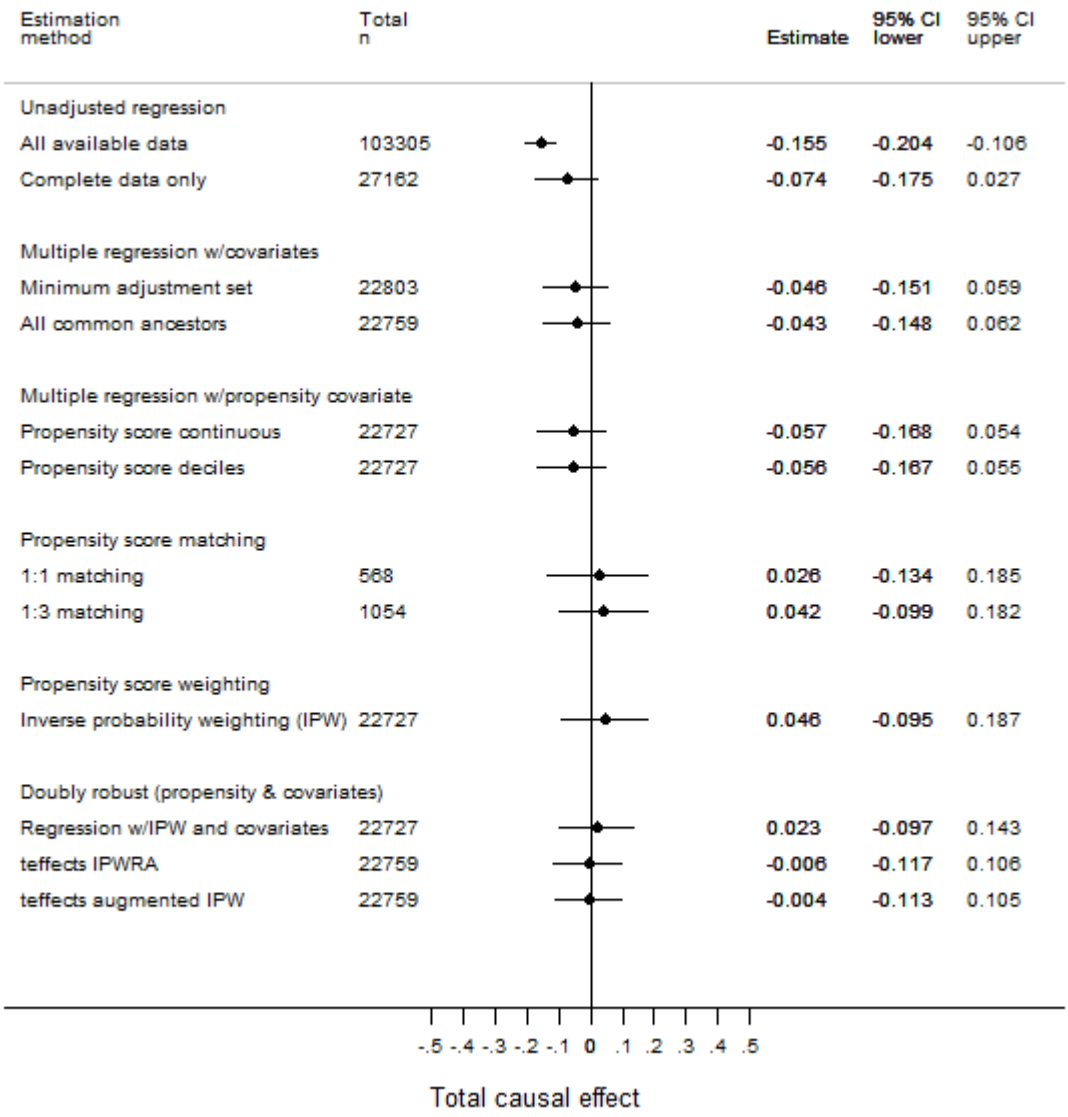


Figure 7.4 - Total causal effect of mania/bipolar disorder on reasoning

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

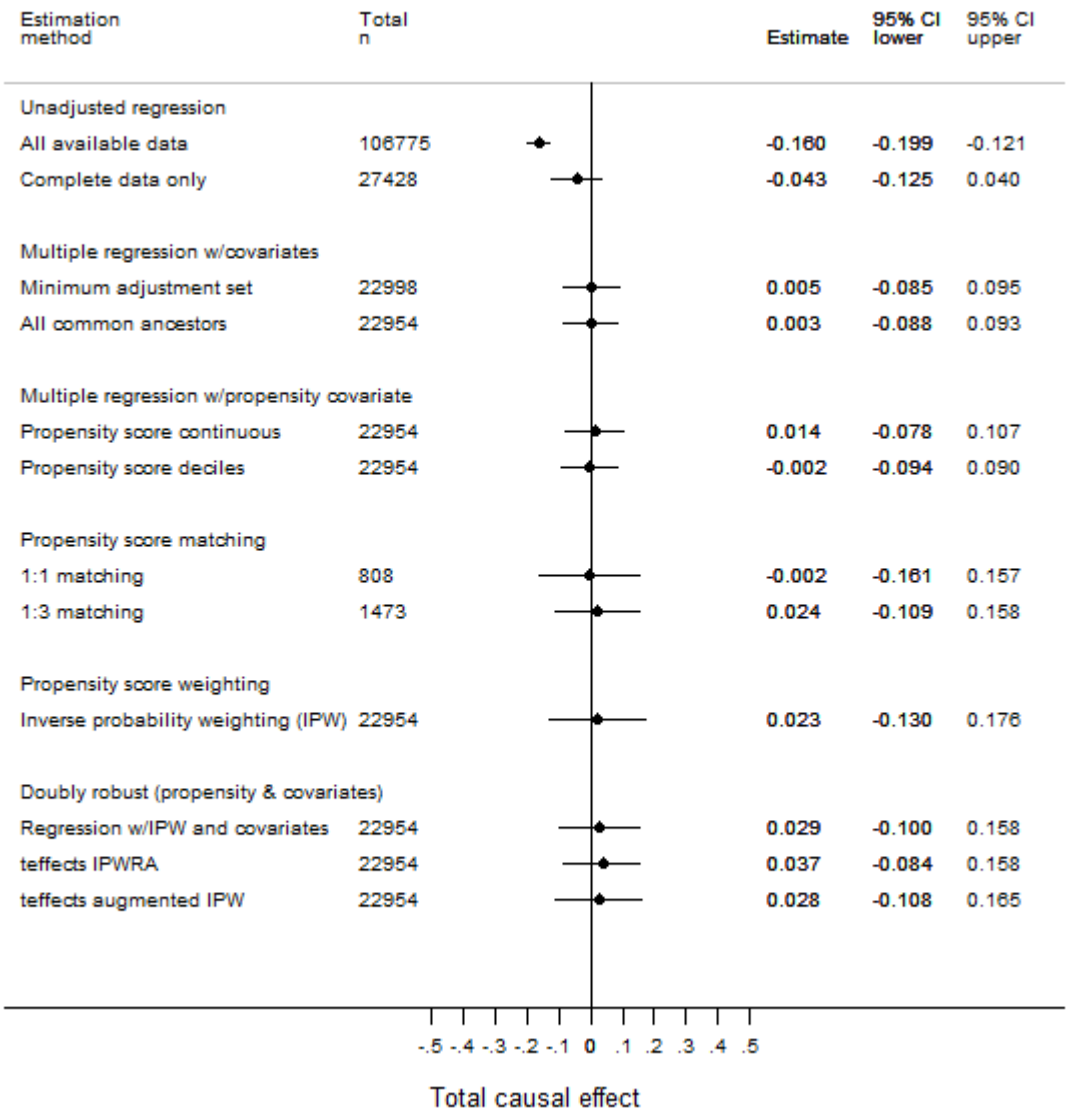


Figure 7.5 - Total causal effect of mania/bipolar disorder on reaction time

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

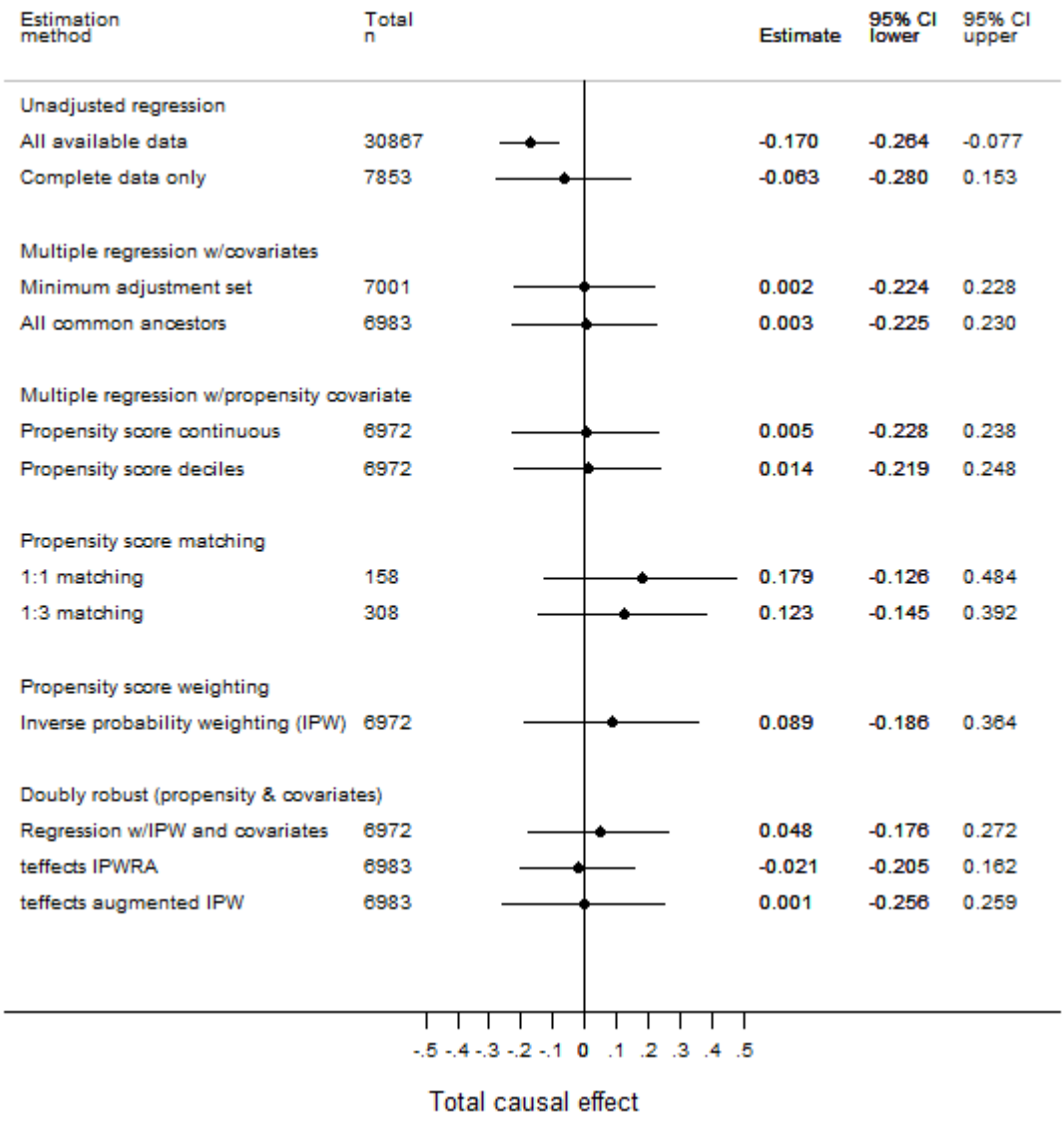


Figure 7.6 - Total causal effect of mania/bipolar disorder on numeric memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

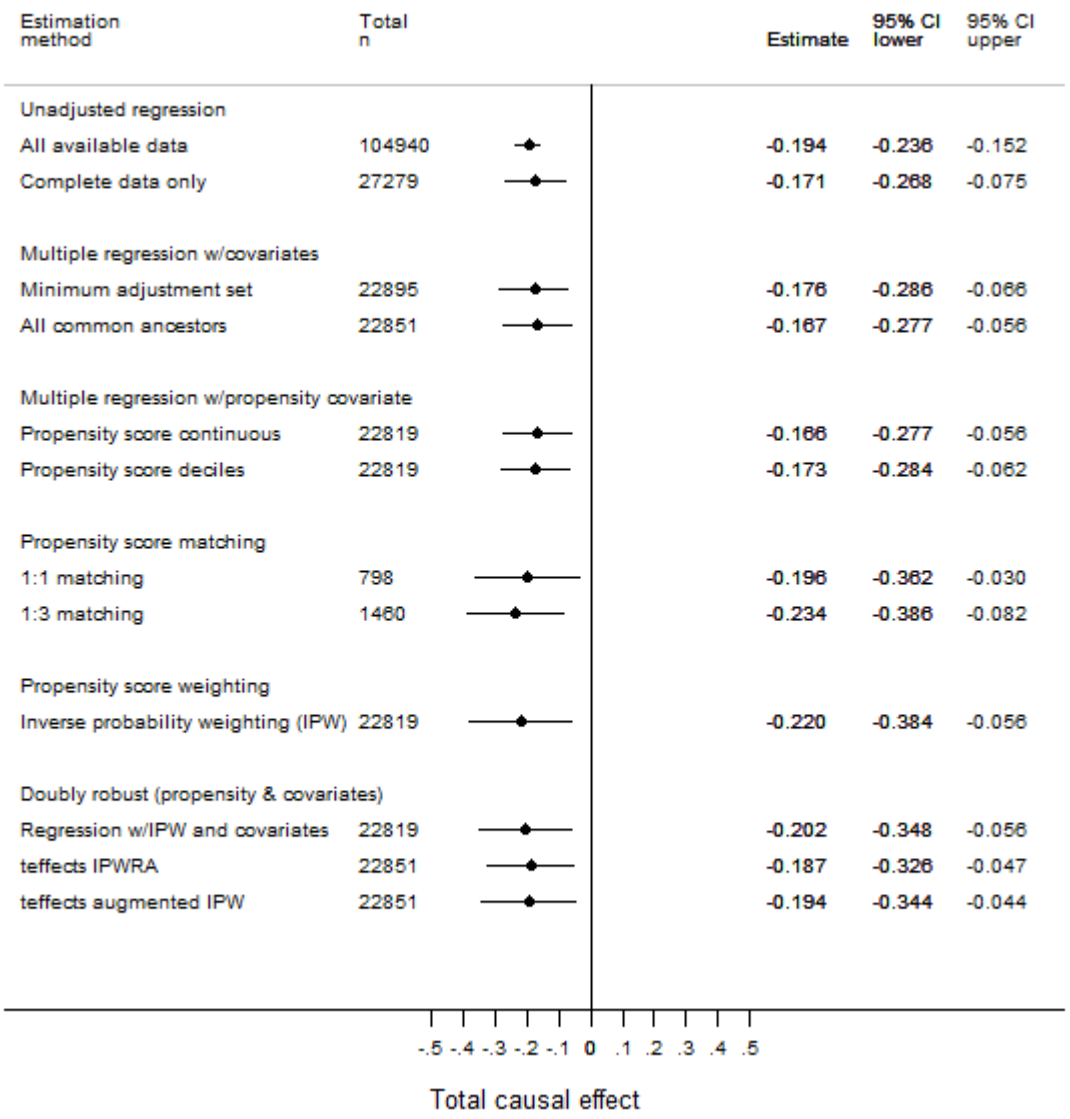


Figure 7.7 - Total causal effect of mania/bipolar disorder on visuospatial memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

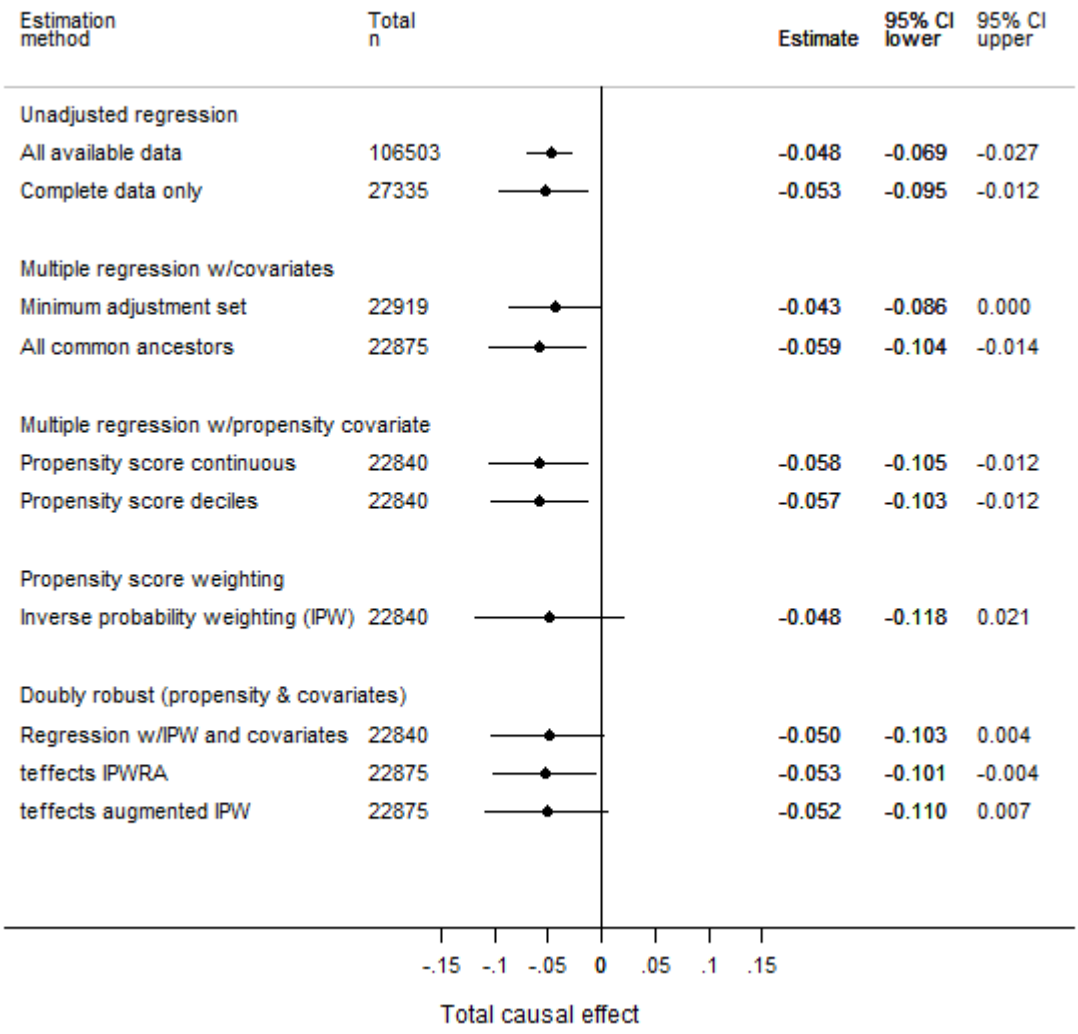


Figure 7.8 - Total causal effect of mania/bipolar disorder on prospective memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are proportions and can be interpreted as risk differences. No estimates are provided from propensity-score matched models as it was not possible to express these as risk differences.

7.2.4 Direct and indirect effects

Structural analysis of the DAG indicated that direct and indirect effects could be decomposed for two potentially modifiable intermediate variables: cardiometabolic disease and psychotropic medication. Effects via other intermediate nodes of interest (e.g. current depressive symptoms) could not be identified, because no covariate adjustment set could be found for the relevant exposure-mediator and/or mediator-outcome paths.

Mediation analyses were conducted to quantify the proportion of the total causal effect of mania/BD on each cognitive outcome that was transmitted via (1) cardiometabolic disease and (2) psychotropic medication. In both sets of analyses, the indirect pathways were affected by intermediate confounding (i.e. the mediator-outcome path was confounded by at least one node that descended from mania/BD), and so the analyses were conducted under the identifying assumption of no interaction between mania/BD and either cardiometabolic disease or psychotropic medication (De Stavola et al., 2015). This assumption was checked by conducting a regression model of each cognitive outcome on mania/BD exposure status, the mediator and all the covariates, including a product term for mania/BD * mediator; there was no evidence of interaction in any of the models (see Table Z.1 in Appendix Z). The mediation model estimates are interpreted as 'randomised interventional analogs' of the natural direct and indirect effects (VanderWeele et al., 2014), as noted in Chapter 6.

7.2.4.1 Cardiometabolic disease

Table 7.4 shows the direct and indirect effect estimates via cardiometabolic disease (binary indicator of any of the following: self-reported diagnosis of angina, hypertension or non-gestational diabetes, or adjudicated diagnosis of myocardial infarction or stroke). The model covariates were age, gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), BMI, alcohol frequency, smoking status, physical activity, and psychotropic medication. The analyses were restricted to participants of white British genetic ancestry, and the residualised GPS scores were entered as deciles. There was no evidence of substantive indirect effects via cardiometabolic disease in any of the models.

Table 7.4 - Mediation of the effect of mania/bipolar disorder on cognitive outcome via cardiometabolic disease

	n	Estimate	95% CI[a]
Reasoning[b]	21,043		
TCE		-0.049	-0.157, 0.059
CDE		-0.048	-0.156, 0.061
NDE		-0.045	-0.153, 0.063
NIE		-0.004	-0.020, 0.013
Reaction time[b]	21,213		
TCE		0.014	-0.082, 0.111
CDE		0.034	-0.063, 0.132
NDE		0.007	-0.090, 0.103
NIE		0.008	-0.011, 0.026
Numeric memory[b]	6,396		
TCE		0.047	-0.213, 0.306
CDE		0.021	-0.237, 0.280
NDE		0.023	-0.235, 0.281
NIE		0.024	-0.008, 0.055
Visuospatial memory[b]	21,124		
TCE		-0.166	-0.285, -0.047
CDE		-0.160	-0.277, -0.044
NDE		-0.164	-0.282, -0.046
NIE		-0.002	-0.022, 0.017
Prospective memory[c]	21,139		
TCE		-0.033	-0.078, 0.012
CDE		-0.038	-0.083, 0.006
NDE		-0.038	-0.083, 0.007
NIE		0.005	-0.002, 0.012

CDE, controlled direct effect when cardiometabolic disease = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; NO_2, nitrogen dioxide; PM_{10}, particulate matter of up to 10μm diameter; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for age, gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), body mass index, alcohol frequency, smoking status, physical activity, and psychotropic medication.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

7.2.4.2 Psychotropic medication

Table 7.5 shows the direct and indirect effect estimates via psychotropic medication (binary variable for any mood stabiliser, antidepressant, antipsychotic, sedative or hypnotic). The model covariates were gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia. The analyses were restricted to participants of white British genetic ancestry, and the residualised GPS scores were entered as deciles.

There was evidence that the previously noted detrimental effect of mania/BD on visuospatial memory was indirectly transmitted via psychotropic medication: of the estimated total effect of -0.19 SD units (95% CI -0.31, -0.08), approximately one quarter was mediated via psychotropic medication (-0.05; 95% CI -0.09, -0.01). Indirect effects were also evident in the reasoning, reaction time and prospective memory models, although the total effects estimates in these models did not show reliable decrements for mania/BD.

Table 7.5 - Mediation of the effect of mania/bipolar disorder on cognitive outcome via psychotropic medication

	n	Estimate	95% CI[a]
Reasoning[b]	21,339		
TCE		-0.065	-0.176, 0.047
CDE		0.016	-0.098, 0.130
NDE		-0.015	-0.130, 0.099
NIE		-0.049	-0.077, -0.021
Reaction time[b]	21,518		
TCE		-0.005	-0.103, 0.094
CDE		0.073	-0.036, 0.182
NDE		0.036	-0.072, 0.144
NIE		-0.040	-0.075, -0.005
Numeric memory[b]	6,547		
TCE		0.033	-0.203, 0.270
CDE		0.058	-0.181, 0.296
NDE		0.075	-0.163, 0.312
NIE		-0.041	-0.097, 0.015
Visuospatial memory[b]	21,424		
TCE		-0.194	-0.311, -0.077
CDE		-0.098	-0.220, 0.024
NDE		-0.140	-0.263, -0.018
NIE		-0.054	-0.094, -0.013
Prospective memory[c]	21,436		
TCE		-0.037	-0.082, 0.009
CDE		-0.028	-0.072, 0.016
NDE		-0.018	-0.062, 0.026
NIE		-0.019	-0.030, -0.007

CDE, controlled direct effect when psychotropic medication = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

7.2.5 Sensitivity analyses

7.2.5.1 Conditional exchangeability

Rosenbaum bounds were calculated to check the sensitivity of the visuospatial memory total effects result to departures from exchangeability. The estimated effect crossed the null at a gamma value of 1.2, which corresponds to the probability of being in the exposed group being approximately 0.45 or 0.55, rather than 0.5 as would be the case if the groups were truly exchangeable. This indicates that the results would not be robust to an unmeasured confounder with even a weak association with group membership.

7.2.5.2 Missing data

There was evidence of missing data bias on several of the outcome measures: much of the attenuation towards the null seen in Figures 7.4 to 7.6 above occurred prior to any multivariable adjustment/matching, simply by restricting the sample to participants with complete covariate data. When the multiple linear regression models for total causal effects were repeated after multiple imputation of the covariate values, the results showed less attenuation (Figure 7.9 below). The effect size estimates in these reasoning, reaction time and numeric memory models were of small magnitude (point estimates approximately -0.10 to -0.15). The estimates in the visuospatial memory models remained similar regardless of which estimation method was used (approximate point estimate -0.19).

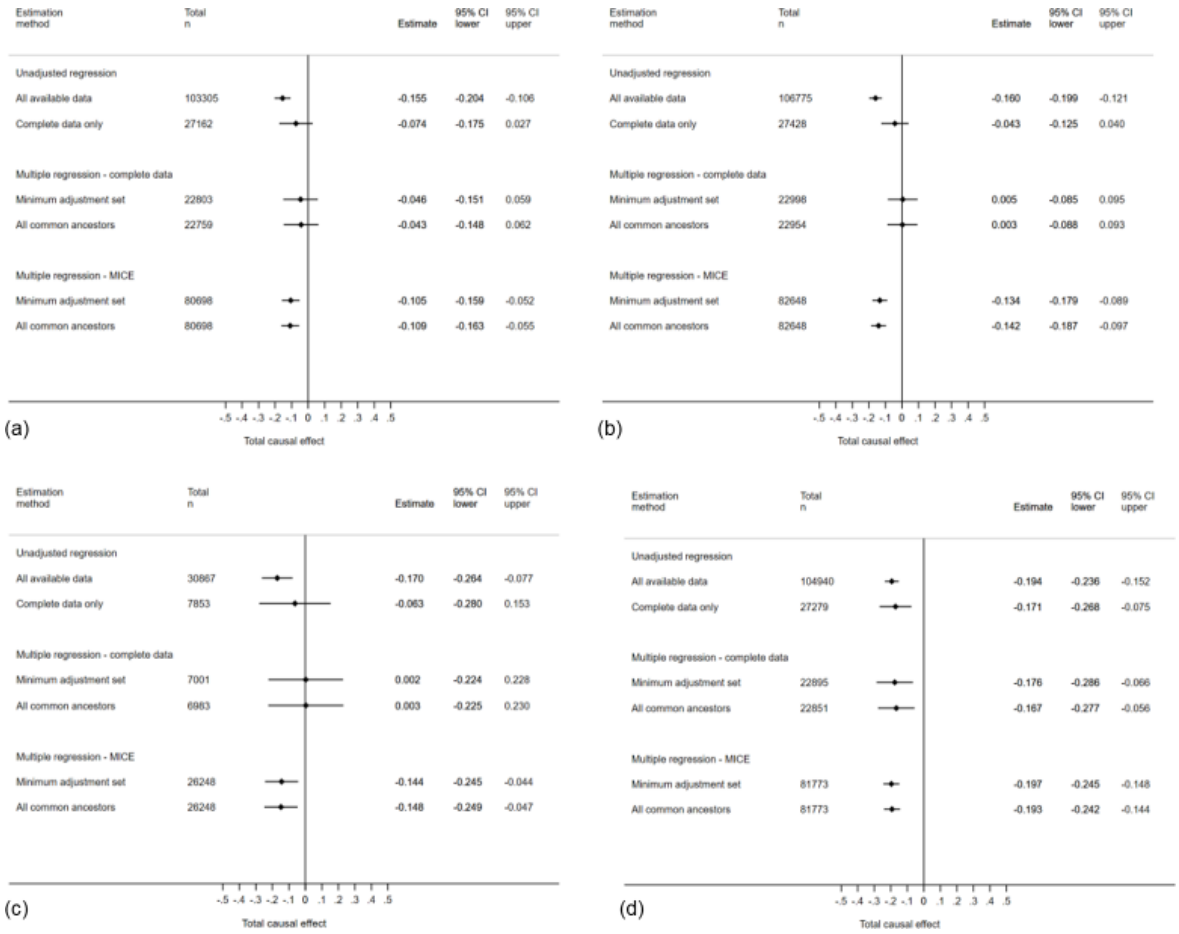


Figure 7.9 - Comparison of missing data approaches in mania/bipolar disorder total effects analyses

CI, confidence interval; MICE, multiple imputation with chained equations. Panels show: (a) reasoning, (b) reaction time, (c) numeric memory and (d) visuospatial memory. Prospective memory not shown as it was not possible to calculate risk differences.

When the mediation models were repeated with imputation of missing mediator and covariate values, the proportion of the total effect transmitted indirectly via cardiometabolic disease remained negligible (Table Z.2 in Appendix Z), whereas there was evidence of indirect effects via psychotropic medication for all the cognitive outcomes (Table Z.3 in Appendix Z).

7.2.5.3 Exposure misclassification

The results of the probabilistic analysis using `episens` indicated that the total effects estimate for visuospatial memory is likely to be sensitive to exposure misclassification. When dichotomised into impaired and unimpaired outcome categories, and assuming no exposure misclassification, the unadjusted relative risk of impairment was 1.70 in the mania/BD group (95% CI 1.45, 2.01). Assuming lower sensitivity to true mania/BD status

among the cognitively impaired (sensitivity range 0.6 to 0.9) versus unimpaired participants (sensitivity range 0.7 to 1.0), the relative risk was estimated as 1.81 (0.22, 11.44).

7.2.5.4 Equivalent models

DAGitty determined that there were six other DAGs that were equivalent to the DAG shown in Figure 7.3 above. In each of these, path directions involving the ancestral history or parental genotype nodes were reversed (see Appendix AA). None of these alternative configurations was causally plausible, owing to temporal order constraints (e.g. it is not possible for parental genotype to be causally influenced by offspring genotype). DAGitty also generated minimum sufficient covariate adjustment sets for the equivalent DAGs, for the total effect of mania/BD on cognitive outcome. The same minimum adjustment set was valid for the analysed DAG and for the six equivalent DAGs. This indicates that the multivariable analyses reported above remain valid, regardless of which model within the equivalence class is correct.

7.3 Discussion

The total causal effect of mania/BD on cognitive performance in the UK Biobank cohort was small, and was principally evident on a test of short-term visuospatial memory. There was evidence of an indirect causal pathway through psychotropic medication, but not through cardiometabolic disease. The effect estimates are likely to be sensitive to even low levels of residual confounding, and to misclassification of mania/BD exposure status. Widespread missingness in the covariates limited the interpretation of the multivariable analyses.

The finding of a specific effect of mania/BD on short-term visuospatial memory was unexpected. This result was robust to covariate adjustment and missingness, and was estimated consistently regardless of the regression or matching approach used. Previous meta-analyses have also reported decrements on visual memory tests in adults with BD (Arts et al., 2008; Bora et al., 2009; Bora et al., 2011; Mann-Wrobel et al., 2011), although with exceptions: one review of first episode BD patients reported no group differences in this domain (Lee et al., 2014). The magnitude of reported effects has been in the medium to large range in previous studies (standardised mean differences -0.5 to -0.8), in contrast with the small effect evident here (approximately -0.2). This may reflect the larger sample size or milder clinical status of the mania/BD group in UK Biobank, compared with clinic-

based samples in previous studies. Deficits in other cognitive domains have also been evident in the literature; for example, group differences in speed/reaction time have been consistently reported, but no such difference was found here. There were indications, however, that the estimates in the present analyses were biased toward the null as a result of missing covariate data. The UK Biobank baseline cognitive assessment did not cover other key domains that have been highlighted in previous research, such as verbal memory, and so a comprehensive cross-domain comparison cannot be made.

Despite evidence of elevated rates of cardiometabolic disease in adults with BD (e.g. Fiedorowicz, He, & Merikangas, 2011)—including in UK Biobank (Martin et al., 2016; see also Table 7.2)—and negative associations between cardiometabolic disease and cognitive function (e.g. Qiu & Fratiglioni, 2015), a mediating effect of cardiometabolic disease on cognitive performance was not found in the present analyses. This may be because of statistical adjustment for closely related sociodemographic and lifestyle covariates (e.g. BMI, smoking), or it may be related to misclassification in the self-reported cardiometabolic disease data, or the relatively young age of the sample compared with previous studies of cardiometabolic disease and cognitive function.

There was clearer evidence of mediation through psychotropic medication. Previous studies of cognitive function in BD have highlighted antipsychotic medication, in particular, as a risk factor for cognitive impairment (Altshuler et al., 2004; Balanzá-Martínez et al., 2010; Frangou et al., 2005; Jamrozinski et al., 2009; Kieseppa et al., 2005). It is important to consider the possibility of confounding by indication: adults with mania/BD who are on particular psychotropic medications may have had a phenotypically different or more severe disease course than those who are not on these medications, and this may not have been adequately controlled for in the present analyses. It should also be noted that the residual dependencies in the DAG (pair-wise correlations >|0.10|; see section 7.2.2 above and Appendix X) involved the psychotropic medication node, which may indicate measurement error or model misspecification, leading to uncontrolled influences of measured or unmeasured background confounders in the analyses involving this variable. It is therefore important that the possible causal role of psychotropic medication is evaluated carefully in future research; where possible, future studies should also analyse different psychotropic classes separately, and should consider opposing directions of effect whereby the potential positive cognitive effects of medication-related affective remission are contrasted against the potential adverse effects of the medications themselves.

The strengths of the present research include the large sample size, with the number of mania/BD participants in this single sample (n = 2,709) approaching that of a recent individual participant data meta-analysis (n = 2,876) of 31 studies (Bourne et al., 2013). A wide range of potential confounders were considered, each measured in a standard way. Multiple planned analyses were conducted, comparing different estimation approaches within a formal causal inference framework. The model specification was carefully evaluated using statistical and graphical methods, to compare the fit of alternative models and to consider the implications of structurally equivalent graphs for the plausibility of the analysed model.

As outlined in previous chapters, a number of limitations must be considered when interpreting the results of cognitive research in UK Biobank. The cohort is not representative of the general population in some respects (Fry et al., 2017), and this may lead to collider bias if both exposure and outcome status influenced participation and/or missing data rates. In the present analyses, it is plausible that milder mental health problems and better cognitive function would increase the likelihood of joining UK Biobank and contributing complete data, which would mean that any true negative relationship between mental illness and cognitive function would be attenuated or reversed. The sensitivity analyses showed that when the impact of missingness was reduced through multiple imputation, the degree of attenuation in the group differences was lessened in the multivariable models, with decrements (albeit very small) now evident in the mania/BD group on multiple cognitive outcomes. These analyses relied on an assumption of missingness at random, however, which is likely to be implausible for at least some of the variables, e.g. past mental state.

Measurement error and misclassification also present challenges for UK Biobank research. The sensitivity analyses reported here indicated that plausible degrees of exposure misclassification would attenuate the visuospatial memory result toward the null. A stricter definition of the mania/BD exposure (e.g. the narrow definition described in Chapter 4) may produce higher estimates of impairment. Conversely, retaining participants with major depression in the unexposed group would likely reduce the average group differences in cognitive performance. The suboptimal reliability of the cognitive outcome measures (Lyall et al., 2016) will have reduced the statistical power, although this is unlikely to have affected the visuospatial memory and reaction time analyses, as these measures were available for the whole cohort. It is possible to address measurement error by using latent variables; for example, a single latent cognitive outcome variable could have been entered

into the analyses, with the measured test scores as indicators. This was not done, however, because it was of interest to investigate whether domain- or test-specific deficits were evident in the mania/BD group, and any such differentiation would have been masked in a latent global variable. It is also possible to use test-retest reliability data to adjust single-indicator measures; repeat data are available in UK Biobank (n ~ 20,000) for most of the exposure, outcome and covariate measures, but these were collected up to 7 years post-baseline, and so may not provide an accurate indication of test-retest reliability in the absence of true change. The validity of some of the covariate measures is open to question, e.g. the information regarding the number of past depressive episodes could not be clearly separated into pre- and post-baseline periods for participants with web-based questionnaire data only, and the GPS that was intended to measure genetic aspects of cognitive function was based on a GWAS of education rather than cognitive ability per se (although these phenotypes have a very high genetic correlation). A further limitation was the absence of data regarding the number of past episodes of mania; combining such information with the data on past depressive episodes would have strengthened the validity of the past mental health measure.

Other limitations of the UK Biobank data used in these analyses include their cross-sectional nature, particularly the absence of information about premorbid cognitive function that would support the assumptions made here regarding the temporal order of the mania/BD exposure and the cognitive outcomes. There were also constraints on which intermediate variables could be examined in mediation analyses; for example, current depressive symptoms could not be analysed in this way, owing to an open back-door path involving the unmeasured ‘brain_health’ node, but this can be addressed in future analyses as the number of participants with neuroimaging data increases. This and other issues—e.g. the transportability of UK Biobank results to external populations—will be considered further in Chapter 9. Chapter 8 continues the present research by applying the same causal inference framework to cognitive performance in major depression and schizophrenia groups in UK Biobank.

Chapter 8 Explaining variation in cognitive function in other psychiatric conditions

The studies presented in this chapter aimed to estimate the total causal effects of major depression and schizophrenia on cognitive function, and to decompose these into direct and indirect effects through potentially modifiable intermediate factors. The analyses were conducted using baseline data from the UK Biobank cohort, and followed the same procedures as were described in Chapter 7. The major depression analyses and their results are presented first, followed by the schizophrenia analyses, and both are discussed together at the end of the chapter. Owing to the absence of data regarding past illness course for UK Biobank participants with neurological conditions, no causal analyses for cognitive outcome in MS or PD were undertaken in this thesis.

8.1 Major depression

8.1.1 Methods

8.1.1.1 Participants

Participants were retained for analysis if they had sufficient data to classify their exposure status and had data on at least one cognitive outcome measure. Adjusted analyses were restricted to participants of white British genetic ancestry; the unadjusted analyses were not restricted by ethnic group. The sample size available for analysis is detailed in section 8.1.2.1.

8.1.1.2 Exposure status

The illness exposure of interest was major depression, classified using the broad definition given in Chapter 4 (i.e. met criteria for single or multiple episodes of major depression according to self-reported diagnosis, or touchscreen mood questionnaire algorithm, or pre-baseline hospital records; participants were classified as exposed if they met the criteria in at least one information source, even if other information sources were missing). The unexposed comparison group comprised participants who had complete self-reported and touchscreen mood questionnaire data indicating that they did not meet the criteria for major depression or mania/BD, and whose pre-baseline hospital records had no primary or secondary diagnoses of major depression or mania/BD. Participants with a self-reported diagnosis or hospital record of schizophrenia (as defined in Chapter 4) were excluded from

the exposed and unexposed groups. Participants who did not meet the above criteria for either the exposed or unexposed groups were not further analysed.

8.1.1.3 Outcome measures and covariates

The cognitive measures were as described previously; reasoning, reaction time, numeric memory and visuospatial memory were analysed as z-scores (higher scores represent better performance), and prospective memory was analysed as a dichotomous variable (correct response at the first attempt or not). Other sociodemographic, local environment, lifestyle, physical, medical and family history, mental health and psychotropic medication measures were as described in Chapter 7, and the same education/cognition GPS was used. A GPS was created for major depression, using summary statistics from the most recent Psychiatric Genomics Consortium GWAS (Wray et al., 2017); the GWAS authors provided reanalysed statistics that excluded UK Biobank participants (participants from the 23andMe data resource were also omitted due to data-sharing restrictions). The major depression GPS was generated for unrelated UK Biobank participants of white British genetic ancestry, using the same quality control criteria as described previously, and the model residuals after adjustment for genotyping array and batch, UK Biobank assessment centre, and the first 20 genetic principal components were used in the analyses. Appendix W shows the R^2 at each *p* threshold, and the odds ratio for the association between deciles of the optimum GPS and the major depression phenotype in the UK Biobank cohort. The R^2 for the optimum GPS was 0.005. This is similar to that reported by the GWAS authors when using their core "anchor" cohort results alone to predict into independent samples, although R^2 values of up to 0.02 were obtained when additional cohort results (including UK Biobank and 23andMe) were included in GWAS meta-analyses (Wray et al., 2017).

8.1.1.4 Data analysis

The DAG used in the major depression analyses was based on the final DAG used in the mania/BD study, with the exception that an arrow was added from gender to major depression; studies have consistently shown higher prevalence of depression in women (e.g. Parker & Brotchie, 2010; Piccinelli & Wilkinson, 2000), and it was assumed in the DAG that this relationship was at least partly causal. No arrows were removed compared with the mania/BD DAG, on the assumption that similar causal relationships might be operating to explain cognitive impairment in both disorders. Two versions of the DAG were evaluated: the first depicted a causal influence of educational attainment on major depression and other psychiatric/neurological conditions, and the second depicted these

relationships in reverse. The testable implied independencies were analysed as per Chapter 7.

The total causal effect of major depression on cognitive performance was identified using the back-door criterion in DAGitty, and a series of regression and matching-based models were then conducted for each cognitive outcome, as described in Chapter 7. The propensity score model was again specified in three ways, and the score that resulted in the best covariate balance was taken forward into the analyses. G-computation was used to estimate the natural direct and indirect effects via intermediate nodes, adjusting for the minimum sufficient covariate sets to block all back-door paths.

8.1.1.5 Sensitivity analyses

These followed the same procedure as described in Chapter 7. The sensitivity of the total effects results to different potential magnitudes of departure from exchangeability were evaluated using Rosenbaum bounds. To investigate the impact of missing covariate data on the results, the multiple linear regression models were repeated using multiply-imputed datasets, and the chained equations imputation option was also implemented in the G-computation models. The effect of different hypothetical levels of major depression exposure misclassification on cognitive outcome (dichotomous impaired status) was assessed using a trapezoidal probability function, assuming differential misclassification. The final DAG was analysed structurally to determine its equivalence class.

8.1.2 Results

8.1.2.1 Characteristics of the sample

Figure 8.1 shows a flowchart of exclusions leading to the final analysis sample, which comprised 50,975 participants with major depression and 102,931 comparison participants.

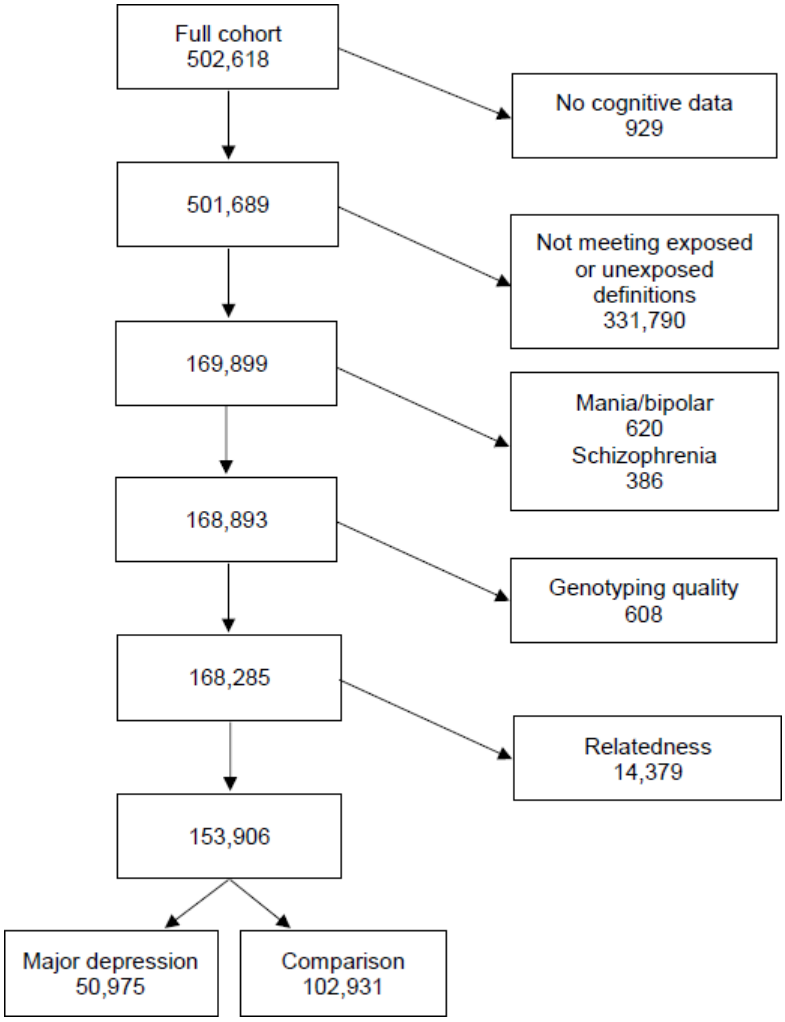


Figure 8.1 - Major depression analysis sample flowchart

Table 8.1 indicates slightly lower scores, on average, in the major depression group (all available data) for reaction time, numeric memory and visuospatial memory, but not for reasoning or prospective memory. Scores appeared higher overall among the participants with complete covariate data. Missingness was most common on the visuospatial memory test (3.2% and 2.8% in the major depression and comparison groups, respectively). Reliability (Cronbach's alpha) was higher for reaction time than for reasoning, and was largely similar between the major depression and comparison groups.

Table 8.1 - Summary of cognitive outcome measures in the major depression and comparison groups

	All available data		Complete covariate data[a]	
	Major depression	**Comparison**	**Major depression**	**Comparison**
n	50,975	102,931	11,662	26,392
Reasoning z-score				
n (%) missing[b]	421 (1.3)	1,584 (1.6)	25 (0.3)	84 (0.3)
Mean (SD)	-0.20 (0.96)	-0.20 (0.97)	0.10 (0.90)	0.13 (0.92)
Cronbach's α	0.69	0.70	0.66	0.68
Reaction time z-score				
n (%) missing	534 (1.1)	1,139 (1.1)	34 (0.3)	68 (0.3)
Mean (SD)	-0.08 (0.98)	-0.03 (0.98)	0.04 (0.94)	0.07 (0.95)
Cronbach's α	0.82	0.82	0.82	0.82
Numeric memory z-score				
n (%) missing[b]	252 (2.6)	733 (2.4)	18 (0.8)	49 (0.7)
Mean (SD)	-0.39 (0.93)	-0.35 (0.94)	-0.24 (0.90)	-0.17 (0.93)
Visuospatial memory z-score				
n (%) missing	1,653 (3.2)	2,843 (2.8)	106 (0.9)	206 (0.8)
Mean (SD)	0.19 (1.05)	0.26 (1.05)	0.30 (1.04)	0.37 (1.03)
Prospective memory[b,c]				
n (%) correct	25,006 (78.0)	78,673 (76.8)	7,003 (85.6)	22,591 (85.7)

GPS, genome-wide polygenic score; SD, standard deviation.
a. Participants with complete data on all the covariates that were entered into the maximally-adjusted total effects models (age, gender, white British genetic ancestry, English-speaking country of birth, degree, comorbid neurological/psychiatric condition, family history of dementia, family history of Parkinson's disease, family history of severe depression, maternal smoking around birth, childhood trauma, education/cognition GPS, major depression GPS).
b. Missing data refers only to the period when this measure was included in the battery.
c. No missing data.

Table 8.2 describes the covariate data in both groups. As was found in Chapter 7, the proportion of missing data was highest, by far, on the childhood trauma variable, and was also relatively high on the family history, current depressive symptoms and physical activity measures. The descriptive information indicated that the major depression group was younger, on average, than the comparison group, had a substantially higher proportion of women, and had higher proportions of current smokers, former drinkers, and participants living in deprived areas. The major depression group had higher proportions with frequent sleeplessness, obesity, cardiometabolic disease, comorbid neurological/psychiatric conditions, family history of severe depression, current psychotropic medication and history of childhood trauma, and they reported more depressed episodes and higher current depressive symptoms on average. The distribution of the major depression GPS score was skewed towards higher values in the major depression group. The subset of participants with complete covariate data appeared different from the full analysis sample across multiple measures, e.g. having a greater proportion of degree-holders and a smaller proportion from the most deprived areas.

Table 8.2 - Summary of covariates in the major depression and comparison groups

	All available data		Complete covariate data[a]	
	Major depression	**Comparison**	**Major depression**	**Comparison**
n	50,975	102,931	11,662	26,392
Sociodemographic				
Age (years)[b]				
Mean (SD)	55.6 (7.9)	57.0 (8.2)	54.9 (7.6)	56.3 (7.9)
Gender[b]				
n (%) female	33,090 (64.9)	51,463 (50.0)	7,996 (68.6)	14,072 (53.3)
Ethnic group				
n (%) missing	220 (0.4)	401 (0.4)	32 (0.3)	71 (0.3)
White, *n* (%)[c]	48,345 (95.3)	93,194 (90.9)	11,347 (97.6)	25,123 (95.5)
Asian/Asian British	814 (1.6)	3,663 (3.6)	77 (0.7)	451 (1.7)
Black/Black British	641 (1.3)	3,159 (3.1)	67 (0.6)	319 (1.2)
Chinese	69 (0.1)	473 (0.5)	11 (0.1)	97 (0.4)
Mixed & other background	886 (1.8)	2,041 (2.0)	128 (1.1)	331 (1.3)
White British genetic ancestry				
n (%) missing	1,518 (3.0)	3,419 (3.3)	0 (0.0)	0 (0.0)
n (%)[c]	41,509 (83.9)	79,097 (79.5)	10,029 (86.0)	22,075 (83.6)
English-speaking country of birth				
n (%) missing	85 (0.2)	149 (0.2)	0 (0.0)	0 (0.0)
n (%)[c]	47,741 (93.8)	91,524 (89.1)	11,068 (94.9)	24,407 (92.5)
Has a degree				
n (%) missing	493 (1.0)	1,012 (1.0)	0 (0.0)	0 (0.0)
n (%)[c]	16,713 (33.1)	36,211 (35.5)	5,570 (47.8)	12,977 (49.2)
Townsend quintile[d]				
n (%) missing	93 (0.2)	159 (0.2)	25 (0.2)	33 (0.1)
Qu1 (least deprived), *n* (%)[c]	8,109 (15.9)	17,672 (17.2)	2,207 (19.0)	5,019 (19.0)
Qu2	9,124 (17.9)	20,807 (20.3)	2,291 (19.7)	5,704 (21.6)
Qu3	9,793 (19.3)	21,291 (20.7)	2,359 (20.3)	5,767 (21.9)
Qu4	11,261 (22.1)	23,193 (22.6)	2,620 (22.5)	5,989 (22.7)
Qu5 (most deprived)	12,595 (24.8)	19,809 (19.3)	2,160 (18.6)	3,880 (14.7)
Local environment				
Home area population density[e]				
n (%) missing	596 (1.2)	951 (0.9)	160 (1.4)	265 (1.0)
England/Wales urban, *n* (%)[c]	42,462 (84.3)	88,651 (86.9)	9,617 (83.6)	22,259 (85.2)
England/Wales town	3,239 (6.4)	6,392 (6.3)	793 (6.9)	1,727 (6.6)
England/Wales village	2,031 (4.0)	4,853 (4.8)	555 (4.8)	1,493 (5.7)
England/Wales hamlet/isolated	842 (1.7)	2,084 (2.0)	223 (1.9)	648 (2.5)
Scotland large urban	1,376 (2.7)	0 (0.0)	253 (2.2)	0 (0.0)
Scotland other urban	300 (0.6)	0 (0.0)	41 (0.4)	0 (0.0)
Scotland small town	66 (0.1)	0 (0.0)	9 (0.1)	0 (0.0)
Scotland rural	63 (0.1)	0 (0.0)	11 (0.1)	0 (0.0)
Proximity to major road (1/m)				
n (%) missing	772 (1.5)	1,290 (1.3)	166 (1.4)	344 (1.3)
Mean (SD)	0.006 (0.014)	0.006 (0.013)	0.005 (0.012)	0.006 (0.010)
Particulate matter ≤10μm (μg/m^3)				
n (%) missing	938 (1.8)	1,668 (1.6)	205 (1.8)	430 (1.6)
Mean (SD)	22.4 (2.9)	22.8 (3.1)	22.5 (3.1)	22.8 (3.2)
Nitrogen dioxide (μg/m^3)				
n (%) missing	772 (1.5)	1,290 (1.3)	166 (1.4)	344 (1.3)
Mean (SD)	30.9 (10.1)	31.9 (10.6)	30.9 (10.6)	31.9 (1.1)
Lifestyle and physical				
Smoking status				
n (%) missing	170 (0.3)	368 (0.4)	8 (0.1)	47 (0.2)
Never, *n* (%)[c]	25,008 (49.2)	58,955 (57.5)	6,222 (53.4)	16,066 (61.0)
Former	18,320 (36.1)	34,658 (33.8)	4,361 (37.4)	8,682 (33.0)
Current	7,477 (14.7)	8,950 (8.7)	1,071 (9.2)	1,597 (6.1)

	All available data		Complete covariate data[a]	
	Major depression	**Comparison**	**Major depression**	**Comparison**
Alcohol frequency				
n (%) missing	100 (0.2)	80 (0.1)	5 (0.1)	5 (0.1)
Daily/almost daily, *n* (%)[c]	9,839 (19.3)	21,729 (21.1)	2,562 (22.0)	6,399 (24.3)
3-4 times per week	10,085 (19.8)	24,196 (23.5)	2,773 (23.8)	7,045 (26.7)
1-2 times per week	11,765 (23.1)	26,111 (25.4)	2,660 (22.8)	6,450 (24.4)
1-3 times per month	6,344 (12.5)	11,118 (10.8)	1,410 (12.1)	2,812 (10.7)
Special occasions only	7,333 (14.4)	11,536 (11.2)	1,378 (11.8)	2,299 (8.7)
Never (former drinker)	3,189 (6.3)	3,096 (3.0)	536 (4.6)	599 (2.3)
Never (not former drinker)	2,320 (4.6)	5,065 (4.9)	338 (2.9)	783 (3.0)
Sleeplessness				
n (%) missing	58 (0.1)	79 (0.1)	5 (0.1)	14 (0.1)
Never/rarely, *n* (%)[c]	8,438 (16.6)	28,617 (27.8)	2,236 (19.2)	7,855 (29.8)
Sometimes	23,090 (45.4)	49,193 (47.8)	5,381 (46.2)	12,485 (47.3)
Usually	19,389 (38.1)	25,042 (24.4)	4,040 (34.7)	6,038 (22.9)
Physical activity (MET h/week)				
n (%) missing	4,617 (9.1)	6,690 (6.5)	583 (5.0)	1,031 (3.9)
Median (Q1, Q3)	26.2 (11.6, 55.9)	29.9 (13.8, 60.2)	25.6 (12.1, 51.5)	29.2 (14.2, 55.8)
Body mass index				
n (%) missing	354 (0.7)	732 (0.7)	26 (0.2)	89 (0.3)
Underweight, *n* (%)[c]	270 (0.5)	496 (0.5)	62 (0.5)	145 (0.6)
Normal	15,166 (30.0)	34,192 (33.5)	4,152 (35.7)	10,534 (40.1)
Overweight	20,289 (40.1)	43,942 (43.0)	4,660 (40.1)	10,842 (41.2)
Obese class I	9,781 (19.3)	17,374 (17.0)	1,897 (16.3)	3,660 (13.9)
Obese class II	3,443 (6.8)	4,588 (4.5)	601 (5.2)	836 (3.2)
Obese class III	1,672 (3.3)	1,607 (1.6)	264 (2.3)	286 (1.1)
Medical and family history				
Cardiometabolic disease				
n (%) missing	121 (0.2)	177 (0.2)	7 (0.1)	21 (0.1)
n (%)[c]	17,459 (34.3)	31,361 (30.5)	2,961 (25.4)	6,185 (23.5)
Comorbid neurological or psychiatric condition[f]				
n (%)	11,542 (22.6)	9,155 (8.9)	2,242 (19.2)	2,035 (7.7)
Family history of dementia				
n (%) missing	7,815 (15.3)	14,476 (14.1)	0 (0.0)	0 (0.0)
n (%)[c]	7,032 (16.3)	15,330 (17.3)	1,880 (16.1)	4,546 (17.2)
Family history of Parkinson's disease				
n (%) missing	9,456 (18.6)	16,090 (15.6)	0 (0.0)	0 (0.0)
n (%)[c]	2,064 (5.0)	4,161 (4.8)	510 (4.4)	1,253 (4.8)
Family history of severe depression				
n (%) missing	7,871 (15.4)	15,347 (14.9)	0 (0.0)	0 (0.0)
n (%)[c]	12,370 (28.7)	10,201 (11.7)	3,014 (25.8)	3,000 (11.4)
Maternal smoking around birth				
n (%) missing	6,933 (13.6)	13,107 (12.7)	0 (0.0)	0 (0.0)
n (%)[c]	14,521 (33.0)	24,125 (26.9)	3,595 (30.8)	6,894 (26.1)
Mental health				
Current depressive symptoms				
n (%) missing	5,028 (9.9)	8,556 (8.3)	672 (5.8)	270 (1.0)
Mean (SD)	3.1 (3.0)	1.2 (1.7)	2.5 (2.6)	1.0 (1.4)
Any psychotropic medication				
n (%) missing	860 (1.7)	1,194 (1.2)	182 (1.6)	1,189 (4.5)
n (%)[c]	20,898 (41.7)	2,577 (2.5)	3,902 (34.0)	490 (1.9)
Number of depressed episodes				
n (%) missing	1,495 (2.9)	6,451 (6.3)	354 (3.0)	1,418 (5.4)
Median (Q1, Q3)	1 (0, 4)	0 (0, 1)	3 (1, 6)	0 (0, 1)
Any childhood trauma[g]				
n (%) missing	34,801 (68.3)	67,160 (65.3)	0 (0.0)	0 (0.0)
n (%)[c]	9,455 (58.5)	15,651 (43.8)	6,515 (55.9)	10,968 (41.6)

	All available data		Complete covariate data[a]	
	Major depression	**Comparison**	**Major depression**	**Comparison**
Genome-wide polygenic scores				
Education/cognition GPS decile[d]				
n (%) missing	1,518 (3.0)	3,419 (3.3)	0 (0.0)	0 (0.0)
D1 (lowest), *n* (%)[c]	5,169 (10.5)	9,728 (9.8)	982 (8.4)	2,163 (8.2)
D2	5,089 (10.3)	9,808 (9.9)	1,089 (9.3)	2,350 (8.9)
D3	5,002 (10.1)	9,895 (9.9)	1,096 (9.4)	2,423 (9.2)
D4	4,931 (10.0)	9,966 (10.0)	1,130 (9.7)	2,572 (9.8)
D5	4,960 (10.0)	9,937 (10.0)	1,142 (9.8)	2,515 (9.5)
D6	4,890 (9.9)	10,007 (10.1)	1,157 (9.9)	2,641 (10.0)
D7	4,914 (9.9)	9,983 (10.0)	1,244 (10.7)	2,753 (10.4)
D8	4,830 (9.8)	10,067 (10.1)	1,185 (10.2)	2,877 (10.9)
D9	4,815 (9.7)	10,082 (10.1)	1,285 (11.0)	2,950 (11.2)
D10 (highest)	4,857 (9.8)	10,039 (10.1)	1,352 (11.6)	3,148 (11.9)
Bipolar disorder GPS decile[d]				
n (%) missing	1,518 (3.0)	3,419 (3.3)	0 (0.0)	0 (0.0)
D1 (lowest), *n* (%)[c]	4,271 (8.6)	10,626 (10.7)	1,038 (8.9)	2,998 (11.4)
D2	4,645 (9.4)	10,252 (10.3)	1,124 (9.6)	2,826 (10.7)
D3	4,690 (9.5)	10,207 (10.3)	1,136 (9.7)	2,791 (10.6)
D4	4,693 (9.5)	10,204 (10.3)	1,070 (9.2)	2,784 (10.6)
D5	4,890 (9.9)	10,007 (10.1)	1,144 (9.8)	2,699 (10.2)
D6	4,887 (9.9)	10,010 (10.1)	1,142 (9.8)	2,551 (9.7)
D7	5,057 (10.2)	9,840 (9.9)	1,198 (10.3)	2,517 (9.5)
D8	5,247 (10.6)	9,650 (9.7)	1,228 (10.5)	2,460 (9.3)
D9	5,313 (10.7)	9,584 (9.6)	1,249 (10.7)	2,460 (9.3)
D10 (highest)	5,764 (11.7)	9,162 (9.2)	1,333 (11.4)	2,306 (8.7)

D, decile; GPS, genome-wide polygenic score; MET, metabolic equivalent of task; Q, quartile; Qu, quintile; SD, standard deviation.
a. Participants with complete data on all the covariates that were entered into the maximally-adjusted total effects models (age, gender, white British genetic ancestry, English-speaking country of birth, degree, comorbid neurological/psychiatric condition, family history of dementia, family history of Parkinson's disease, family history of severe depression, maternal smoking around birth, childhood trauma, education/cognition GPS, major depression GPS).
b. No missing data.
c. Missing excluded from denominator.
d. Based on data distribution in the whole UK Biobank cohort.
e. Scottish psychiatric hospital records were unavailable, which meant no Scotland-based participants could be classified in the comparison group; therefore all locations for comparison participants are in England/Wales.
f. Apart from mood disorder or schizophrenia; not possible to distinguish between missing data and self-report of no condition, therefore both classified as No.
g. From the web-based questionnaire, which was completed by 157,366 (31.3%) of the cohort.

8.1.2.2 Evaluation of the graphical model

The different predicted independencies implied by the two specifications of the DAG (educational attainment as an antecedent, or a consequence, of major depression and other psychiatric/neurological conditions) were tested, and better fit was evident in the first specification (see Appendix BB). Fifteen percent (21 of 137) of the partial correlation coefficients were above |0.1|, but most of these were below |0.2| and the largest coefficient was |0.22|. This model, depicted in Figure 8.2, was used to plan the causal effect models reported below.

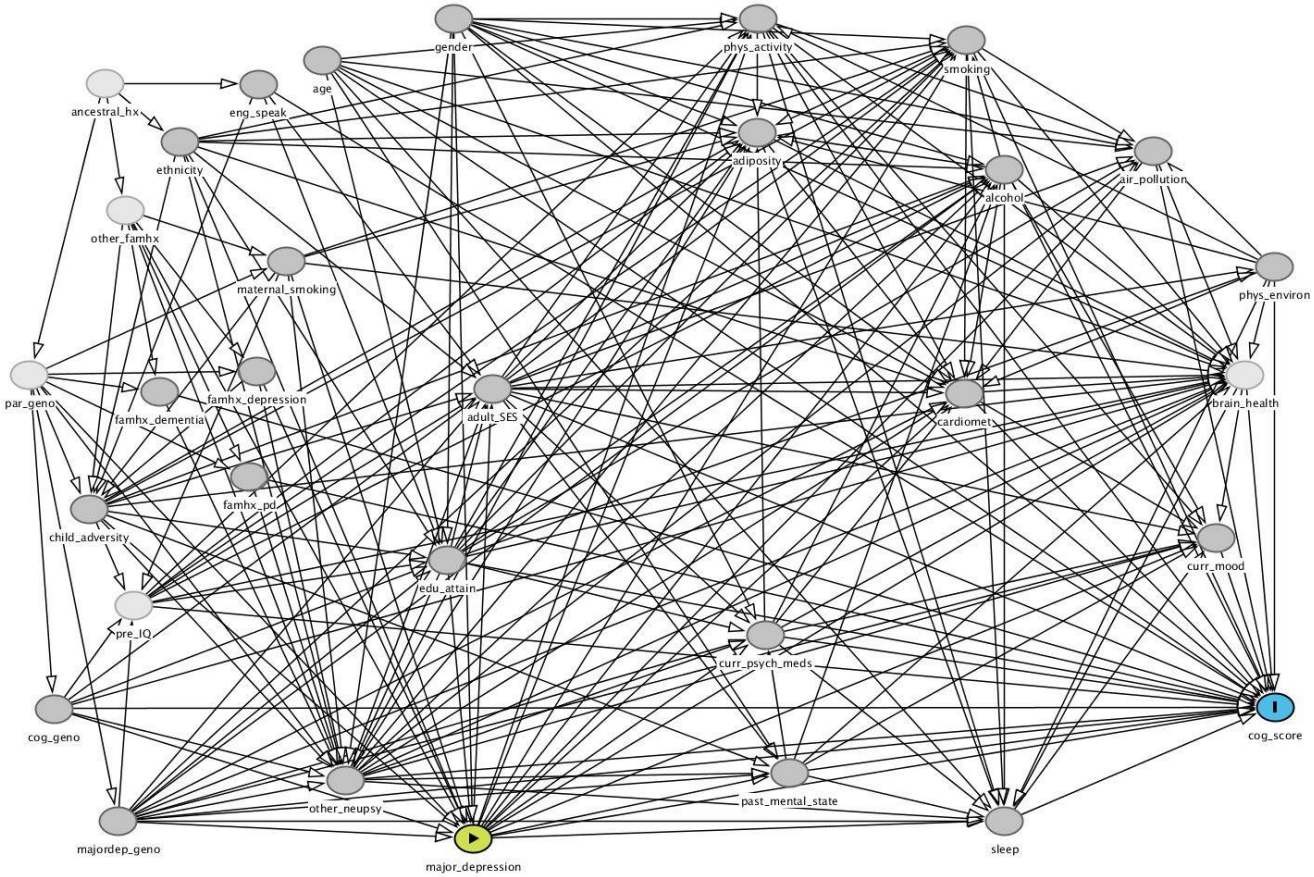


Figure 8.2 - Directed acyclic graph for the effect of major depression on cognitive outcome

Cardiomet, cardiometabolic disease; cog, cognitive; curr, current; edu, educational; eng_speak, English speaking birth country; famhx, family history; geno, genotype; hx, history; majordep, major depression; other_neupsy, other neurological/psychiatric condition; par, parental; PD, Parkinson's disease; phys, physical; pre_IQ, premorbid intelligence; psych_meds, psychotropic medications; SES, socioeconomic status. Green node is the exposure and blue node is the outcome. Light nodes represent unmeasured constructs and darker nodes represent measured constructs. This graph and the underlying code are publicly available online at dagitty.net/mqiL72i

8.1.2.3 Total causal effects

Applying d-separation criteria to the DAG indicated that the minimum sufficient adjustment set for the total effect of major depression on cognitive performance comprised gender, educational attainment, English-speaking birth country, ethnicity, education/cognition GPS, major depression GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, and other psychiatric/neurological conditions. The extended adjustment set encompassed additional measured nodes that were antecedents of the exposure, namely age and family history of depression. This extended adjustment set was also used as the predictor set for the propensity score model. Ethnicity was accounted for in all the multivariable analyses and in the propensity score estimation by restricting these to participants of white British genetic ancestry. The GPS score residuals were entered as deciles, based on the distribution in the full analysis sample.

The best covariate balance was obtained using the first propensity score model with no interaction terms, as illustrated in Table 8.3. This was used in all the outcome models that involved propensity score adjustment or matching, or inverse probability weighting.

Figures 8.3 to 8.7 show the results of all the total effects models. Only the visuospatial memory test indicated a detrimental effect of major depression that remained evident in the multivariable models. The effect size was very small, with the major depression group scoring approximately 0.07 SD lower than the comparison group. The visuospatial memory estimates showed little change between the unadjusted and adjusted/matched models, whereas the estimates for the other cognitive measures generally showed attenuation towards the null.

Table 8.3 - Summary of covariates in matched major depression and comparison groups

	Major depression	Comparison
	Mean (SD)	
Age (years)	55.1 (7.5)	55.2 (7.5)
	%	
Female gender	68.4	69.5
English-speaking country of birth	98.3	98.2
Has a degree	46.0	46.5
Comorbid neurological or psychiatric condition	19.3	19.3
Family history of dementia	16.4	16.4
Family history of Parkinson's disease	4.4	4.3
Family history of severe depression	25.2	24.7
Maternal smoking around birth	31.2	30.7
Any childhood trauma	54.8	54.7
Education/cognition GPS		
D1 (lowest)	8.6	8.2
D2	9.4	9.4
D3	9.5	9.5
D4	9.7	9.5
D5	9.7	9.8
D6	9.8	9.8
D7	10.6	10.9
D8	10.1	10.1
D9	11.1	11.1
D10 (highest)	11.5	11.7
Major depression GPS		
D1 (lowest)	8.6	8.8
D2	9.6	9.7
D3	9.9	9.8
D4	9.0	8.8
D5	9.8	10.0
D6	9.8	9.3
D7	10.2	10.6
D8	10.6	10.7
D9	10.8	10.9
D10 (highest)	11.7	11.4

D, decile; GPS, genome-wide polygenic score; SD, standard deviation.
These results are from the matched samples used in the 1:3 matched model for the total effect of major depression on reaction time (major depression n = 9,381; comparison n = 13,538).

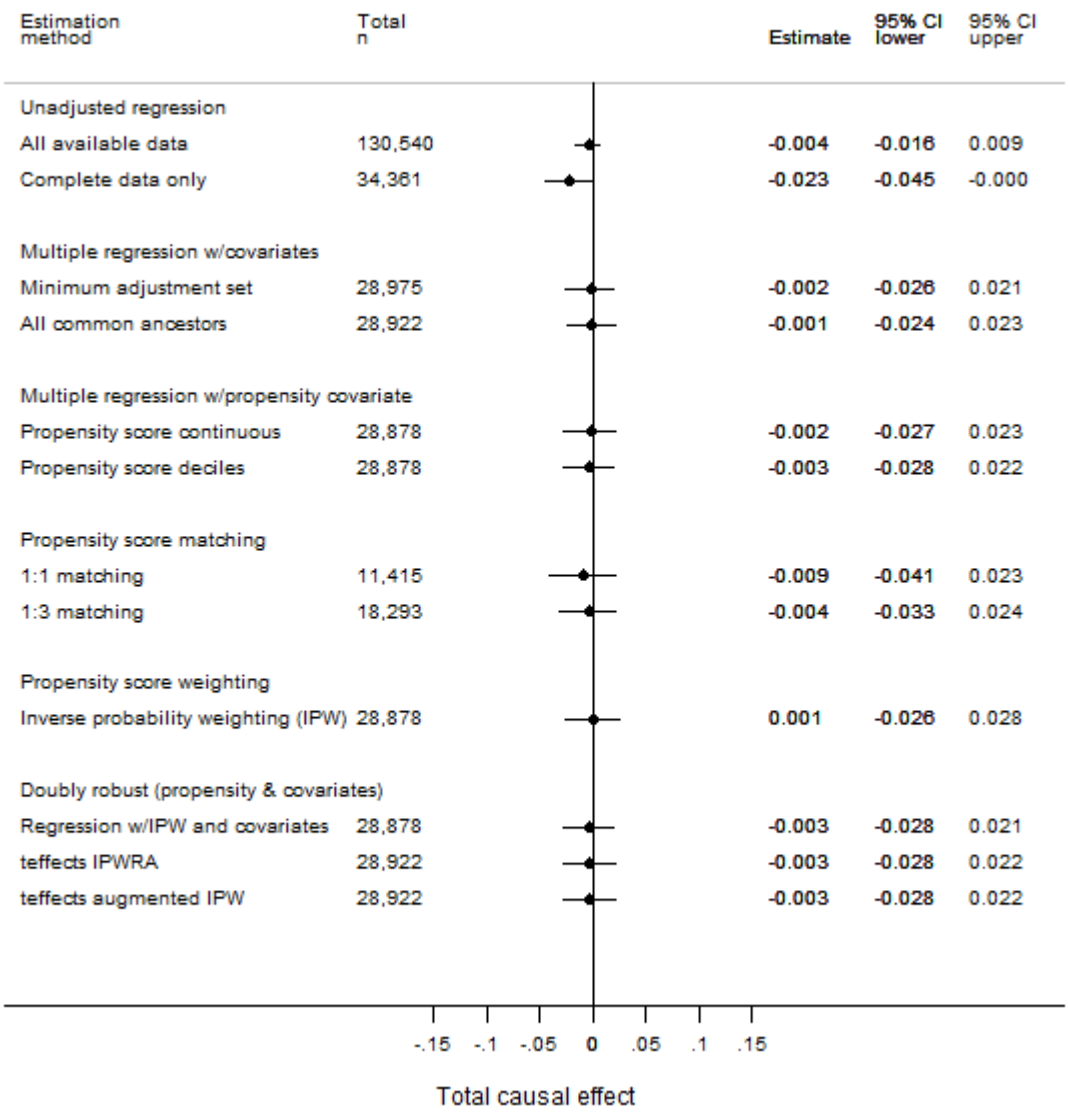


Figure 8.3 - Total causal effect of major depression on reasoning

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

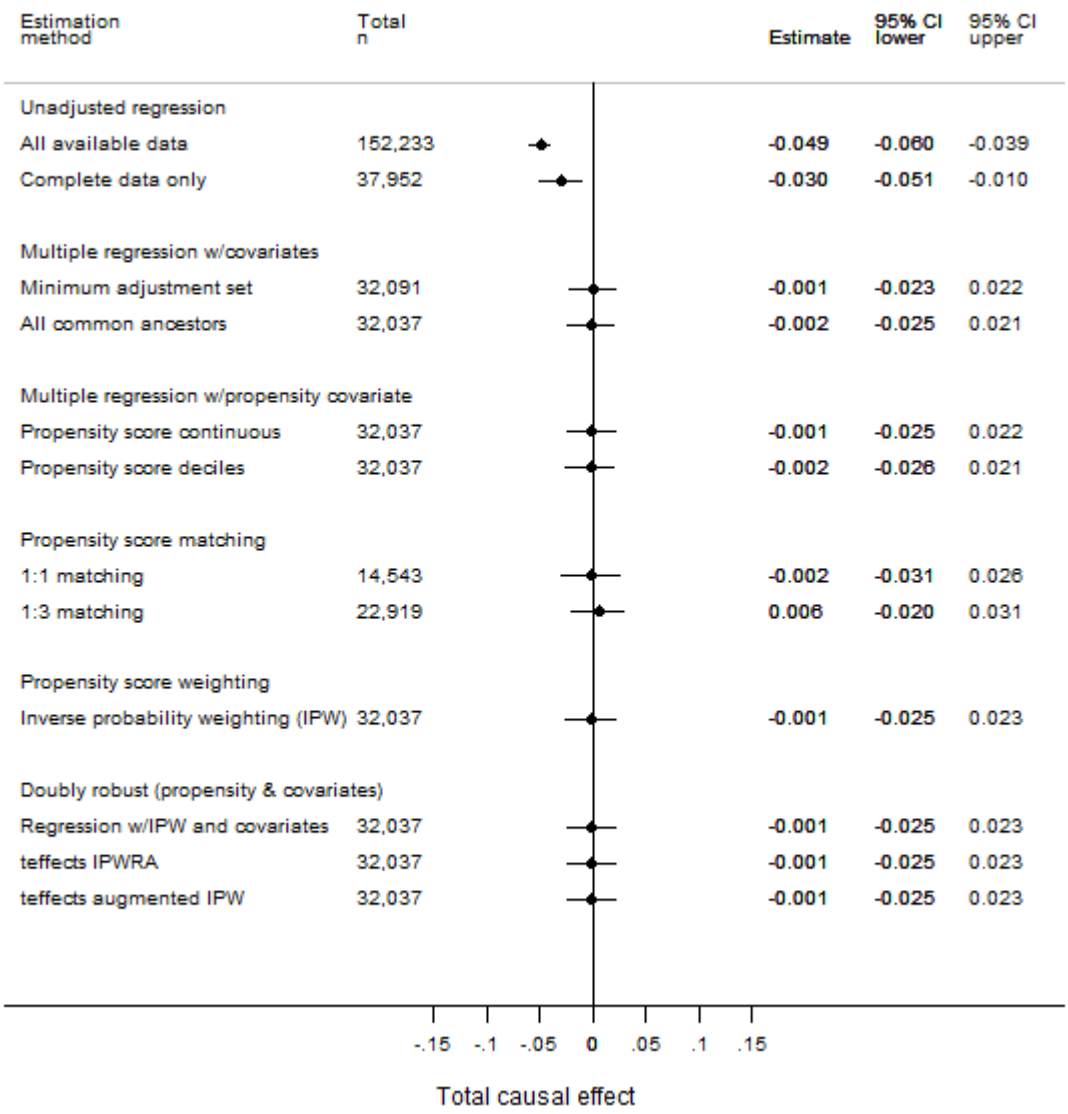


Figure 8.4 - Total causal effect of major depression on reaction time

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

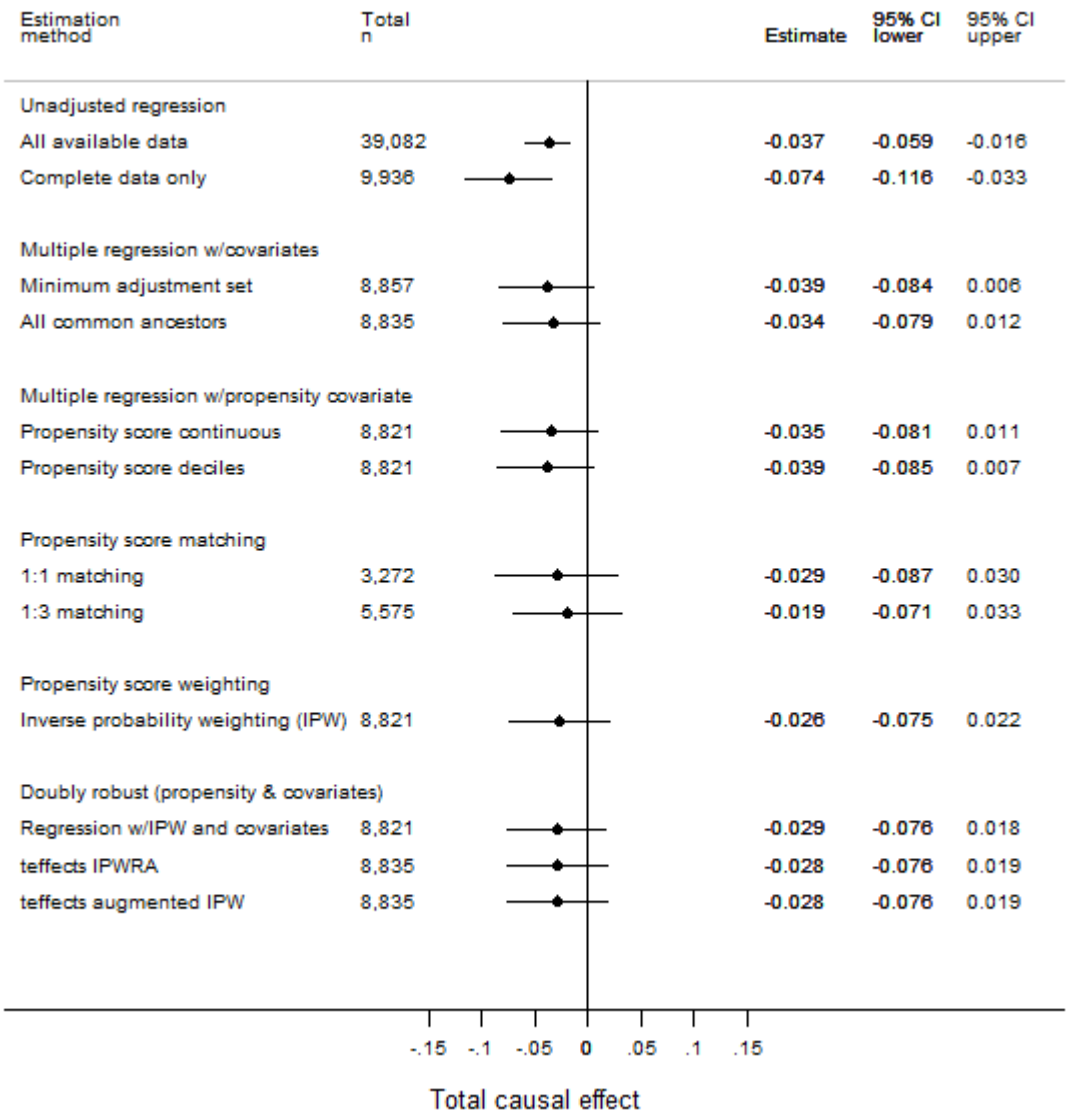


Figure 8.5 - Total causal effect of major depression on numeric memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

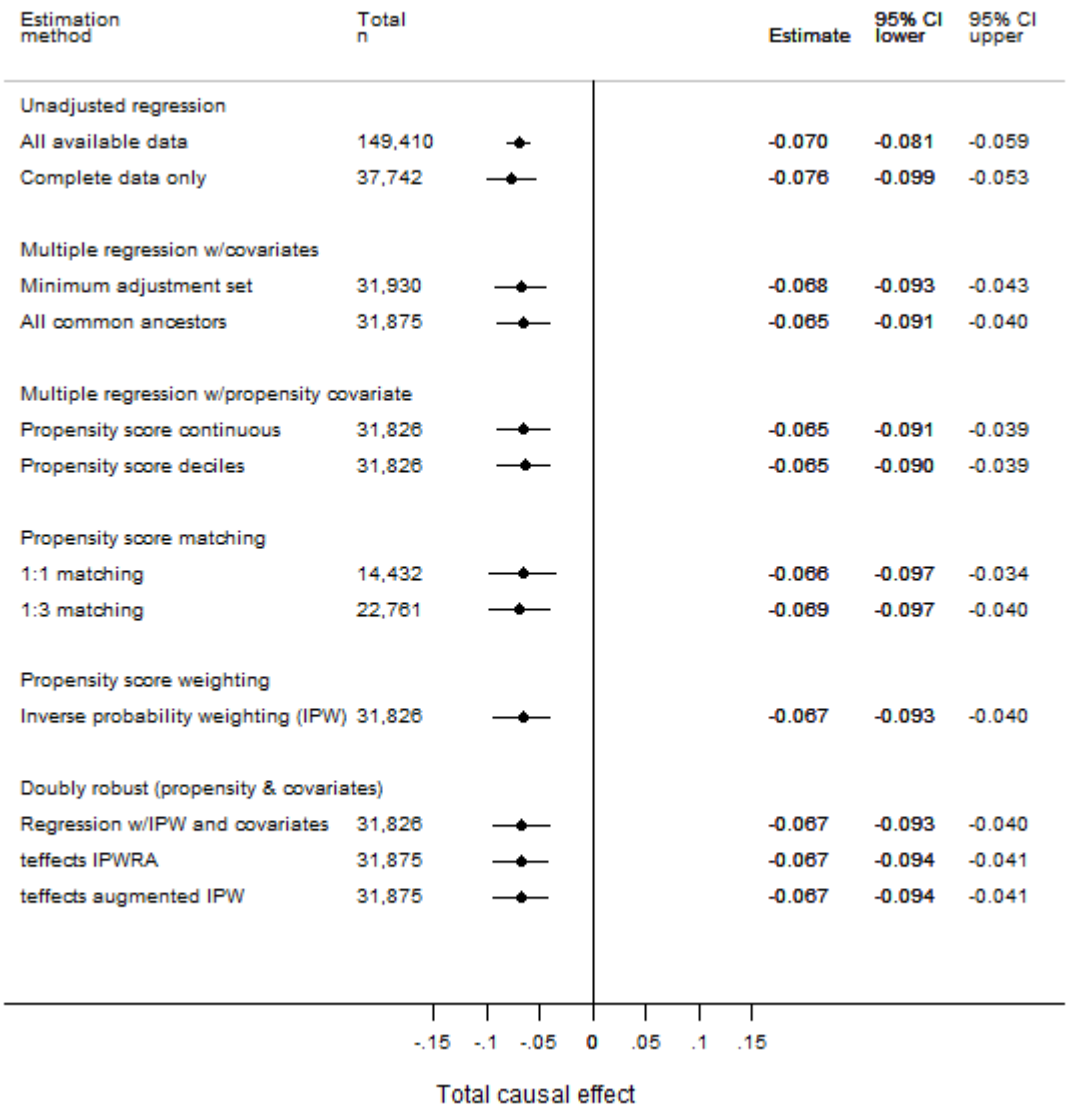


Figure 8.6 - Total causal effect of major depression on visuospatial memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are in z-score units and can be interpreted as standardised mean differences.

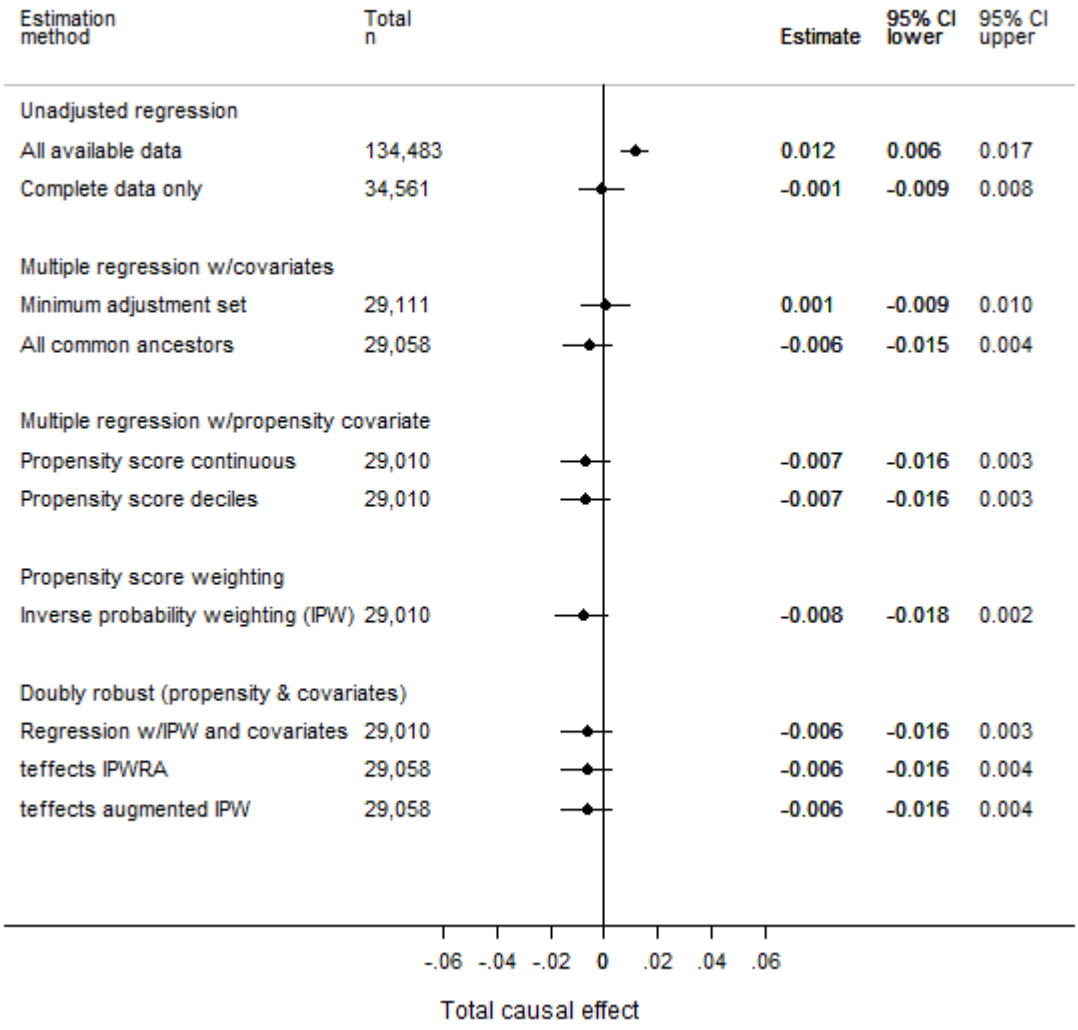


Figure 8.7 - Total causal effect of major depression on prospective memory

CI, confidence interval; IPW, inverse probability weighting; IPWRA, inverse probability weighting with regression adjustment; teffects, Stata `teffects` package. Estimates are proportions and can be interpreted as risk differences. No estimates are provided from propensity-score matched models as it was not possible to express these as risk differences.

8.1.2.4 Direct and indirect effects

As with the mania/BD study, structural analysis of the DAG indicated that direct and indirect effects could be decomposed for cardiometabolic disease and psychotropic medication. In both sets of analyses, the indirect pathways were affected by intermediate confounding, and so an identifying assumption was firstly made of no interaction between major depression and either cardiometabolic disease or psychotropic medication (Robins & Greenland, 1992). This assumption was checked by conducting a regression model of each cognitive outcome on major depression exposure status, the mediator and all the covariates, including a product term for major depression * mediator. There was no evidence of interaction in the cardiometabolic disease mediation models, but some evidence of interaction in the psychotropic medication models (see Table CC.1 in Appendix CC). The alternative identifying assumption proposed by Petersen et al. (2006) was therefore made for the psychotropic medication models; following De Stavola et al. (2015), this was checked by testing for interactions between major depression and each of the intermediate confounders (deprivation, and lifetime number of episodes of depressed mood/anhedonia), in regression models that included major depression, psychotropic medication and their product, along with all the other model covariates. There was little evidence of interaction between major depression and the intermediate confounders (Table CC.2 in Appendix CC), and so the Petersen et al. (2006) identifying assumption was deemed reasonable and these mediation models were estimated with a product term included between major depression and psychotropic medication.

Table 8.4 shows the direct and indirect effect estimates via cardiometabolic disease. The model covariates were age, gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), BMI, alcohol frequency, smoking status, physical activity, and psychotropic medication. The analyses were restricted to participants of white British genetic ancestry, and the residualised GPS scores were entered as deciles. There was no evidence of substantive indirect effects via cardiometabolic disease in any of the models.

Table 8.4 - Mediation of the effect of major depression on cognitive outcome via cardiometabolic disease

	n	Estimate	95% CI[a]
Reasoning[b]	26,679		
TCE		-0.009	-0.038, 0.019
CDE		-0.029	-0.056, -0.001
NDE		0.005	-0.023, 0.033
NIE		-0.014	-0.028, 0.000
Reaction time[b]	29,422		
TCE		-0.015	-0.042, 0.013
CDE		-0.006	-0.036, 0.024
NDE		-0.009	-0.037, 0.020
NIE		-0.006	-0.021, 0.009
Numeric memory[b]	8,085		
TCE		-0.034	-0.087, 0.019
CDE		-0.032	-0.085, 0.021
NDE		-0.035	-0.089, 0.018
NIE		0.001	-0.027, 0.029
Visuospatial memory[b]	29,284		
TCE		-0.074	-0.107, -0.042
CDE		-0.077	-0.109, -0.046
NDE		-0.079	-0.111, -0.047
NIE		0.005	-0.012, 0.022
Prospective memory[c]	26,789		
TCE		-0.010	-0.022, 0.001
CDE		-0.001	-0.012, 0.010
NDE		-0.011	-0.022, 0.000
NIE		0.001	-0.012, 0.010

CDE, controlled direct effect when cardiometabolic disease = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; NO_2, nitrogen dioxide; PM_{10}, particulate matter of up to 10μm diameter; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for age, gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), body mass index, alcohol frequency, smoking status, physical activity, and psychotropic medication.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

Table 8.5 shows the direct and indirect effect estimates via psychotropic medication. The model covariates were gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of PD, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia. The analyses were restricted to participants of white British genetic ancestry, and the residualised GPS scores were entered as deciles. There was little evidence of mediation via psychotropic medication: approximately one third of the total effect on visuospatial memory (itself of very small magnitude, at -0.058) was estimated to be indirect, but the confidence interval included the null (-0.019; 95% CI -0.040, 0.003).

Table 8.5 - Mediation of the effect of major depression on cognitive outcome via psychotropic medication

	n	Estimate	95% CI[a]
Reasoning[b]	27,263		
TCE		-0.009	-0.037, 0.019
CDE		-0.011	-0.040, 0.018
NDE		-0.010	-0.039, 0.020
NIE		0.001	-0.016, 0.017
Reaction time[b]	30,189		
TCE		0.002	-0.026, 0.029
CDE		0.005	-0.025, 0.036
NDE		0.026	-0.004, 0.055
NIE		-0.024	-0.042, -0.005
Numeric memory[b]	8,338		
TCE		-0.023	-0.078, 0.032
CDE		-0.021	-0.077, 0.035
NDE		-0.019	-0.076, 0.038
NIE		-0.004	-0.035, 0.028
Visuospatial memory[b]	30,038		
TCE		-0.058	-0.088, -0.028
CDE		-0.066	-0.100, -0.031
NDE		-0.039	-0.073, -0.006
NIE		-0.019	-0.040, 0.003
Prospective memory[c]	27,381		
TCE		0.001	-0.011, 0.012
CDE		0.001	-0.010, 0.013
NDE		0.001	-0.011, 0.012
NIE		0.000	-0.006, 0.006

CDE, controlled direct effect when psychotropic medication = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia. All models included a product term for major depression * psychotropic medication.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

8.1.2.5 Sensitivity analyses

Rosenbaum bounds were calculated to check the sensitivity of the visuospatial memory total effect result to departures from exchangeability. The estimated effect crossed the null at a gamma value of 1.07, i.e. the point where the probability of being in the exposed group is approximately 0.48 or 0.52. The results would therefore not be robust to an unmeasured confounder with even a very weak association with group membership.

There was evidence of missing data bias, in that the unadjusted total effects estimates shifted (towards or away from the null) when the sample was restricted to participants with complete covariate data. When the multiple linear regression models for total causal effects

were repeated using imputed covariate values, the estimates for reaction time indicated a very small detrimental effect (point estimate approximately -0.02 to -0.03) in the major depression group (Figure 8.8(b) below), which was not evident in the complete case analyses. The estimates for the other cognitive outcomes were similar between the complete case analyses and those using multiple imputation.

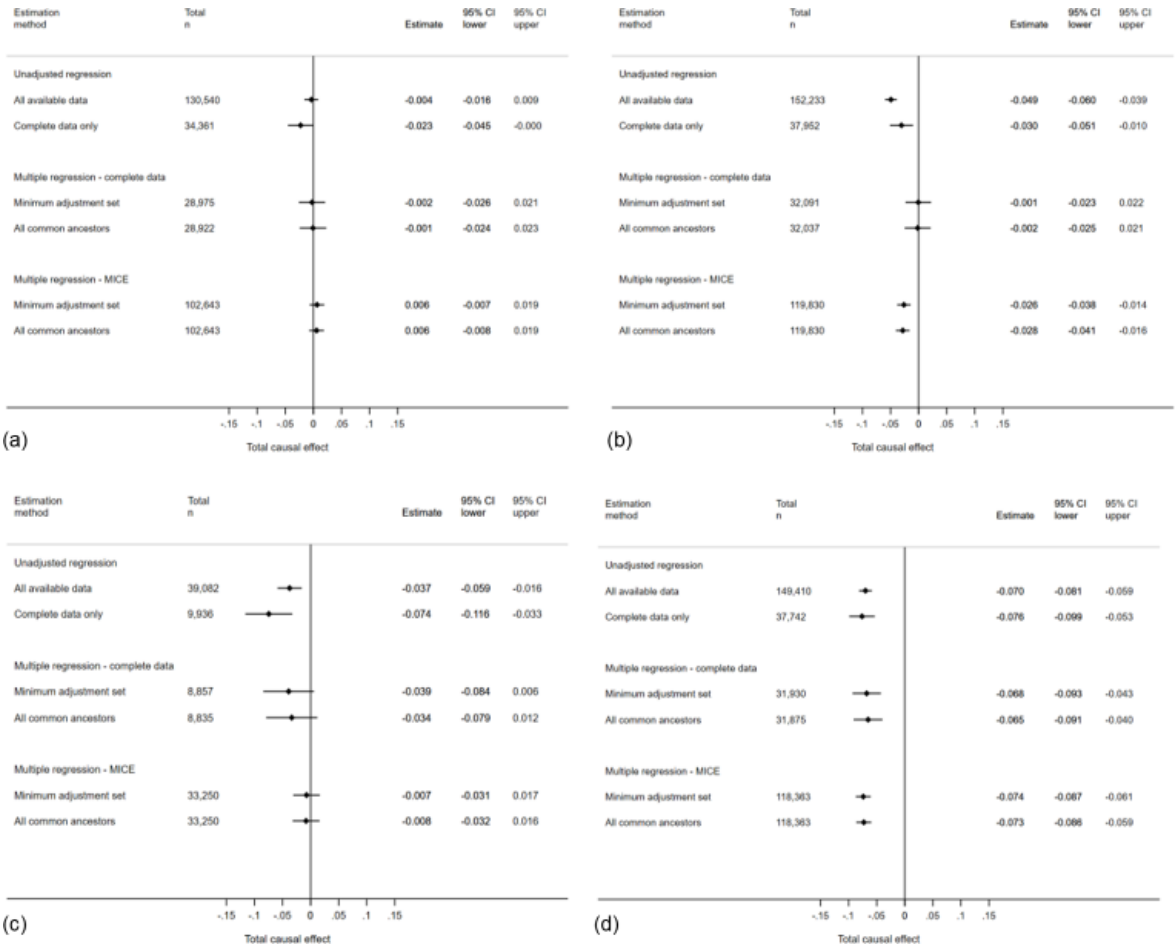


Figure 8.8 - Comparison of missing data approaches in major depression total effects analyses

CI, confidence interval; MICE, multiple imputation with chained equations. Panels show: (a) reasoning, (b) reaction time, (c) numeric memory and (d) visuospatial memory. Prospective memory not shown as it was not possible to calculate risk differences.

When the mediation models were repeated with imputation of missing mediator and covariate values, there remained no evidence of indirect effects via cardiometabolic disease (Table CC.3 in Appendix CC). There was evidence of an indirect effect via psychotropic medication on visuospatial memory performance (Table CC.4 in Appendix CC), accounting for approximately 18% of the total effect.

The results of the probabilistic analysis using `episens` indicated that the total effects estimate for visuospatial memory would be biased away from the null if there were differential misclassification of the exposure. When dichotomised into impaired and unimpaired outcome categories, and assuming no exposure misclassification, the unadjusted relative risk of impairment was 1.14 in the major depression group (95% CI 1.07, 1.20). Assuming lower sensitivity to true major depression status among the cognitively impaired (sensitivity range 0.6 to 0.9) versus unimpaired participants (sensitivity range 0.7 to 1.0), the relative risk was estimated as 1.37 (1.06, 1.75).

DAGitty determined that there were six other DAGs that were equivalent to the DAG shown in Figure 8.2 above. In each of these, path directions involving the ancestral history or parental genotype nodes were reversed (see Appendix DD). None of these alternative configurations was causally plausible, owing to temporal order constraints. The same minimum adjustment set was valid for the analysed DAG and for the six equivalent DAGs, for estimating the total effect of major depression on cognitive outcome.

8.2 Schizophrenia

8.2.1 Methods

8.2.1.1 Participants and exposure status

Participants were retained for analysis if they had sufficient data to classify their exposure status and had data on at least one cognitive outcome measure. The illness exposure was schizophrenia, classified using the broad definition given in Chapter 4 (i.e. met criteria according to self-reported diagnosis or pre-baseline hospital records; participants were classified as exposed if they met the criteria in at least one information source, even if the other information source was missing). The unexposed comparison group included participants who had complete self-reported data indicating no diagnosis of schizophrenia, and whose pre-baseline hospital records had no primary or secondary diagnosis of schizophrenia. Participants meeting the criteria for mania/BD or major depression (as defined earlier), according to the self-reported data, touchscreen mood questionnaire algorithm or hospital records, were excluded from the exposed and unexposed groups. Participants who did not meet the above criteria for either the exposed or unexposed groups were not further analysed.

8.2.1.2 Outcome measures and covariates

The cognitive outcome measures and covariates were as described in section 8.1.1.3 above, with the exception of the past mental state data: in addition to the information regarding past depressive episodes collected at baseline or in the subsequent web-based questionnaire (as described in Chapter 7), data were also collected in the web-based questionnaire regarding past "unusual experiences". Participants were asked "Did you ever": "…see something that wasn't really there, that other people could not see?"; "…hear things that other people said did not exist, like strange voices coming from inside your head talking to you or about you, or voices coming out of the air when there was no one around?"; "…believe that a strange force was trying to communicate directly with you by sending special signs or signals that you could understand but that no one else could understand (for example through the radio or television)?"; or "…believe that that there was an unjust plot going on to harm you or to have people follow you, and which your family and friends did not believe existed?". If they answered affirmatively, they were asked to state how many times each type of experience had occurred, when they were "not dreaming, not half-asleep, and not under the influence of alcohol or drugs", and this was recorded ordinally as 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, >10 and 'too many to count'. For the present analyses, the item (seeing things; hearing things; signs; plot) that had the highest reported frequency was used as a single index of the number of past unusual experiences; frequencies were not summed across the four items, on the assumption that these experiences may have happened concurrently. It was not possible to distinguish between experiences that occurred before or after the baseline assessment date. This variable was then summed with the variable coding the number of past episodes of depressed mood or anhedonia (described in Chapter 7), to produce a single ordinal variable representing the number of past episodes of depressed or psychotic-like experiences.

A GPS was created for schizophrenia, using summary statistics from the most recent GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). The GPS was generated for unrelated UK Biobank participants of white British genetic ancestry, using the same quality control criteria as described previously, and the model residuals after adjustment for genotyping array and batch, UK Biobank assessment centre, and the first 20 genetic principal components were used in the analyses. Appendix W shows the R^2 at each *p* threshold, and the odds ratio for the association between deciles of the optimum GPS and the schizophrenia phenotype in the UK Biobank cohort. The R^2 for the optimum GPS was 0.023. This is notably lower than the R^2 of 0.184 reported by the

GWAS authors when predicting into an independent sample (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

8.2.1.3 Data analysis

The DAG used in the schizophrenia analyses was the same as the final DAG used in the major depression analyses reported above (with the exposure and genotype nodes re-named as appropriate). Two versions of the DAG were evaluated: the first depicted a causal influence of educational attainment on schizophrenia and other psychiatric/neurological conditions, and the second depicted these relationships in reverse. The testable implied independencies were analysed as before. Total effects and mediation analyses were planned using the back-door criterion in DAGitty, as before.

8.2.2 Results

8.2.2.1 Analysis sample and evaluation of the graphical model

Figure 8.9 shows a flowchart of exclusions leading to the final analysis sample, which comprised 351 participants with schizophrenia and 360,122 comparison participants.

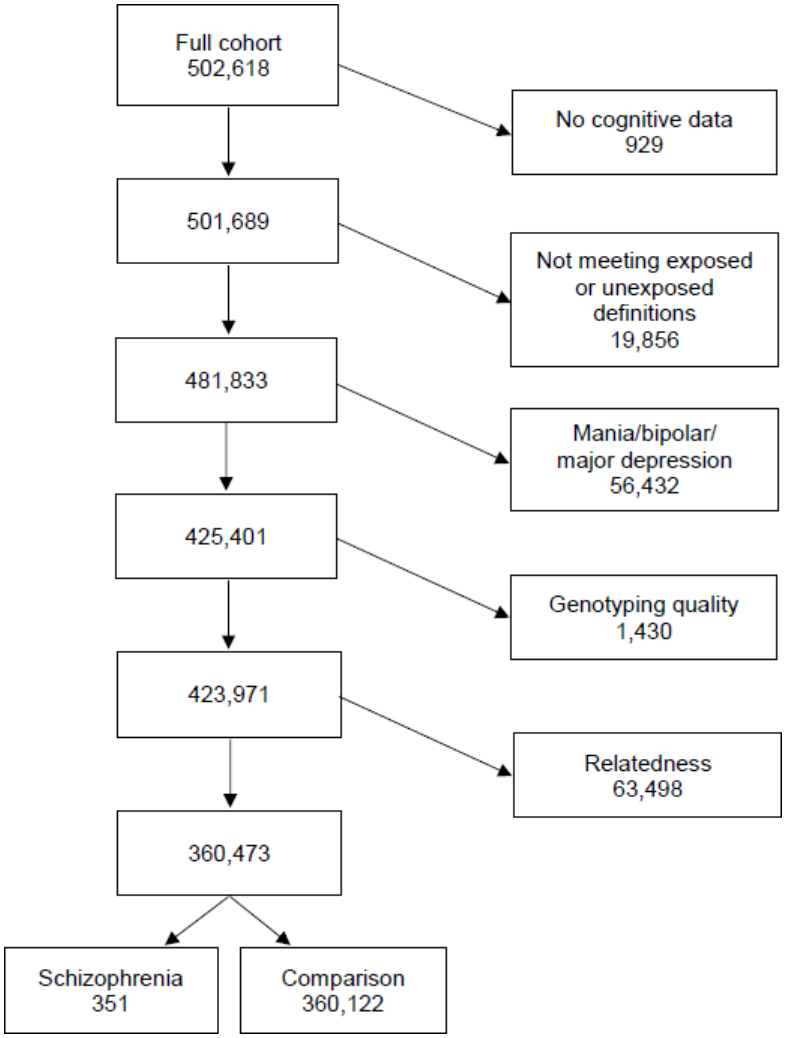


Figure 8.9 - Schizophrenia analysis sample flowchart

The different predicted independencies implied by the two specifications of the DAG were tested (see Appendix EE). Twenty percent (28 of 137) of the predicted independencies were above |0.10| in the better fitting model (educational attainment as an antecedent of schizophrenia and other psychiatric/neurological conditions), with two being above |0.40|. More detailed evaluation of these results confirmed that most of the highest correlations were reliably different from the null. Modifications to the DAG were therefore considered. Given the strong evidence of a relationship between cannabis use and schizophrenia (Gage et al., 2017), cannabis use (measured dichotomously as an indicator for ever having taken cannabis, using the web-based questionnaire data) was incorporated into the DAG (Appendix EE). The conditional independencies were tested again twice, with cannabis depicted as either an antecedent or a consequence of schizophrenia and other psychiatric/neurological conditions. The fit in both modified DAGs was worse than the original DAG that did not include cannabis (Appendix EE).

8.2.2.2 Causal estimation

Based on structural analysis of the best-fitting DAG, the covariate adjustment set for the total effect of schizophrenia on cognitive outcome was the same as that used in the major depression analyses described above (substituting the schizophrenia GPS and the expanded measure of depressive and psychotic-like episodes). An attempt was made to estimate a propensity score model for these covariates, but this could not proceed because only 24 participants in the schizophrenia group had complete covariate data, and only 15 of these were of white British genetic ancestry (a necessary model restriction). In light of the poor fit of the graphical model and the predominance of missing data, it was concluded that estimation of the causal effect of schizophrenia on cognitive outcome was not feasible in this dataset, and no further analyses were undertaken.

8.3 Discussion

The total causal effect of major depression on cognitive performance in the UK Biobank cohort followed a similar pattern to that of mania/BD, though with an effect size of around one-third the magnitude. This echoes previous evidence of a ‘dose-response’-type relationship between mood disorders and cognitive function, with major depression groups showing an intermediate level of impairment on average, between BD groups and those with no psychiatric history (Bora et al., 2013; Szmulewicz et al., 2017). There was equivocal evidence of an indirect causal pathway through psychotropic medication, but no evidence of mediation via cardiometabolic disease. The effect estimates are likely to be sensitive to even very low levels of residual confounding. As with the mania/BD analyses, widespread missingness in the covariates limited the interpretation of the results.

The finding of a robust—though very small—deficit in visuospatial memory but not on other cognitive tests was at odds with previous studies, which have reported deficits in a range of cognitive domains. The meta-analysis of Bora et al. (2013) found worse performance compared with healthy participants in every cognitive domain, with effect sizes in the medium range (-0.4 to -0.6), far exceeding the very small group difference observed here (-0.07). After multiple imputation of missing covariate data, additional evidence emerged of a very small group difference in reaction time (-0.02), although the validity of this result rests on the plausibility of the missing-at-random assumption, which may not be justifiable.

As with the mania/BD results reported in Chapter 7, the lack of evidence of a mediating effect via cardiometabolic disease was somewhat surprising, and may be explained by measurement error, adjustment for closely related risk factors, or the age of the sample, as noted previously. The evidence of mediation through psychotropic medication that was found in the mania/BD analyses was not so clearly apparent here: the point estimate of the proportion of the total effect transmitted through this path was similar to that in the mania/BD group (for the visuospatial memory outcome), but the confidence interval was wide and the estimate was reliably different from the null only after missing data imputation. A further issue to be considered here is the evidence for interaction between major depression and psychotropic medication; this indicates that any causal effect of major depression on cognitive outcome that is transmitted through psychotropic medication is non-linear. In other words, if non-depressed participants who are not on psychotropic medication were the reference group, the observed effect of both having depression and being on psychotropic medication would be different from the linear combination of the two separate effects of having depression but not being on psychotropic medication, and not having depression but being on psychotropic medication (Richiardi, Bellocco, & Zugna, 2013). This means that the mediation model results reported in this chapter represent a population average effect over both levels of the mediator, but these results will be sensitive to the population prevalence of being on psychotropic medication.

The strengths of the major depression analyses reported here include the sample size—exceeding 50,000 participants with depression, more than 50 times as large as the most recent meta-analysis of cognitive function in euthymic depression (Bora et al., 2013)—and the consideration of multiple consistently-measured covariates. The statistical analyses were informed by the graphical model and took account of interaction effects. As previously outlined in earlier chapters, there are a number of limitations in using the UK Biobank resource for this type of research, including assumptions regarding temporal order, measurement error and misclassification, and the possibility of collider bias arising from patterns of participation and missingness that may be related to mental health and cognitive status.

Extensive missing data in the schizophrenia group precluded any causal estimation analyses. The proportion with complete data on all the measures required for the multivariable analyses was only 7%, compared with 23% in the comparison group (and 19% and 23% in the mania/BD and major depression groups, respectively). Aside from the impact on study power, this pattern calls into question the validity of attempting inferential

analyses, given the non-representativeness of the sub-group with complete data, and the likelihood that the missingness mechanism was non-random and not amenable to statistical approaches such as multiple imputation. This illustrates the limitation of a general population cohort such as UK Biobank for conducting this type of research in a group with severe mental illness; accurate and representative assessment in this area may be better accomplished through smaller-scale clinic-based studies or whole population registries, although such studies are typically limited in statistical power (in the former case) and in depth of phenotyping (in the latter).

Chapter 9 Discussion

This chapter presents an overall summary and discussion of the findings of this thesis, in light of the research questions addressed, the methods chosen, and the strengths and limitations of the work undertaken. Implications for the clinical understanding of cognitive impairment in psychiatric and neurological conditions are considered, and future research directions are outlined.

9.1 Summary of findings

9.1.1 Prevalence of cognitive impairment in bipolar disorder and other psychiatric and neurological conditions

The aim of the systematic review presented in Chapter 2 was to determine the prevalence of cognitive impairment in euthymic adults with BD, and to ascertain which clinical, sociodemographic or other factors were associated with cognitive impairment in this population. Fifteen articles provided prevalence data, using a wide variety of cognitive tests and impairment thresholds. Taking the 5th percentile threshold as the reference, impairment prevalence ranges were: executive function 5.3% to 57.7%; attention/working memory 9.6% to 51.9%; speed/reaction time 23.3% to 44.2%; verbal memory 8.2% to 42.1%; visual memory 11.5% to 32.9%. Sample sizes were generally small, and the individual study estimates were consequently imprecise. There was some evidence that the prevalence or severity of cognitive impairment was associated with illness severity or duration. Associations were also reported with antipsychotic medication use, although these must be interpreted with caution, given the complex interplay between treatment indication, adherence and responsiveness.

In Chapter 5, baseline data from the UK Biobank cohort were analysed to quantify the prevalence of cognitive impairment in adults with a history of mania or BD, and to compare these results with estimates from adults with a history of major depression, schizophrenia, MS and PD. Using the 4th-5th percentile performance level in a single unexposed comparison group as the reference, the age- and gender-standardised prevalence of impairment across four cognitive tests ranged from 6.6% (95% CI 5.7, 7.5) to 10.4% (95% CI 7.7, 13.1) in the (broadly-defined) mania/BD group. Estimates were lower in the major depression group, ranging from 4.3% (95% CI 4.1, 4.5) to 5.7% (95% CI 5.3, 6.2), and were higher in schizophrenia, at 9.7% (95% CI 7.7, 11.7) to 20.6% (95% CI 10.6,

30.6). In MS, prevalence ranged from 3.2% (95% CI 1.8, 4.7) to 15.2% (95% CI 13.6, 16.8), and in PD, the range was 5.9% (95% CI 3.2, 8.7) to 12.7% (95% CI 6.0, 19.4). The prevalence of impairment was also elevated in most groups on a pass/fail test of prospective memory: compared with a failure prevalence of 22.8% in the unexposed comparison group, the standardised prevalence was 29.7% (95% CI 27.6, 31.7) in mania/BD, 22.8% (95% CI 22.4, 23.3) in major depression, 45.9% (95% CI 40.5, 51.3) in schizophrenia, 26.2% (95% CI 22.8, 29.7) in MS and 27.8% (95% CI 22.8, 32.9) in PD.

The relative pattern of impairment by clinical group and cognitive domain was in accordance with previous research, e.g. the schizophrenia group showed the highest prevalence on all tests (Bortolato et al., 2015) except reaction time, on which the MS group had the highest proportion with impairment (Rao et al., 2014). The absolute prevalence was lower than expected in each clinical group, however, compared with estimates from previous clinic-based studies. This may reflect truly lower cognitive impairment prevalence in general population-based illness groups than in those attending specialist clinics, or it may indicate collider bias arising from non-representative patterns of cohort participation and data completion among the more cognitively able people with these conditions. Collider bias is discussed in more detail in section 9.3.2 below.

Although the difference in prevalence (compared with the unexposed group) of cognitive impairment in major depression was relatively small, lifetime prevalence of major depression is approximately ten times that of BD and schizophrenia and fifty times that of MS and PD, which meant that the population attributable prevalence of cognitive impairment was highest overall for this group.

9.1.2 Explaining variation in cognitive function in bipolar disorder and other conditions

Chapter 7 presented the results of causal analyses to estimate the total effect of mania/BD on baseline cognitive performance in the UK Biobank cohort, and to estimate the proportion of any such effect that was transmitted indirectly through potentially modifiable intermediate factors. Analysis of the independencies implied by the structural causal graph indicated moderate fit to the data, once modifications had been made to take account of the influence of other psychiatric/neurological comorbidities. When background confounders were incorporated into the matched or adjusted analyses, a total causal effect of mania/BD was only evident on a test of short-term visuospatial memory. The magnitude of this effect was small, with the point estimates across the various matched/adjusted models being in

the range -0.23 to -0.17 standard deviation units (95% CI range across all estimates -0.39, -0.03). The small effect size relative to previous published research may reflect the larger sample size or less severe clinical status of the mania/BD group in UK Biobank, compared with clinic-based samples in other studies. There was evidence of an indirect causal pathway through psychotropic medication (accounting for approximately one-quarter of the total effect: -0.05; 95% CI -0.09, -0.01), but (perhaps surprisingly) not through cardiometabolic disease. The effect estimates are likely to be sensitive to residual confounding and misclassification, and they may be biased toward the null as a result of missing covariate data.

This work was extended to other psychiatric conditions in Chapter 8. For major depression, the graphical model was adapted to take account of gender differences in depression propensity, and the fit of the model in terms of implied independencies was deemed reasonable. The total causal effect of major depression on baseline cognitive performance in UK Biobank followed a similar pattern to that of mania/BD, though with an effect size of around one-third the magnitude: the point estimate across the various models was approximately -0.07 (95% CI range across all estimates -0.10, -0.03). The point estimate of the proportion of the total effect transmitted through psychotropic medication was similar to that in mania/BD, but the confidence interval was wide and the estimate was reliably different from the null only after imputation of missing data. Again, potential biases from residual confounding, misclassification and collider stratification are likely.

Similar models for cognitive performance in UK Biobank participants with schizophrenia were planned as part of the work reported in Chapter 8. The fit of the structural model was poorer in this study population, and problems were encountered as a result of widespread missing data. In light of this, multivariable causal models could not be conducted for the effect of schizophrenia on cognitive performance.

9.2 Contribution of this work to the literature

This thesis presents the first systematic review of the prevalence of cognitive impairment in euthymic BD. It complements and extends the findings of previous systematic reviews and meta-analyses (Arts et al., 2008; Bortolato et al., 2015; Bostock et al., 2017; Bourne et al., 2013; Dickinson et al., 2017; Mann-Wrobel et al., 2011; Raucher-Chene et al., 2017; Robinson et al., 2006), which have focused on the magnitude of between-group differences in cognitive test scores. Although mean score differences are important for understanding

the nature and extent of cognitive performance variation associated with BD, quantifying the number who have clinically relevant levels of cognitive impairment is essential in order to measure and understand the 'cognitive footprint' of this disorder, and to target clinical resources towards interventions and support for those who need it most. This work is also relevant to recent efforts to identify distinct clusters of people with BD on the basis of their cognitive performance patterns (Burdick et al., 2014; Sparding et al., 2017; Sparding et al., 2015), which have also highlighted the existence of heterogeneous sub-groups with different difficulties and needs.

This contribution was further extended by examining the prevalence of cognitive impairment in adults with mania/BD within a large population cohort, and comparing this with impairment prevalence in cohort sub-groups with a history of other psychiatric and neurological conditions. This is the first study to directly compare across these conditions, using consistent assessment methods and a common unexposed comparison group. Multiple sources of information were used to classify exposure status, and direct standardisation permitted like-for-like comparisons across conditions. This also enabled direct comparisons of population attributable prevalence, highlighting the contribution of both exposure prevalence and cognitive impairment prevalence to the overall 'cognitive footprint' of behavioural and brain disorders. The findings underscore the importance of making cognitive impairment a key focus of research and treatment in mood disorders, as is already the case in less prevalent neurological conditions such as MS and PD.

This thesis makes a novel contribution to the literature by applying, for the first time, a formal causal inference framework to identify and estimate total and indirect effects of mood disorders on cognitive performance. The complexity inherent in this area was acknowledged and addressed as far as possible, by developing and evaluating comprehensive graphical models and incorporating a broad range of genetic, sociodemographic, environmental, lifestyle and clinical measures in the analyses. Model estimation was conducted in multiple ways, and a series of quantitative and graphical sensitivity analyses were carried out to investigate the robustness of the results to various assumptions. The sample sizes were substantially larger than those used in previous studies in the field, allowing small effect sizes to be estimated with precision.

The results of the causal analyses indicated a reliable decrement, of small magnitude, in visuospatial memory in adults with a history of mania/BD, and a very small decrement on the same task in those with major depression. This is congruent with previous reports of a

gradation of severity of cognitive impairment in mood disorders (Szmulewicz et al., 2017). Any group differences seen in other cognitive functions such as reasoning and reaction time attenuated towards the null when background confounders were included in the analyses. The absence of differences on these tasks was surprising, in light of previous research showing multi-domain impairments, and it remains unclear to what extent this reflects insufficient adjustment for confounding in previous studies, the characteristics of the UK Biobank cohort, or the possibility that episodic memory is a particularly sensitive marker of cognitive function in mood disorders.

The results also contribute to the evidence base on the relationship between psychotropic medication and cognitive impairment, which has been repeatedly highlighted in previous observational studies (Balanzá-Martínez et al., 2010). Mediation analyses, taking account of intermediate confounders such as past depressive episodes, indicated that an appreciable proportion of the detrimental effect of mania/BD on visuospatial memory performance was carried via this pathway. The interplay between reasons for prescribing—especially of antipsychotic medications—and affective remission in understanding this relationship is not yet understood, but the present results appear to confirm that psychotropic medications warrant closer study as potential contributors to cognitive impairment in mood disorders.

9.3 Critical appraisal

9.3.1 Limitations of the prevalence studies

The issue of burden of cognitive impairment is complex and cannot be fully addressed by investigating the prevalence of sub-threshold test performance alone. Furthermore, there is no clear consensus regarding the best threshold to define impairment. It is likely that the disabling impact of cognitive impairment varies substantially even among those who score below a certain threshold (for example, depending on individual circumstances and levels of social support), and it is also possible that individuals can score above a threshold and yet experience functional disability related to cognitive difficulties. It was not possible to address these issues within the scope of this thesis. Also beyond the scope, and deserving of further study, is the identification and understanding of superior cognitive function in some people with psychiatric conditions, as previously highlighted in BD (Burdick et al., 2014).

The UK Biobank prevalence study was limited by the information sources available to identify the exposed and unexposed groups, which relied largely on self-reported data.

Large population registry-based studies have the advantage of access to a range of routinely recorded healthcare data; only hospital in-patient and day case records are currently available in UK Biobank, although primary care records will be incorporated into the resource in the coming years. Registry-based studies of whole populations remain ideally placed to investigate clinical outcomes such as dementia diagnoses in exposure groups that have also been identified through clinical records, but direct cognitive testing is not undertaken routinely in adulthood, and so prevalence studies of cognitive impairment still rely on research cohorts.

9.3.2 Limitations of the causal inference studies

The analyses were necessarily limited by the variables measured in the UK Biobank resource. A broad range of background and intermediate variables were available, but many of the measures were brief (e.g. four items to measure current mood state), and very little information was available regarding early life factors. Assumptions were made here about the temporal order of the variables, but these could not be verified empirically. For example, it was not possible to tell when a participant's cognitive performance reached the level at which it was measured at the UK Biobank baseline assessment, or whether this represented a change from a previous level of ability. The introduction of a premorbid ability estimate (vocabulary test) at the imaging follow-up visit will go some way to addressing this, for at least a subset of the cohort. The overall range of cognitive domains assessed has also been increased at the imaging visit, which will offer an improvement over the limited and idiosyncratic baseline assessments that were available for analysis in this thesis.

Assumptions were also made regarding the conditional exchangeability of the exposed and unexposed groups in the causal models, but sensitivity analyses using Rosenbaum bounds indicated that plausibly small levels of residual confounding would alter the findings. The SUTVA assumption may be deemed questionable, given the lack of a clear definition of what would constitute a well-defined contrast between having and not having a psychiatric condition, and how a hypothetical intervention on this could be conceptualised. This opens up criticisms regarding the true applicability of the analyses in the real world (Hernán, 2005).

The issue of collider stratification bias was noted throughout the thesis, and the problems this poses for causal inference in a cohort such as UK Biobank should not be

underestimated (Munafò et al., 2017). It is likely that people with less severe psychiatric and neurological phenotypes and better cognitive function will have joined the cohort, and this will have been amplified further in the patterns of missingness across the cognitive outcome measures and the covariates. The missing-at-random assumption that is required for multiple imputation is arguably not valid for some of the measures in these analyses, given the probability that, for example, missingness on mental health-related measures will be influenced by true mental health status. What is unknown, however, is the magnitude of the bias arising from collider stratification, and how this compares with similar or opposing biases from residual confounding.

Difficulties arising from measurement error and misclassification were also noted in earlier chapters. Quantitative sensitivity analysis methods are still being developed in this area, to encompass the range of issues arising from differential and non-differential misclassification of exposures, outcomes, mediators and covariates. SEM estimation approaches have a major advantage over other modelling methods here, in that they incorporate latent measures as an intrinsic part of model estimation. This does require additional data in the form of multiple indicators for each latent variable, or reliability coefficients which can be used to constrain error variances, but unfortunately there is little scope for applying these methods in UK Biobank, owing to the limited number of measures of each construct and the lack of reliability data.

A further advantage of SEM in causal modelling more generally is that the role of multiple mediators, and the inter-relationships between these and both background and intermediate confounders, are naturally accommodated in a single estimation model. This requires every path to be identified, however, which is not realistic in complex models with unmeasured nodes, such as those depicted in the DAGs in Chapters 7 and 8 here. Estimation can instead be approached on a localised basis, as was the case in this thesis, but this is not optimal when multiple mediators are of interest and would ideally be considered simultaneously in one model. Conducting estimation one mediator at a time may lead to erroneous conclusions about the contribution of each mediator to the overall effect, if the sum of the individual mediated effects does not equal the joint mediated effect (as will be the case when mediators cause, or interact with, each other) (VanderWeele & Vansteelandt, 2014; Vansteelandt & Daniel, 2017).

A final issue to note concerns the external validity of the results, including generalisability to the population from which the sample was drawn, and transportability to different

populations (Westreich et al., 2017). Weighting and imputation-based methods can be used to estimate the magnitude of bias in study estimates compared with a reference population, provided that data are available for comparison. Limited sociodemographic data regarding non-participating UK Biobank invitees have been retained (Fry et al., 2017), but these are not generally available to researchers. There is potential to use routine healthcare records from the same sampling frame as the study sample, to construct synthetic datasets via imputation within strata defined by combinations of outcome and confounder characteristics (Gray et al., 2013), but data on cognitive function are not captured in routine records. A further consideration is that the external validity of the estimates reported here is influenced by the population prevalence of variables that act in an interactive manner to produce outcomes, as noted in Chapter 8 with regard to the interaction between major depression exposure and psychotropic medication.

9.4 Clinical and public health implications

This thesis emphasises the importance of cognitive impairment as a persistent problem for a substantial proportion of adults with mood disorders, as was already recognised to be the case for those with schizophrenia and neurological conditions. Whether arising from the general background vulnerability and inequalities that affect these populations, or intrinsic to the nature of the disorders themselves, cognitive impairment may contribute to the overall picture of clinical need and should be recognised by practitioners and service managers as part of routine follow-up and service planning.

Even the crude prevalence estimates and test score differences reported here may be of interest to clinicians and service planners because, although they may have arisen in part from non-causal pathways (e.g. shared antecedents of both mood disorder and cognitive impairment), this has a real impact on the needs of the patients who attend clinics, and therefore shapes the services that should be provided in response. A different perspective may be taken by public health professionals, who will be more concerned with identifying modifiable causal pathways, on which interventions may be implemented to prevent or ameliorate adverse outcomes. These are conceptualised at the population level rather than the individual level, in that intervening on a certain causal pathway may be predicted to alter average outcomes in the population, even if the specific individuals who will benefit cannot be predicted in advance. Causal modelling and understanding of intermediate pathways are, of course, also important to professionals and patients alike, if they shed light on the fundamental mechanisms that explain the illness experience.

A greater understanding of the causes and consequences of cognitive impairment in mood disorders will have potentially widespread implications. There is scope to educate individuals with mood disorders and their families about the nature and impact of cognitive difficulties, as part of the broader picture of symptom recognition and self-management, and this in turn would have beneficial effects in dealing with the challenges of employment support and social participation. Clinical psychologists, occupational therapists, community mental health nurses and social care staff should be aware of the influence of cognitive impairment on day to day functioning, and take this into account when planning rehabilitation and support. The possible adverse effects of psychotropic medications should be of concern to psychiatrists and general practitioners; getting the balance right in terms of adequate symptom remission and minimal adverse consequences is difficult, and cognitive impairment should be recognised as part of that. Finally, consideration should be given to incorporating indicators of cognitive impairment and associated disability within routine outcome measurements, for the purposes of service evaluation and informing policy development.

9.5 Future directions

More work needs to be done to develop a meaningful measure of the 'cognitive footprint', including analysis of a wide range of health, social, educational and economic data to build a comprehensive picture of the population burden of chronic cognitive disability (Rossor & Knapp, 2015). Clinical researchers can aid these efforts by focusing on representative samples and by reporting numbers of participants falling below impairment thresholds of varying strictness, in addition to reporting between-group score differences.

Our understanding of causal pathways towards cognitive impairment in psychiatric and neurological conditions will improve as the UK Biobank cohort is followed up over time, and other prospective cohorts such as the Avon Longitudinal Study of Parents and Children (ALSPAC; Boyd et al., 2013; Golding, 1990) and the US National Longitudinal Study of Adolescent to Adult Health (Add Health; Klein, 1997) mature into adulthood. The availability of a fuller range of background and intermediate data, including early life factors, premorbid cognitive ability measures and brain imaging, will expand the kinds of causal effects that can be identified in models such as those proposed here. Linkage with prescribing data will permit more detailed investigation of the role of different classes of psychotropic medication, and combinations thereof, in explaining adverse cognitive outcomes.

A current barrier to progress in this area is that research cohorts often lack representativeness but offer deep and broad measurement of key data, while routine records offer population coverage but lack information about context, behavioural and psychosocial factors, and clinical data beyond that which can be captured in standardised coding systems. Significant investment is required to bridge this gap. It is essential that research cohorts are adequately resourced to achieve more representative recruitment and to maintain follow-up rates over time. This, in turn, requires research investment to improve our understanding of the determinants of participation and the optimal strategies which should be implemented in research cohorts. These efforts can be supplemented by improving approaches to sample weighting, and understanding and dealing with missing data. This is especially important for cognitive research, because cognitive function is likely to be a key driver of non-random missingness mechanisms. Improving linkage between research cohorts and routine administrative records should also be a priority, including taking advantage of the wealth of data contained in text-based records (e.g. clinical notes and correspondence; radiology reports). The coming years are likely to see continuing advances in text analytics and machine learning, and healthcare researchers must be ready to use these methods to enrich research capabilities. In the meantime, efforts to validate centrally-held code-based records against detailed local data will provide essential evidence regarding the quality and accuracy of the mental health data currently available to the research community (Davis et al., 2016; McIntosh et al., 2016; Stewart & Davis, 2016).

Even if the full potential of routine data for mental health research could be realised, however, investment in dedicated cohort studies will remain crucial. This is because research of the type undertaken in this thesis will always require detailed, dimensional measures of psychological traits and states, and multi-domain cognitive abilities, at multiple points in time. Furthermore, cheaper and more easily accessible MRI scanning means that neuroimaging assessments are no longer constrained to small studies, and should be viewed as a core component of large-scale observational research in mental health. Genomic data will also increasingly be seen as essential rather than optional; the recent step-change in GWAS discoveries and the growing appreciation of pleiotropy between psychological and physical traits means that no study can hope to provide a comprehensive causal account of cognitive function in mental health without incorporating genomic data.

Cognitive function is a fundamental phenotype in psychiatric and epidemiological research. As an outcome in its own right, as a transdiagnostic dimensional construct in ongoing efforts to improve psychiatric nosology (Insel et al., 2010), and as a determinant of health behaviours, morbidity and mortality (Deary, 2009), it deserves increased research attention.

Appendices

A – PROSPERO registration

UNIVERSITY of York
Centre for Reviews and Dissemination

NHS
National Institute for Health Research

PROSPERO International prospective register of systematic reviews

Systematic review of prevalence and correlates of cognitive impairment in euthymic bipolar disorder

Breda Cullen, Ian Deary, Jill Pell, Daniel Smith, Jonathan Evans

Citation

Breda Cullen, Ian Deary, Jill Pell, Daniel Smith, Jonathan Evans. Systematic review of prevalence and correlates of cognitive impairment in euthymic bipolar disorder. PROSPERO 2015:CRD42015017558 Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42015017558

Review question(s)

What is the prevalence of cognitive impairment in euthymic adults with a history of bipolar disorder (BD)?

What clinical, sociodemographic and lifestyle factors are significantly associated with cognitive impairment, in addition to diagnosis of BD?

Searches

The following electronic databases will be searched:

• Web of Science (Thomson Reuters), including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Current Contents Connect, Data Citation Index, MEDLINE and SciELO Citation Index

• PubMed (NCBI), including MEDLINE, PubMed Central and in-process/ahead-of-print citations

• EBSCOhost (EBSCO), including CINAHL and PsycINFO.

Searches will be restricted to English language publications from 01.01.1994 to 24.02.2015. The full search strategy is available in the protocol.

Where available, the 'cited by' function will also be used within individual electronic records of relevant articles. Reference lists of relevant articles and recent review papers will be hand-searched. Key authors in the field will be contacted in order to identify additional studies.

Types of study to be included

Cross-sectional studies (standalone, or part of a prospective study).

Condition or domain being studied

Cognitive impairment (evidence of impaired performance on one or more objective cognitive tests). The threshold for impairment will be as defined by the study authors, but will at a minimum require performance to be at least 1SD below the mean performance of a healthy comparison group (e.g. based on published test norms, or an appropriate external comparison group recruited to the study), or to be in the fail range on a pass/fail test.

Participants/ population

Euthymic community-dwelling adults aged 18-70.

Intervention(s), exposure(s)

History of bipolar disorder type I or II, meeting DSM/ICD criteria.

Comparator(s)/ control

Euthymic community-dwelling adults without a history of bipolar disorder.

Outcome(s)

Primary outcomes

Prevalence of cognitive impairment (percent of sample with performance below a defined threshold).

Secondary outcomes

Any clinical, sociodemographic or lifestyle factor that is reported by the authors to be significantly associated with cognitive impairment.

Data extraction, (selection and coding)

Study titles and/or abstracts will be screened for relevance by the first author. Screening will be carried out with reference to a detailed checklist of eligibility criteria; this will have been piloted by two researchers independently against a sample of initial search results, and refined as required. The sensitivity of the search strategy will be checked by testing whether a sample of papers that are known to be relevant are detected by the search. Reproducibility will be assessed by a second researcher, who will independently run the search in one electronic database and will screen a sample of titles and/or abstracts for relevance.

Full text will be obtained for all potentially relevant papers that remain. These will be assessed by the first author using the eligibility checklist, with a second researcher independently assessing a sample of papers for comparison. Discrepancies will be resolved by consensus, involving a third researcher as required. Reasons for exclusion will be documented.

A pro-forma Excel spreadsheet will be used for extracting data from included papers, having been piloted by two researchers independently against a sample of five papers. This will include:

• Citation details: authors, year, title, journal, volume, pages

• Study funding source

• Setting: general community or clinical out-patient

• Comparison group: none/internal/external; matched or not

• Method of recruitment for clinical group, and comparison group if included

• Eligibility criteria applied (e.g. exclusion of other diagnoses) for clinical group, and comparison group if included

• Power/sample size calculation done: yes/no

• Sample size of clinical group, and comparison group if included

• Sample demographics for clinical group, and comparison group if included: age (mean and SD, or median and IQR), sex, ethnicity, country, language, education level, socioeconomic status

• Qualifications and training of study assessors

• BD definition

• Euthymia definition

• Age at onset of BD (mean and SD, or median and IQR)

• Number of previous depressive episodes

• Number of previous manic episodes

• Current medication status: % of sample(s) currently taking psychotropic medication; details of medication type and dosage

• Cognitive assessments used: test names and source references; domains covered

• Cut-offs applied for impaired/not impaired

• Prevalence of impairment in clinical group, and comparison group if included, across each cognitive measure

• Associated factors: details of correlates tested; method of testing for significance; result of test for significance

• Proportion of missing data and how this was dealt with in the analysis.

Where studies appear to have collected data that could be used to report prevalence of impairment, but have not reported prevalence explicitly in the paper (e.g. articles only reporting group differences on test scores), the study authors will be contacted in order to request prevalence results using an appropriate cut-off.

Data extraction will be carried out by the first author. A second researcher will compare a random sample of the data extraction forms against the source papers to check for accuracy and completeness. Any ambiguities in completing the forms will be discussed with other members of the research team until consensus is reached.

Risk of bias (quality) assessment

Each included study will be assessed for methodological quality and bias using a critical appraisal tool for prevalence studies (Munn, Moola, Riitano, & Lisy, 2014). Reporting bias will be assessed using the STROBE checklist for cross-sectional studies (von Elm et al., 2007). Assessments will be made by the first author, and a second researcher will independently rate a random sample of papers for comparison. Rating discrepancies will be resolved by consensus, involving a third researcher as required.

These bias assessments will be considered in the review synthesis and discussion in order to comment on study quality in this field and to aid interpretation of the validity of the results.

Strategy for data synthesis

Where one dataset has been reported on in two or more eligible articles, only the article reporting the largest sample will be considered in the data synthesis.

A narrative synthesis will be presented together with a tabulated summary of study characteristics and results extracted from each article. The synthesis may be subdivided according to key factors such as number of illness episodes, or different definitions of cognitive impairment. Significant correlates of impairment will be described, and consistency in these findings compared across studies.

We expect that variation in cognitive tests used, and in cut-offs applied to define presence of impairment, will preclude meta-analysis of prevalence estimates.

Analysis of subgroups or subsets

The synthesis may be subdivided according to key factors such as number of illness episodes, or different definitions of cognitive impairment.

Dissemination plans

The results will be submitted for publication in peer reviewed journals and will be included in the first author's PhD thesis.

Contact details for further information

Dr Cullen

R4, University Section,

Ground Floor, Office Block

Queen Elizabeth University Hospital,

Glasgow G51 4TF

breda.cullen@glasgow.ac.uk

Organisational affiliation of the review
University of Glasgow

http://www.gla.ac.uk/researchinstitutes/healthwellbeing/

Review team
Dr Breda Cullen, University of Glasgow
Professor Ian Deary, University of Edinburgh
Professor Jill Pell, University of Glasgow
Professor Daniel Smith, University of Glasgow
Professor Jonathan Evans, University of Glasgow

Collaborators
Mr Joey Ward, University of Glasgow
Dr Nicholas Graham, University of Glasgow

Anticipated or actual start date
02 February 2015

Anticipated completion date
31 December 2015

Funding sources/sponsors
Breda Cullen is supported by a personal fellowship from the Scottish Executive Health Department (DTF/14/03).

Conflicts of interest
None known

Language
English

Country
Scotland

Subject index terms status
Subject indexing assigned by CRD

Subject index terms
Bipolar Disorder; Cognition; Cognition Disorders; Cyclothymic Disorder; Humans; Prevalence

Stage of review
Completed and published

Date of registration in PROSPERO
16 March 2015

Date of publication of this revision
25 April 2017

Details of final report/publication(s)
Cullen B, Ward J, Graham NA, Deary IJ, Pell JP, Smith DJ, Evans JJ. Prevalence and correlates of cognitive

impairment in euthymic adults with bipolar disorder: a systematic review. Journal of Affective Disorders. 2016. 205:165-181
http://dx.doi.org/10.1016/j.jad.2016.06.063

DOI
10.15124/CRD42015017558

Stage of review at time of this submission	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	Yes	Yes

B – Search strategy for systematic review

Search strategy implemented in Web of Science

Search set-up: English; publication year 1994 to date

1	TS=((bipolar NEAR/3 depress*) OR (bipolar NEAR/3 disorder*) OR (manic NEAR/0 depress*))
2	TS=(cogniti* OR neuro-cogniti* OR neurocogniti* OR neuro-psycholog* OR neuropsycholog* OR speed OR reaction OR attention OR memory OR learning OR *spatial OR executive OR reasoning OR IQ OR intelligence)
3	TS=(impair* OR dysfunction* OR declin* OR deteriorat* OR defici*)
4	(#1 AND #2 AND #3)
5	TI=(therapy OR CBT OR cognitive-behavio* OR (cognitive NEAR/0 behavio*))
6	#4 NOT #5
7	TS=((nursing NEAR/0 home*) OR (care NEAR/0 home*))
8	#6 NOT #7
9	TS=(dement*)
10	#8 NOT #9

Results limits: Refined by: Databases: (WOS OR SCIELO OR MEDLINE OR CCC OR DRCI) AND DOCUMENT TYPES: (ARTICLE OR CLINICAL TRIAL OR DATA SET OR UNSPECIFIED OR OTHER OR DATA STUDY) AND LANGUAGES: (ENGLISH)

C – Eligibility checklist for systematic review

Eligibility checklist used when screening titles, abstracts and full text

No.	Criteria and definitions
1	**Original research published in peer-reviewed journals, from 1994 until date of search** ▪ Not editorials, opinion papers, reviews, meta-analyses (but 'individual patient data meta/mega-analyses' that involve re-analysis of raw data [rather than group effect sizes] are acceptable) ▪ Not conference proceedings, books, book chapters, academic theses ▪ Not single case studies or case series (must be group studies) ▪ If unsure whether journal is peer-reviewed, check the journal website or look up the title at https://ulrichsweb.serialssolutions.com/ and look for the 'refereed' symbol:  ▪ Publication year is 1994 or later
2	**Articles published in English** ▪ Full text of the article must be available in English
3	**Studies of community-dwelling adults aged 18 to 70 years inclusive** ▪ Out-patient service or general population setting (this refers to the setting in which participants were recruited and/or the study took place) ▪ BD participants may be recruited from clinic attendees, or from a population study (e.g. large-scale population registry) ▪ Participants must have been living in the community at time of assessment ▪ Not currently a hospital in-patient (but hospital/clinic out-patient is acceptable) ▪ Not living in a nursing/care home ▪ Participants are adults of minimum age 18 years and maximum age 70 years (abstracts describing samples as 'adolescent' or 'elderly' are presumed to be ineligible unless they also mention a separate sample of 'adults' in the study) ▪ There may be a healthy comparison group in the study, but this is not a requirement
4	**Cross-sectional study design** ▪ The key criterion is that the assessment of mood disorder status (whether the person has euthymic BD – the 'exposure') and the assessment of cognitive performance (the 'outcome') took place at the same time (maximum allowable time gap between establishing euthymic status and carrying out cognitive assessment = 2 weeks) ▪ The mood disorder may have been diagnosed in the past; this is acceptable (as long as the confirmation of current euthymic status took place at the same time as the cognitive assessment – see section 8 below) ▪ A cross-sectional study of this type may be part of a prospective/longitudinal/cohort study or a treatment trial; this is acceptable as long as separate (standalone) results are presented for the cross-sectional component that we are interested in ▪ Studies investigating whether cognitive status predicts *future* remission of mood disorder are not eligible
5	**If clinical out-patient service setting, samples must be consecutively recruited** ▪ Where BD participants have been recruited via a clinical service, recruitment must have been consecutive and representative of the target group ▪ This means participants should not have been selectively approached (e.g. on the basis of their cognitive function or some other characteristic) in a way that might bias the results of the study ▪ All eligible patients in the target group should have had an equal chance of being approached ▪ For example, if the target group was patients with euthymic BD, and the researchers considered all such patients attending the clinic between time X and time Y (or a randomly chosen subset of these), then that would be acceptable
6	**Primary diagnosis of bipolar disorder (BD)** ▪ Primary means the disorder has been diagnosed in its own right, not secondary to another illness ▪ BD = History of bipolar disorder (type I or II or unspecified), meeting DSM or ICD criteria ▪ Evidence for meeting diagnostic criteria may come from direct assessment as part of the study (often including a semi-structured interview schedule), or from a doctor's diagnosis recorded in the medical notes ▪ Questionnaire measures of mood state alone (without reference to a diagnostic reference system) are not acceptable

	▪ The diagnosis may have been made at any time in the person's life up until the time of the study
7	**Euthymic at time of assessment** ▪ Not meeting DSM or ICD criteria for a depressive or manic episode at time of cognitive assessment; or as otherwise defined by the study authors based on an appropriate clinical measure ▪ For example, the authors may define euthymia as being below X threshold on a depression or mania rating scale ▪ The concept of euthymia may be referred to as remitted/remission, or recovery, or stable on treatment, or treatment-responsive (or other similar phrase) ▪ Non-euthymic patients may be described as acutely manic/unwell, or treatment-resistant (or other similar phrase) ▪ Baseline samples within treatment trial studies are presumed to be NOT euthymic unless the abstract says otherwise
8	**Assessed using at least one standardised cognitive measure** ▪ The measure should be an objective test, on which the participant's performance is assessed directly ▪ Self-report questionnaires are not acceptable (e.g. the participant rates how good they think their cognition is) ▪ Informant-rated questionnaires are not acceptable (e.g. the participant's spouse rates how good they think the participant's cognition is) ▪ Informal behavioural observations are not acceptable (e.g. researcher observes participant without using any standardised rating scale) ▪ The cognitive test may cover one or more of the following abilities/domains: · Global/overall function (cognitive/neuropsychological) · Processing speed/psychomotor speed/reaction time · Attention/vigilance/alertness/concentration · Working memory/memory/learning/encoding/recall/recognition/retrieval · Spatial/visuospatial ability · Language/naming/comprehension · Executive function (including planning, strategy-formation, problem-solving, decision-making, initiation, self-monitoring, self-regulation, mental control, goal management, goal neglect, inhibition, response suppression, fluency, word generation, perseveration, set-shifting, rule-shifting, flexibility, impulsivity, sequencing, dual-tasking, multi-tasking) · Reasoning/abstraction/concept formation/IQ/intelligence · Social cognition (including theory of mind, meta-cognition – e.g. judgement of other people's thoughts/behaviours) ▪ Studies of non-conscious learning (e.g. classical conditioning and extinction) are not eligible ▪ Studies using only experimental neuroimaging tasks (e.g. oddball/continuous performance tasks that are not interpretable in their own right) are not eligible ▪ Studies of basic emotional processing only, without an explicit social cognition aspect, (e.g. reaction times to emotional faces) are not eligible ▪ The test should yield a numeric score, or a rating (e.g. pass/fail, or poor/fair/good, or impaired/unimpaired) ▪ *NOTE*: The mood disorder literature also contains many studies about cognitive distortions/biases. For example, a study may look at biased thinking patterns, attitudes, or rumination. This can be thought of as *cognitive style*, which is not the same as cognitive function in the neuropsychological sense outlined above. Therefore studies which are solely about these cognitive distortions, including treatments such as CBT aimed at changing these distortions, are not relevant to this review. ▪ Similarly, some studies may focus on *formal thought disorder* (e.g. tangentiality, flight of ideas); these are not eligible unless a cognitive assessment of the type outlined above is also conducted.
9	**Samples NOT selected on the basis of presence of cognitive impairment (known or suspected)** ▪ Participants (BD participants, or healthy comparison participants) should have been recruited (approached/selected) based on their exposure status (having or not having BD), and NOT based on their outcome status (having or not having cognitive impairment) ▪ It is possible that potential participants who were approached to take part might have been more or less likely to agree depending on their cognitive status; this is outside the researcher's control and so is not the focus here. Rather, the issue is that the *initial approach/selection by the researcher* should not be based on cognitive status

D – Data extraction form for systematic review

Instructions for data extraction

Spreadsheet field no.	Description	Entry format
1	Publication ID	Number
2	Sub-ID (use if study contains separate results for >1 BD sample)	Number-letter
3	Authors (e.g. Bloggs, A.B., Jones, C.D. & Smith, E.F.)	Free text
4	Year	XXXX
5	Journal	Free text
6	Volume	Free text
7	Pages	XX-YY
8	Title	Free text
9	Corresponding author's name and email	Free text
10	Study funding source	Free text or Not stated
11	Any other conflicts of interest declared	Free text or Not stated
12	Study setting: type	1=general community (e.g. population cohort) 2=clinical service 3=other 4=not stated
13	Study setting: details	Free text or Not stated
14	Population from which BD sample was drawn (e.g. all BD patients known to clinical service between date X and date Y; all families with 2 or more BD patients, etc)	Free text or Not stated
15	Healthy comparison group included	1=none (no healthy controls) 2=internal (from same population e.g. BD and controls taken from same general pop cohort) 3=external (from different population e.g. BD from clinic and controls from posters in community)
16	Comparison group matching (NB this refers to planned *matching during recruitment* – i.e. matched study *design* - not whether some characteristics happened to be similar after analysis)	1=not matched 2=matched to BD at group level only (unpaired design) 3=matched to BD individually (paired design) 4=NA 5=not stated
17	Comparison group: characteristics matched by	Free text or NA or Not stated
18	Recruitment procedure for BD group	Free text or Not stated
19	Recruitment procedure for comparison group	Free text or NA or Not stated
20	Inclusion criteria for BD group	Free text or Not stated
21	Exclusion criteria for BD group	Free text or Not stated
22	Inclusion criteria for comparison group	Free text or NA or Not stated
23	Exclusion criteria for comparison group	Free text or NA or Not stated
24	Informed consent obtained from all participants	1=yes 2=no 3=not stated
25	Power/sample size calculation reported for the study	1=yes 2=no
26	Country where study took place	Free text or Not stated
27	Language in which cognitive test was administered (assume English if test name and citation are for English version AND study is from English-speaking country)	Free text or Not stated

28	BD sample: sample size	n or Not stated
29	BD sample: age (mean and SD, or median and IQR, or range)	Free text or Not stated
30	BD sample: sex (n and % male)	Free text or Not stated
31	BD sample: ethnicity (n and % in each category)	Free text or Not stated
32	BD sample: education (info re years or qualifications)	Free text or Not stated
33	BD sample: socioeconomic status (how measured, and status of sample)	Free text or Not stated
34	Comparison sample: sample size	n or NA or Not stated
35	Comparison sample: age (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
36	Comparison sample: sex (n and % male)	Free text or NA or Not stated
37	Comparison sample: ethnicity (n and % in each category)	Free text or NA or Not stated
38	Comparison sample: education (info re years or qualifications)	Free text or NA or Not stated
39	Comparison sample: socioeconomic status (how measured, and status of sample)	Free text or NA or Not stated
40	Qualifications/training of person who made BD diagnosis	Free text or Not stated
41	Qualifications/training of person who did cognitive assessment	Free text or Not stated
42	BD definition (e.g. DSM IV/ICD-10 criteria)	Free text or Not stated
43	Euthymia definition (e.g. score less than X on named questionnaire, or clinician judgement of remission, etc)	Free text or Not stated
44	Euthymia confirmed at time of cognitive assessment	1=yes 2=no 3=not stated
45	Age at onset of BD (mean and SD, or median and IQR, or range)	Free text or Not stated
46	Duration since onset of BD (mean and SD, or median and IQR, or range)	Free text or Not stated
47	Number of previous illness episodes, all types combined (mean and SD, or median and IQR, or range)	Free text or Not stated
48	Number of previous depressive episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
49	Number of previous manic/hypomanic episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
50	Number of previous mixed episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
51	BD sample: n and % currently taking any psychotropic medication	Free text or Not stated
52	BD sample: further details of medication type and dosage	Free text or Not stated
53	Comparison sample: n and % currently taking any psychotropic medication	Free text or NA or Not stated
54	Comparison sample: further details of medication type and dosage	Free text or NA or Not stated
55	Cognitive assessment: name of test	Free text
56	Cognitive assessment: source reference (journal citation; or name of publishing company)	Free text or Not stated
57	Cognitive assessment: domain covered (e.g. verbal memory)	Free text or Not stated

58	Cognitive assessment: definition of impairment threshold (e.g. 2SD below control mean score on one test; or 2SD below control mean score on at least two tests; or 1.5SD below published norms; or scores less than X; etc)	Free text or Not stated
59	Prevalence of cognitive impairment in BD sample (n and % who meet the stated definition of impairment given in field 58)	Free text or Not stated
60	Prevalence of cognitive impairment in comparison sample (n and % who meet the stated definition of impairment given in field 58)	Free text or NA or Not stated
61	Source of prevalence info (or explanation for missing info)	1=provided in paper 2=author provided on request 3=info missing: author replied but did not provide 4=info missing: author did not reply
62	Statistical analysis method to test/adjust for confounders/covariates (e.g. multiple regression)	Free text or NA or Not stated
63	Confounders/covariates (apart from other cognitive measures) that were significantly associated with cognitive performance in the analysis (e.g. higher age was a significant independent predictor of worse performance)	Free text or NA or Not stated
64	Missing data: brief description of level of missing data in the analysis (e.g. 10% of BD patients were missing a memory score)	Free text or Not stated
65	Missing data: brief description of how authors dealt with missing data (e.g. complete case analysis, or all available data, or imputation)	Free text or Not stated
66	Any other comments	Free text
67 *	Group cognitive score: BD sample (unadjusted group average) (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
68 *	Group cognitive score: comparison sample (unadjusted group average) (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
69 *	Cohen's *d* (if M and SD available in fields 67 and 68). Minus sign = BD worse; calculate using http://www.uccs.edu/~lbecker/	Free text or NA or Not stated
70	Full list of all confounders/covariates tested (apart from other cognitive measures), whether significant or not (additional info for field 63)	Free text or NA or Not stated

* Optional fields – only complete these for papers for which information is available regarding prevalence of impairment (fields 59 and 60), otherwise enter NA

E – Eligible articles excluded from systematic review synthesis

Supplementary references for eligible articles not included in synthesis

Antila, M., Partonen, T., Kieseppa, T., Suvisaari, J., Eerola, M., Lonnqvist, J., & Tuulio-Henriksson, A. (2009). Cognitive functioning of bipolar I patients and relatives from families with or without schizophrenia or schizoaffective disorder. *Journal of Affective Disorders, 116*(1-2), 70-79. doi:10.1016/j.jad.2008.11.006

Antila, M., Tuulio-Henriksson, A., Kieseppa, T., Eerola, M., Partonen, T., & Lonnqvist, J. (2007). Cognitive functioning in patients with familial bipolar I disorder and their unaffected relatives. *Psychological Medicine, 37*(5), 679-687. doi:10.1017/s0033291706009627

Braw, Y., Sitman, R., Sela, T., Erez, G., Bloch, Y., & Levkovitz, Y. (2012). Comparison of insight among schizophrenia and bipolar disorder patients in remission of affective and positive symptoms: Analysis and critique. *European Psychiatry, 27*(8), 612-618. doi:10.1016/j.eurpsy.2011.02.002

Donaldson, S., Goldstein, L. H., Landau, S., Raymont, V., & Frangou, S. (2003). The Maudsley Bipolar Disorder Project: The effect of medication, family history, and duration of illness on IQ and memory in bipolar I disorder. *Journal of Clinical Psychiatry, 64*(1), 86-93.

Frangou, S. (2005). The Maudsley Bipolar Disorder Project. *Epilepsia, 46*, 19-25.

Goswami, U., Sharma, A., Khastigir, U., Ferrier, I. N., Young, A. H., Gallagher, P., . . . Moore, P. B. (2006). Neuropsychological dysfunction, soft neurological signs and social disability in euthymic patients with bipolar disorder. *British Journal of Psychiatry, 188*, 366-373. doi:10.1192/bjp.188.4.366

Haldane, M., & Frangou, S. (2005). The Maudsley Bipolar Disorder Project: Insights into the role of the prefrontal cortex in bipolar disorder I. *Revista de Psiquiatria do Rio Grande do Sul, 27*(3), 1-20.

Kauer-Sant'Anna, M., Yatham, L. N., Tramontina, J., Weyne, F., Cereser, K. M., Gazalle, F. K., . . . Kapczinski, F. (2008). Emotional memory in bipolar disorder. *British Journal of Psychiatry, 192*(6), 458-463. doi:10.1192/bjp.bp.107.040295

Martinez-Aran, A., Penades, R., Vieta, E., Colom, F., Reinares, M., Benabarre, A., . . . Gasto, C. (2002). Executive function in patients with remitted bipolar disorder and schizophrenia and its relationship with functional outcome. *Psychotherapy and Psychosomatics, 71*(1), 39-46. doi:10.1159/000049342

Mur, M., Portella, M. J., Martinez-Aran, A., Pifarre, J., & Vieta, E. (2009). Influence of Clinical and Neuropsychological Variables on the Psychosocial and Occupational Outcome of Remitted Bipolar Patients. *Psychopathology, 42*(3), 148-156. doi:10.1159/000207456

Palsson, E., Figueras, C., Johansson, A. G. M., Ekman, C.-J., Hultman, B., Ostlind, J., & Landen, M. (2013). Neurocognitive function in bipolar disorder: a comparison between bipolar I and II disorder and matched controls. *BMC Psychiatry, 13*, 165. doi:10.1186/1471-244x-13-165

Strejilevich, S. A., & Martino, D. J. (2013). Cognitive function in adulthood and elderly euthymic bipolar patients: A comparison to test models of cognitive evolution. *Journal of Affective Disorders, 150*(3), 1188-1191. doi:10.1016/j.jad.2013.05.012

Torralva, T., Gleichgerrcht, E., Torrente, F., Roca, M., Strejilevich, S. A., Cetkovich, M., . . . Manes, F. (2011). Neuropsychological functioning in adult bipolar disorder and ADHD patients: A comparative study. *Psychiatry Research, 186*(2-3), 261-266. doi:10.1016/j.psychres.2010.08.007

Trivedi, J. K., Goel, D., Sharma, S., Singh, A. P., Sinha, P. K., & Tandon, R. (2007). Cognitive functions in stable schizophrenia & euthymic state of bipolar disorder. *Indian Journal of Medical Research, 126*(5), 433-439.

van Gorp, W. G., Altshuler, L., Theberge, D. C., & Mintz, J. (1999). Declarative and procedural memory in bipolar disorder. *Biological Psychiatry, 46*(4), 525-531. doi:10.1016/s0006-3223(98)00336-9

van Gorp, W. G., Altshuler, L., Theberge, D. C., Wilkins, J., & Dixon, W. (1998). Cognitive impairment in euthymic bipolar patients with and without prior alcohol dependence - A preliminary study. *Archives of General Psychiatry, 55*(1), 41-46. doi:10.1001/archpsyc.55.1.41

F – Critical appraisal ratings for systematic review

	1. Representative sample	2. Appropriate recruitment	3. Sample size	4. Subjects & setting described	5. Analysis coverage	6. Objective measure	7. Reliable measure	8. Statistical analysis	9. Confounding factors	10. Objective subgroupings
Altshuler 2004	N	Y	N	Y	?	Y	?	Y	Y	Y
Arslan 2014	N	Y	N	Y	?	Y	Y	Y	Y	Y
Barrera 2013	N	Y	N	N	?	Y	?	?	Y	NA
Cavanagh 2002	Y	Y	N	Y	Y	Y	Y	?	Y	NA
Cheung 2013	N	Y	Y	Y	?	Y	Y	Y	Y	Y
Daban 2012	Y	Y	Y	N	?	Y	?	Y	Y	Y
Doganavsargil-Baysal 2013	N	Y	N	N	?	Y	?	Y	Y	Y
Elshahawi 2011	Y	Y	Y	Y	?	Y	?	Y	Y	Y
Fakhry 2013	Y	Y	N	Y	?	Y	?	Y	Y	Y
Ferrier 1999	N	Y	N	Y	?	Y	?	Y	Y	Y
Frangou 2005	Y	Y	N	Y	?	Y	Y	Y	Y	NA
Goswami 2009	Y	Y	N	N	?	Y	Y	Y	Y	Y
Ibrahim[a] 2009	Y	Y	N	Y	?	Y	Y	?	Y	?
Jamrozinski 2009	Y	Y	N	N	?	Y	?	Y	Y	Y
Juselius[b] 2009	N	Y	N	Y	?	Y	?	Y	Y	NA
Kieseppä[b] 2005	N	Y	N	Y	Y	Y	?	Y	Y	Y
Lopera-Vásquez 2011	?	Y	N	N	?	Y	?	Y	Y	Y
López-Jaramillo 2010	N	Y	N	Y	?	Y	?	Y	Y	Y
Martino[c] 2008	N	Y	N	N	?	Y	Y	Y	Y	Y
Martino[c] 2011a	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2011b	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2011c	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2014	N	Y	Y	Y	?	Y	Y	Y	Y	Y
Mur 2007	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Normala[a] 2010	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Osher 2011	Y	Y	Y	Y	?	Y	Y	Y	Y	Y
Pirkola[b] 2005	N	Y	N	Y	?	Y	Y	Y	Y	NA

	1. Representative sample	2. Appropriate recruitment	3. Sample size	4. Subjects & setting described	5. Analysis coverage	6. Objective measure	7. Reliable measure	8. Statistical analysis	9. Confounding factors	10. Objective subgroupings
Sánchez-Morla 2009	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA
Sparding 2015	Y	Y	Y	Y	?	Y	?	Y	Y	Y
van der Werf-Eldering 2010	?	Y	N	Y	Y	Y	Y	Y	Y	Y

a. Studies contain overlapping samples.
b. Studies contain overlapping samples.
c. Studies contain overlapping samples.

Y	Yes
N	No
?	Unclear/unsure
NA	Not applicable

G – STROBE ratings for systematic review

	1a. Design in abstract/title	1b. Abstract	2. Background	3. Objectives	4. Design	5. Setting	6a. Participants	7. Variables	8. Data sources/measurement	9. Bias	10. Study size	11. Quantitative variables	12a. Describe statistics	12b. Subgroups/interactions	12c. Missing data	12d. Account for sampling	12e. Sensitivity analyses	13a. Participant numbers at each stage[a]	13b. Non-participation reasons[a]	13c. Flow diagram	14a. Sample characteristics	14b. Number missing data	15. Outcome data	16a. Estimates & precision	16b. Category boundaries	16c. Absolute risk	17. Other analyses	18. Key results	19. Limitations	20. Interpretation	21. Generalisability	22. Funder and role
Altshuler 2004	N	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	N	Y	Y	Y	N
Arslan 2014	N	N	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	NA	N	NA	NA	Y	Y	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	N
Barrera 2013	N	N	Y	Y	N	N	N	Y	N	N	N	Y	N	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	NA	N	N	Y	N	N
Cavanagh 2002	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	NA	N	N	Y	Y	Y	N	Y	N	Y	Y	NA	NA	Y	Y	Y	Y	Y	N
Cheung 2013	N	Y	Y	Y	N	N	N	Y	Y	N	N	N	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	Y	Y	Y
Daban 2012	N	Y	Y	Y	N	N	N	Y	Y	N	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	N	N	N
Doganavsargil-Baysal 2013	N	N	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	Y	N	NA	NA	Y	Y	N	Y	N	Y	N	NA	NA	Y	Y	N	Y	N	Y
Elshahawi 2011	N	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	N	N	Y	Y
Fakhry 2013	Y	N	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	Y
Ferrier 1999	N	N	Y	Y	N	N	Y	Y	Y	N	N	N	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	NA	Y	N	Y	N	N
Frangou 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	NA	N	Y	NA	Y	Y	N	Y	N	Y	Y	NA	N	NA	Y	N	Y	N	N
Goswami 2009	N	N	Y	Y	N	N	Y	N	N	Y	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	N	NA	Y	Y	Y	Y	N	N
Ibrahim[b] 2009	Y	N	Y	Y	Y	N	N	Y	Y	N	N	N	N	Y	N	NA	NA	N	N	N	Y	N	Y	N	N	N	N	Y	Y	N	N	N
Jamrozinski 2009	Y	N	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	N	Y	N	N
Juselius[c] 2009	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	N	N	N	Y	N	Y	Y	Y	NA	Y	N	Y	Y	Y	N
Kieseppä[c] 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	N	Y	N	Y	N	Y	Y	NA	NA	Y	N	Y	Y	Y	N
Lopera-Vásquez 2011	N	N	Y	Y	N	N	N	Y	N	N	N	Y	Y	NA	N	NA	NA	N	N	N	N	N	Y	N	NA	NA	NA	Y	N	N	N	N

	1a. Design in abstract/title	1b. Abstract	2. Background	3. Objectives	4. Design	5. Setting	6a. Participants	7. Variables	8. Data sources/measurement	9. Bias	10. Study size	11. Quantitative variables	12a. Describe statistics	12b. Subgroups/interactions	12c. Missing data	12d. Account for sampling	12e. Sensitivity analyses	13a. Participant numbers at each stage[a]	13b. Non-participation reasons[a]	13c. Flow diagram	14a. Sample characteristics	14b. Number missing data	15. Outcome data	16a. Estimates & precision	16b. Category boundaries	16c. Absolute risk	17. Other analyses	18. Key results	19. Limitations	20. Interpretation	21. Generalisability	22. Funder and role
López-Jaramillo 2010	N	N	Y	Y	Y	N	N	Y	N	Y	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	NA	Y	N	Y	N	N
Martino[d] 2008	N	Y	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	N	Y	Y	Y	N
Martino[d] 2011a	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	Y	Y	N	N
Martino[d] 2011b	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	Y	Y	N	N
Martino[d] 2011c	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	Y	Y	Y	N
Martino[d] 2014	N	Y	Y	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	N
Mur 2007	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	NA	NA	Y	Y	Y	Y	N	Y	N	NA	NA	Y	Y	N	Y	Y	N
Normala[b] 2010	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	Y	Y	Y	N
Osher 2011	N	N	Y	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	N	N	N	N
Pirkola[c] 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	NA	N	Y	Y	N	N	N	Y	N	Y	Y	Y	NA	NA	Y	Y	Y	Y	N
Sánchez-Morla 2009	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	Y	Y	N
Sparding 2015	N	N	Y	Y	N	N	Y	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	Y	NA	NA	Y	Y	Y	N	Y	Y
van der Werf-Eldering 2010	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	NA	Y	Y	Y	N	Y	N	Y	Y	Y	NA	Y	Y	N	Y	Y	Y

a. Rated Yes if information reported for bipolar disorder sample, at a minimum.
b. Studies contain overlapping samples.
c. Studies contain overlapping samples.
d. Studies contain overlapping samples.

Y	Yes
N	No
NA	Not applicable

H – Characteristics of the study samples in the systematic review

Table H.1 Characteristics of the BD-I samples

Author Year	Age (years), M (SD)		Male gender, *n* (%)		Education (years), M (SD)		BD onset age and/or duration (years), M (SD)	BD past illness episodes, M (SD)	Psychotropic medication in BD sample
	BD	HC	BD	HC	BD	HC			
Altshuler 2004	49.9 (13.9)	51.8 (12.6)	40 (100%)	22 (100%)	15.5 (2.4)	14.9 (2.0)	Onset 26.7 (9.2) Duration 24.9 (11.1)	Not stated	Lithium 63%, anticonvulsant 30%, antidepressant 10%, benzodiazepine 8%, typical antipsychotic 15%, anticholinergic 5%
Cavanagh 2002	43.6 (14.2)	42.2 (14.7)	10 (50%)	10 (50%)	Not stated	Not stated	Duration 16 (12.5)	Depressive 6 (6) Manic 6 (7)	SSRI 35%, typical antipsychotic 40%, lithium 40%, carbamazepine 25%
Cheung 2013	38.57 (10.70)	37.76 (10.27)	19 (36.5%)	19 (36.5%)	12.0 (2.94)	14.04 (3.11)	Onset 24.63 (7.6) Duration 13.3 (8.3)	Depressive 5.1 (5.2) Manic 5.2 (5.0)	Monotherapy (sodium valproate, lithium, antipsychotic, carbamazepine or lamotrigine) 48.1%, combination therapy 48.1% (inc. anticholinergic 15.3%); Two patients received short-acting low-dose benzodiazepine 12h before assessment
Fakhry 2013 S1: recent manic episode	32.27 (7.43)	31.47 (5.93)	17 (56.7%)	15 (50%)	Middle 3 (10%) Secondary 12 (40%) University 15 (50%)	Middle 3 (10%) Secondary 12 (40%) University 15 (50%)	Not stated	Not stated	Antipsychotic n=30, mood stabiliser other than lithium n=28
Fakhry 2013 S2: recent depressive episode	31.60 (6.43)	As above	13 (43.3%)	As above	Middle 3 (10%) Secondary 15 (50%) University 12 (40%)	As above	Not stated	Not stated	Antipsychotic n=28, mood stabiliser other than lithium n=26, antidepressant n=22
Ibrahim 2009[a] & Normala 2010[a]	Mdn 37.5 IQR 20.0	Mdn 27.0 IQR 15.0	19 (47.5%)	10 (25%)	Not stated	Not stated	Onset Mdn 21 IQR 13 Duration 10.95 (9.04)	All types 3.83 (3.07)	Mood stabiliser only 27.5%, mood stabiliser + antipsychotic 55%, mood stabiliser + antidepressant 2.5%, antipsychotic only 15%
Juselius 2009[b]	44.2 (1.6)	47.8 (0.6)	15 (57.7%)	55 (48.2%)	Level attained[c] 4.1 (0.5)	Level attained[c] 4.1 (0.2)	Duration 20.5 (5.8)	Manic 3.8 (0.45)	Not stated
Osher 2011	41.3 (13.2)	53.7 (18.9)	25 (49%)	181 (37%)	12.8 (2.0)	15.3 (3.2)	Onset 24.0 (8.0)	Not stated	Mood stabiliser monotherapy n=12, neuroleptic monotherapy n=8, polytherapy n=31

BD, bipolar disorder; BD-I, bipolar disorder type I; HC, healthy comparison; M, mean; Mdn, median; IQR, interquartile range; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.

a. Same sample.

b. All participants were twins. BD sample (n = 26) included 20 individuals whose co-twin did not have BD plus 3 pairs (6 individuals) concordant for BD. HC sample included n = 114 twins (46 pairs + 22 individuals) with no history of BD in the participant or their co-twin.

c. Level 4 is "vocation school or equivalent", after graduating high school.

Table H.2 Characteristics of the mixed BD samples

Author Year	% with BD-I	Age (years), M (SD)		Male gender, *n* (%)		Education (years), M (SD)		BD onset age and/or duration (years), M (SD)	BD past illness episodes, M (SD)	Psychotropic medication in BD sample
		BD	HC	BD	HC	BD	HC			
Barrera 2013	58%	48.21 (11.24)	46.04 (12.30)	12 (100%)	12 (100%)	12.33 (2.67)	12.50 (2.71)	Onset 25 (7.93)	Depressive 10.63 (13.5) Manic 5 (3.8)	Mood stabiliser 83.3%, antipsychotic 50%, anxiolytic 30%
Daban 2012	Not stated	41.12 (10.87)	46.53 (13.99)	23 (43.4%)	20 (33.3%)	14.21 (3.05)	12.49 (2.75)	Not stated	Not stated	Not stated
Martino 2014	51%	39.55 (10.83)	40.28 (12.03)	36 (36%)	12 (30%)	14.36 (2.36)	13.88 (2.77)	Onset 27.65 (9.49) Duration 11.18 (6.67)	Depressive 3.46 (2.01) Manic 3.18 (2.09)	Mood stabiliser 100%, antidepressant 38%, benzodiazepine 55%, antipsychotic 55%
Mur 2007	Not stated	42.6 (13.0)	42.2 (12.4)	22 (50%)	23 (50%)	10.5 (3.2)	12.5 (3.4)	Onset 25.6 (11.5) Duration 16.9 (11.67)	Manic 2.45 (2.5)	Lithium monotherapy n=20, lithium plus other n=24
Sánchez-Morla 2009	75%	43.5 (10.4)	43.8 (11.2)	30 (41.1%)	31 (46.3%)	12.5 (3.9)	14.1 (3.5)	Onset 26.2 (9.3) Duration 17.3 (10.5)	All types 13.3 (11.2) Manic 6.0 (5.6)	Lithium 39.8%, anticonvulsant 26.0%, lithium + anticonvulsant 30.1%, SSRI 30.1%, benzodiazepine 23.3%, typical antipsychotic 8.2%, atypical antipsychotic 26.0%
van der Werf-Eldering 2010	83%	Not stated[a]	40.8 (14.4)	Not stated[a]	27 (36%)	Not stated[a]	Level attained[b] 3.7 (1.1)	Not stated[a]	Not stated[a]	Not stated[a]

BD, bipolar disorder; HC, healthy comparison; M, mean; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. This information not reported in article for euthymic group separately.
b. Ranging from 1 = primary school to 6 = PhD or higher degree obtained.

I – Cognitive impairment in BD-I vs BD-II

Prevalence of cognitive impairment in BD-I versus BD-II samples, from Sparding et al. (2015)

Cognitive measure	Impairment prevalence[a,b]	
	BD-I[c]	BD-II[d]
D-KEFS TMT 2 number sequencing (executive/speed)	34%	19%
D-KEFS TMT 3 letter sequencing (executive/speed)	22%	14%
D-KEFS TMT 4 number-letter switching (executive)	48%	43%
D-KEFS Verbal Fluency 2 category (executive/language)	38%	39%
D-KEFS Verbal Fluency 3 category switching total (executive/language)	34%	27%
D-KEFS Verbal Fluency 3-v category switching accuracy (executive)	18%	11%
D-KEFS Color-Word 3 inhibition (executive)	29%	29%
D-KEFS Color-Word 4 inhibition/switching (executive)	27%	15%
D-KEFS Design Fluency 3 switching (executive)	32%	17%
D-KEFS Tower rule violations (executive)	34%	Not stated
WAIS-III Matrix Reasoning (abstract reasoning)	21%	30%
WAIS-III Similarities (abstract reasoning)	13%	11%
WAIS-III Letter-Number Sequencing (working memory)	22%	28%
WAIS-III Arithmetic (working memory)	37%	31%
WAIS-III Symbol Search (speed)	28%	19%
WAIS-III Digit-Symbol Coding (speed)	19%	11%
WAIS-III Digit Symbol Coding copy (speed)	35%	19%
WAIS-III Block Design (speed/visuospatial)	40%	44%
RCFT time to copy (speed/visuospatial)	23%	24%
RCFT 3-minute recall (visual memory)	23%	21%
RCFT recognition (visual memory)	22%	21%
WAIS-III Digit Symbol Coding free recall (visual memory)	25%	16%
WAIS-III Digit Symbol Coding pairing (visual memory)	26%	27%
WAIS-III Picture Completion (visuospatial)	14%	16%

BD-I, bipolar disorder type I; BD-II, bipolar disorder type II; D-KEFS, Delis-Kaplan Executive Function System; HC, healthy comparison; RCFT, Rey Complex Figure Test; SD, standard deviation; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition.
a. Impairment threshold is 1.25 SD from HC mean; HC impairment prevalence is 10.57% by definition. HC sample (n = 86) was frequency matched to BD samples based on age, gender and years of education; details not reported in article.
b. Standardised mean difference between BD and HC groups is not given in table because SD was not reported in article.
c. BD-I (n = 64) sample characteristics: age M = 38 (SD = 14); 48% male; education level M = 3.7 (SD = 1.1) (where 3 = 12 years, 4 = 13–15 years); age of onset M = 19 (SD = 9); past illness episodes (all types) M = 19 (SD = 26); medications: lithium 68%, antipsychotic 32%, antidepressant 31%, anticonvulsant 32%.
d. BD-II (n = 44) sample characteristics: age M = 35 (SD = 12); 45% male; education level M = 3.9 (SD = 1.2) (where 3 = 12 years, 4 = 13–15 years); age of onset M = 18 (SD = 11); past illness episodes (all types) M = 18 (SD = 18); medications: lithium 48%, antipsychotic 11%, antidepressant 41%, anticonvulsant 32%.

J – Factors associated with cognitive impairment in the systematic review

	Age ↑	Gender	Education ↓	Premorbid ability ↓	Ethnicity	Parental socioeconomic status ↓	Age at BD onset ↓	Duration of illness ↑	Residual affective symptoms ↑	Duration of euthymia ↓	Bipolar disorder subtype (BD-I)	No. of hospitalisations ↑	No. of episodes (all types) ↑	No. of episodes (manic) ↑	No. of episodes (depressed) ↑	History of psychotic symptoms	Any psychotropic medication	Mood stabiliser	Antipsychotic	Antidepressant	Benzodiazepine	Electroconvulsive therapy	Inflammatory biomarkers ↑	Alcohol/substance use disorder	Comorbid ADHD	Obstetric complications	Family history psychiatric illness
Altshuler 2004																											
Arslan 2014																											
Barrera 2013																											
Cavanagh 2002																											
Cheung 2013																											
Daban 2012		M< F																									
Doganavsargil-Baysal 2013																											
Elshahawi 2011		M< F																									
Ferrier 1999													a														
Frangou 2005																											
Goswami 2009																											
Ibrahim[b] 2009																											
Jamrozinski 2009																		c									

	Age ↑	Gender	Education ↓	Premorbid ability ↓	Ethnicity	Parental socioeconomic status ↓	Age at BD onset ↓	Duration of illness ↑	Residual affective symptoms ↑	Duration of euthymia ↓	Bipolar disorder subtype (BD-I)	No. of hospitalisations ↑	No. of episodes (all types) ↑	No. of episodes (manic) ↑	No. of episodes (depressed) ↑	History of psychotic symptoms	Any psychotropic medication	Mood stabiliser	Antipsychotic	Antidepressant	Benzodiazepine	Electroconvulsive therapy	Inflammatory biomarkers ↑	Alcohol/substance use disorder	Comorbid ADHD	Obstetric complications	Family history psychiatric illness
Juselius[d] 2009																c											
Kieseppä[d] 2005																											
Lopera-Vásquez 2011																											
López-Jaramillo 2010																											
Martino[e] 2008																											
Martino[e] 2011a																											
Martino[e] 2011b																											
Martino[e] 2011c																											
Martino[e] 2014																											
Mur 2007																											
Normala[b] 2010																											
Osher 2011		M> F																									
Pirkola[d] 2005		M> F																									
Sánchez-Morla 2009																											
Sparding 2015																											

ADHD, attention deficit hyperactivity disorder; BD, bipolar disorder; F, female; M, male.
a. Analysis based on number of recent episodes as well as time to recovery.
b. Studies contain overlapping samples.
c. Positive association with cognitive performance on some scores.
d. Studies contain overlapping samples.
e. Studies contain overlapping samples.

Significantly associated with impairment
Not significantly associated with impairment
Not analysed/measured

K – UK Biobank project approval

Annex II: Material Transfer Agreement with the Applicant for data and/or samples

biobank uk
Improving the health of future generations

Dear Dr Cullen,

UK Biobank is pleased to approve your Application Reference Number 11332 to use the UK Biobank Resource. Execution of this Material Transfer Agreement (MTA) and payment of the Access Charges are the final steps before access is granted. UK Biobank's approval of this Application is valid for 90 days, after which the Applicant Principal Investigator (PI) will need to re-apply for access. The content of UK Biobank's standard MTA, and the conditions contained within it, are non-negotiable.

Parties

This is an agreement between UK Biobank Limited on the one hand and the Applicant Institution (University of Glasgow) on the other hand. The Applicant PI is not a party to the MTA, however, UK Biobank requires that the Applicant PI acknowledges that the provisions of this MTA have been "read and understood" by the Applicant PI so that they are fully aware of their Institution's obligations to both UK Biobank and to UK Biobank's participants.

The Applicant Institution will be responsible for the conduct of any and all of the Applicant Researchers involved in this Research Project. The Applicant Institution shall not be responsible for the conduct of the Collaborating Institution(s), the Collaborating Investigator(s) or the Collaborating Researcher(s).

Structure of agreement

The MTA will become effective on receipt by UK Biobank of:

(i) A copy of this MTA Agreement (and copies of the relevant executed MTAs from all Collaborating Institution(s)) executed by an authorised signatory of the Applicant Institution and confirmed as "read and understood" by the Applicant PI; and
(ii) Cleared funds covering the Access Charges from the Applicant Institution.

UK Biobank will then promptly send a dated confirmatory email.

Provision of samples and/or data

Annex A summarises the data and/or samples that UK Biobank will make available to the Applicant in accordance with their approved Application Reference Number 11332. The timeframe and methodology by which the data and/or samples will be dispatched is also set out in Annex A.

Payment

The Access Charges which are payable are set out in Annex B. This also serves as an invoice on which VAT will be included (as appropriate). The derivation of these Access Charges is also set out in Annex B.

This payment should be submitted in cleared funds to Barclays Bank PLC, Account name: UK Biobank Limited, Account number: 33069427 and Sort code: 20-24-09.

Standard terms and schedules

This Agreement incorporates the attached terms and conditions (including any documents and/or materials that are referred to in them), the Annexes and where applicable the contents of the Preliminary and Main Application Forms with Reference Number 11332.

Yours faithfully

For and on behalf of UK Biobank/Effective Date
(Jonathan Sellors / Company Solicitor)

30th March 2015

Accepted and agreed

For and on behalf of Applicant Institution
(Please sign and print your name and position)

Paul Ellis
Senior Contracts Manager
Research Support Office
University of Glasgow

Read and understood by the Applicant Principal Investigator
(Please sign and print your name and position)

DR BREDA CULLEN, RESEARCH FELLOW

16th February 2015

Annex II: Material Transfer Agreement with the Collaborator for data and/or samples

biobank uk
Improving the health of future generations

Dear Professor Deary,

UK Biobank is pleased to approve the Application Reference Number 11332 to use the UK Biobank Resource. Execution of this Material Transfer Agreement (MTA) and payment of the Access Charges are the final steps before access is granted. UK Biobank's approval of this Application is valid for 90 days, after which the Applicant Principal Investigator (PI) (Dr Cullen) will need to re-apply for access. The content of UK Biobank's standard MTA, and the conditions contained within it, are non-negotiable.

Parties

This is an agreement between UK Biobank Limited on the one hand and the Collaborating Institution (University of Edinburgh) on the other hand. The Collaborating Investigator is not a party to the MTA, however, UK Biobank requires that the Collaborating Investigator acknowledges that the provisions of this MTA have been "read and understood" by the Collaborating Investigator so that they are fully aware of their Institute's obligations to both UK Biobank and to UK Biobank's participants.

The Collaborating Institution shall be responsible for the conduct of any and all of the Collaborating Researchers involved in this Research Project. The Collaborating Institution shall not be responsible for the obligations and responsible for the conduct of either (a) the Applicant Institution, the Applicant Principal Investigator or the Applicant Researchers or (b) third party Collaborating Institution(s), third party Collaborating Investigator(s) or third party Collaborating Researcher(s).

In the event that the related MTA with the Applicant is terminated (for whatever reason) then UK Biobank shall have the right to terminate this MTA forthwith on the provision of written notice to the Collaborator.

Structure of agreement

The MTA will become effective on receipt by UK Biobank of:

(i) A copy of this MTA Agreement (and a copy of the relevant executed MTA from the Applicant Institution and copy(s) of the relevant executed MTA(s) from third party Collaborating Institutions) executed by an authorised signatory of the Collaborating Institution and confirmed at "read and understood" by the Collaborating Investigator; and

(ii) Cleared funds covering the Access Charges from the Applicant Institution.

UK Biobank will then promptly send a dated confirmatory email.

Provision of samples and/or data

Annex A summarises the data and/or samples that UK Biobank will make available to the Collaborator in accordance with the approved Application Reference Number 11332. The timeframe and methodology by which the data and/or samples will be dispatched is also set out in Annex A.

Standard terms and schedules

This Agreement incorporates the attached terms and conditions (including any documents and/or materials that are referred to in them), the Annexes and where applicable the contents of the Preliminary and Main Application Forms with Reference Number 11332.

Yours faithfully

For and on behalf of UK Biobank / Effective Date
(Jonathan Sellors / Company Solicitor)

30th March 2015

Accepted and agreed

For and on behalf of Collaborating Institution
(Please sign and print your name and position)

CLARE WHITTAKER
CONTRACTS MANAGER

Read and Understood by the Collaborating Investigator
(Please sign and print your name and position)

IAN J. DEARY
PROFESSOR OF DIFFERENTIAL PSYCHOLOGY

16th February 2015

L – NHS Research Ethics Committee approval

NHS
Health Research Authority

North West - Haydock Research Ethics Committee
3rd Floor - Barlow House
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0207 104 8012

13 May 2016

Dr Tim Peakman
UK Biobank Limited
1-4 Spectrum Way
Adswood
Stockport
Cheshire
SK3 0SA

Dear Dr Peakman

Title of the Research Tissue Bank:	**UK Biobank: a large scale prospective epidemiological resource**
REC reference:	**16/NW/0274**
Designated Individual:	**Dr Tim Peakman**
IRAS project ID:	**200778**

The Research Ethics Committee reviewed the above application at the meeting held on 10 May 2016. Thank you for attending with Mr Jonathan Sellors and Ms Nicola Doherty to discuss the application.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Ms Rachel Katzenellenbogen, nrescommittee.northwest-haydock@nhs.net.
Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Favourable opinion

The members of the Committee present gave a favourable ethical opinion of the above research tissue bank on the basis described in the application form and supporting documentation, subject to the conditions specified below.

The Committee has also confirmed that the favourable ethical opinion applies to all research

projects conducted in the UK using tissue or data supplied by the tissue bank, provided that the release of the tissue or data complies with the attached conditions. It will not be necessary for these researchers to make project-based applications for ethical approval. They will be deemed to have ethical approval from this committee. You should provide the researcher with a copy of this letter as confirmation of this. The Committee should be notified of all projects receiving tissue and data from the tissue bank by means of an annual report.

This application was for the renewal of a Research Tissue Bank application. The previous REC Reference number for this application was 11/NW/0382.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Research governance

Under the Research Governance Framework (RGF), there is no requirement for NHS research permission for the establishment of research tissue banks in the NHS. Applications to NHS R&D offices through IRAS are not required as all NHS organisations are expected to have included management review in the process of establishing the research tissue bank.

Research permission is also not required by collaborators at tissue collection centres (TCCs) who provide tissue or data under the terms of a supply agreement between the organisation and the research tissue bank. TCCs are not research sites for the purposes of the RGF.

Research tissue bank managers are advised to provide R&D offices at all TCCs with a copy of the REC application for information, together with a copy of the favourable opinion letter when available. All TCCs should be listed in Part C of the REC application.

NHS researchers undertaking specific research projects using tissue or data supplied by the research tissue bank must apply for permission to R&D offices at all organisations where the research is conducted, whether or not the research tissue bank has ethical approval.

Site-specific assessment (SSA) is not a requirement for ethical review of research tissue banks.

Summary of discussion at the meeting

Social or scientific value; scientific design and conduct of the study

The Committee were pleased to see that UK Biobank was constantly re-evaluating itself with regards to new technology and data collection. This meant that new tests were undertaken and new data and tissue collected allowing UK Biobank to grow and develop as a resource.

The Committee were very pleased that this resource was open access and also that researchers had to register to use it. The Committee noted that while UK Biobank owned the resource they had no preferential access. The Committee very happy to note that all research results had to be sent back to UK Biobank as part of a transparency agenda.

The Committee noted that the data was being used by a broad range of researchers and asked how use would be maximised in the future.

> You explained that originally UK Biobank had been designed to be used in case control studies. However, you had been able to demonstrate that centralised generation of large datasets had advantages of cost, standardisation and a lack of gaps. This meant that it was being used in more than just case control studies.
>
> You said that genotyping was being done on all participants and that they were currently measuring 34 biomarkers with the data available to all. You said that you were currently working up a proposal to measure 40 markers of infectious disease and were also looking at developing strategies to look at proteins and metabolites. You said it was important to maximise the tissue so that, for example, you wouldn't use tissue simply to measure glucose, but if you could run tests that delivered a lot of data, including glucose, then the data would be gathered in a good way.
>
> You said that the data was linked to various registers, including deaths, cancer and hospital visit. 30% of English participants had primary care information and this was a lot higher for Welsh and Scottish participants. This meant that you would be able to create a plausible calendar as to when the data would be mature for more common conditions and then you would put out a call for researchers.

The Committee noted that one of the criteria for accessing the biobank was that the research be "in the public interest". The Committee asked if any applications had been turned down because they had not been in the public interest.

> You said that no applications had been turned down because they were not in the public interest. In fact, only 2 or 3 requests for samples had been turned down and that was because they had either requested too much or actually did not need to turn to a biobank to do their research.

The Committee agreed that it had been an exemplary submission and had led to an interesting and informative discussion. The Committee looked forward to the publication regarding imaging and the reporting of findings and hoped the researchers would advise them of when it was published and how it could be accessed.

Care and protection of research participants; respect for potential and enrolled participants' welfare and dignity

The Committee agreed that the systems in place to avoid identifying participants were robust. Always growing and considering and developing.

The Committee noted that UK Biobank was regularly in touch with participants via newsletter and held an Annual General Meeting. The Committee agreed this was very important if participants were to stay motivated and interested as without this no new data or tissue could be added.

The Committee noted that the imaging Participant Information Sheet and consent form said that GPs would be contacted if anything clinically significant was discovered. The Committee noted that UK Biobank had had a policy of not feeding back findings and wondered if this policy had now changed. The Committee also agreed that they needed to know more about how significant

clinical findings were determined. For example, carotid arteries narrowed as people aged, so would all narrowing be reported or just ones with a certain percentage of narrowing.

The Committee asked what the current position was regarding feeding back clinically significant findings.

> You said that the position had not changed, although it was reconsidered on a regular basis. When participants came for their baseline visit in 2007-2010 if something was spotted during the visit, then it was fed back. However, assay or other research findings were not fed back.
>
> With regard to imaging, which could lead to acute findings such as cancer, you explained that you had spent 5-6 years working out the best protocol for that. The end result was that if the radiographer observed something that concerned them it was flagged and a radiologist would assess it. If the radiologist determined that it was significant then it was reported to the GP.
>
> You explained that during the imaging pilot you had run two protocols, the one that is in current use, and a second one that involved a radiologist screening all of the images. After follow up it became clear that this was hugely problematic, not because of cost or expediency, but because it had led to 200 false positives. At the extreme end there had been a lung section and a removal of ovaries for people with false positives. Scaling this up to 100,000 people meant there could be 20,000 false positives.
>
> You said that you had spoken to participants and to imaging projects and it had been agreed that while the radiographers might miss things, the best protocol was to have radiologists only look at images flagged by radiographers. You also said that you would be publishing the results of this research shortly.
>
> You said that, in short, the feedback policy was that anything of clinical significance discovered during data acquisition would be fedback but any other findings would not be.

The Committee agreed that this was acceptable, especially as it was all made very clear to participants in information sheets.

The Committee asked why radiologists were diagnosing so many false positives.

> You said that the images were research scans which, despite what many participants had thought, were not more detailed than ones taken for clinical purposes. Additionally, the radiologists did not have any of the other information they would have in normal diagnosis.

The Committee agreed that the level of commitment required from participants was high and the Committee agreed they would like to know how many participants had withdrawn and how many had simply been lost to contact. However, the Committee was impressed with the way UK Biobank kept participants informed of new developments and asked how many participants had been lost to contact or withdrawal.

> You said that just over 1,000 participants had withdrawn with about 600 of them having requested all tissue and data be removed from the resource. The Committee said that

while annual communications always sparked some withdrawals, the benefits of the communication far outweighed that problem.

You said that most communication was by email, including web based questionnaires. However, it was easier to keep in touch with people by post because if they moved you could usually find their new address. Also it was impossible to know how many emails were opened and read, so no one knew who actually read the newsletter.

You explained that response rates to questionnaires had actually gone up over time and that there had been a 50% response rate to the request for participants willing to wear an accelerometer. In fact, you had managed to recruit 100,000 participants to do that.

You said that you were now also starting to use mobile technology to contact participants.

Other ethical issues were raised and resolved in preliminary discussion before your attendance at the meeting.

Duration of ethical opinion

The favourable opinion is given for a period of five years from the date of this letter and provided that you comply with the standard conditions of ethical approval for Research Tissue Banks set out in the attached document. You are advised to study the conditions carefully. The opinion may be renewed for a further period of up to five years on receipt of a fresh application. It is suggested that the fresh application is made 3-6 months before the 5 years expires, to ensure continuous approval for the research tissue bank.

Research Tissue Bank Renewals

The Research Tissue Bank has been renewed for a further five years from the end of the previous five year period. The previous five year period ran from 17 June 2011 to 17 June 2016. This Research Tissue Bank may be renewed for further periods of five years at a time by following the process described in the above paragraph.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Human Tissue Authority licence [HTA Licence 12002 & 12624]		26 July 2010
IRAS Checklist XML [Checklist_27042016]		27 April 2016
Other [Table 1: Comparison of the sample collection for the baseline assessment and imaging pilot]	1.0	02 March 2016
Other [Table 2: Progress with key cohort-wide linkages Q1-2 2016]	1.0	02 March 2016
Other [UK Biobank Ethics & Governance Framework]	3.0	01 October 2007
Other [Figure 1: Submitted Access Applications by areas of interest]	1.0	02 March 2016
Other [Table 3: Biochemistry assays being performed in all 500,000 participants]	1.0	02 March 2016
Other [Participant Withdrawal Form]	1.1	10 February 2012

Other [UK Biobank Newsletter June 2015]	1.0	22 June 2015
Other [Data Dictionary Showcase Sept15]	Sept 2015	24 March 2016
Other [Curriculum vitae - Timothy Peakman]	March 2016	24 March 2016
Other [RTB Report March 2016]	1.0	24 March 2016
Other [Appendix: Occupational Questionnaire]	1.0	27 August 2014
Other [Appendix: Occupational Questionnaire Invitation Text]	1.0	27 August 2014
Other [Appendix: Occupational Questionnaire Reminder Invitation Text]	1.0	27 August 2014
Other [Revised Imaging Invitation Email]	1.0	06 October 2014
Other [Imaging Reminder Text & SMS]	1.0	18 November 2014
Other [Feedback in the UK Biobank Imaging pilot study]	Jan 2014	29 January 2014
Other [Invitation letter for deliberative group interviews]	1.0	06 October 2014
Other [Imaging 2nd Invite email HTML]	0.1	01 January 2016
Other [Imaging 2nd Invite email PLAIN]	0.2	01 January 2016
Other [Imaging Participant pre-screening questionnaire]	1.3	27 October 2015
Other [Imaging Exit Survey]	0.1	01 January 2016
Other [Invite email reminder 6-month questionnaire HTML]	0.1	01 October 2015
Other [Invite email reminder 6-month questionnaire PLAIN]	0.1	01 October 2015
Other [Invite email reminder 6-week questionnaire HTML]	0.1	01 October 2015
Other [Invite email reminder 6-week questionnaire PLAIN]	0.1	01 October 2015
Other [Invite email reminder understanding consent questionnaire HTM]	0.1	01 October 2015
Other [Invite email reminder understanding consent questionnaire PLAIN]	0.1	01 October 2015
Other [Appendix 1: Mental Health Questionnaire]	1.2	23 March 2016
Other [Appendix 2: Rationale and tools used in Mental Health Questionnaire]	1.1	04 March 2016
Other [Appendix 3: Invitation email Mental Health Questionnaire]	1.2	11 March 2016
Other [Appendix 4: Reminder email Mental Health questionnaire]	1.2	11 March 2016
Other [Appendix 5: Reminder partial responder email Mental Health questionnaire]	1.1	11 March 2016
Other [Appendix 6: Last chance email Mental Health questionnaire]	1.0	11 March 2016
Other [Repeat Assessment email invitation]	1.0	09 August 2012
Other [Repeat Assessment invite letter]	1.0	26 March 2012
Other [Repeat Assessment confirmation letter]	1.0	11 July 2012
Other [Confirmation of imaging appointment letter]	1.0	08 April 2016
Other [Activity Monitor Information Letter]	26/03/2012	26 March 2012
Other [Activity Monitor Invitation Letter]	26/03/2012	26 March 2012
Other [Activity Monitor Return Reminder]	26/03/2012	26 March 2012
Other [UK Biobank Assessment form]	20061124	24 November 2006
Other [Diet Questionnaire]	1.0	11 April 2016
Other [UK Biobank Participant Invite letter]	1.0	11 April 2016
Other [Touch-screen questionnaire]	1.0	11 April 2016

Other [Touch-screen questionnaire addendum]	1.0	11 April 2016
Other [Cognitive Function tests]	1.0	26 March 2013
Other [Cognitive Function Web Questionnaire email invitation]	1.0	26 March 2013
Other [Cognitive Function Web Questionnaire email reminder]	1.0	26 March 2013
Other [Cognitive Function Web Questionnaire email reminder partial responder]	1.0	26 March 2013
Other [UK Biobank Protocol]	21/03/2007	21 March 2007
Other [UK Biobank Protocol addendum 1]	09/04/2009	09 April 2009
Other [UK Biobank Protocol addendum 2]	02/07/2009	02 July 2009
Other [Text Message to request email address]	1.0	20 April 2016
Other [UK Biobank TIME study invitation]	2.2	15 April 2016
Other [Imaging Questionnaire to assess participant understanding of consent]	January 2014	01 January 2014
Other [Imaging Participant Questionnaire sent at 6 weeks to assess IF]	January 2014	01 January 2014
Other [Imaging Participant Questionnaire sent at 6 months to assess impact of IF]	January 2014	01 January 2014
Other [Imaging Questionnaire sent to participants who did not receive IF feedback]	January 2014	01 January 2014
Other [Imaging Letters notifying participant and participant's GP of potentially serious incidental finding]	1.0	01 January 2014
Other [Imaging GP questionnaire sent at 6 months to assess the later impact of feedback of IF]	1.0	01 April 2015
Participant consent form [UK Biobank Consent form]	20061124	24 November 2006
Participant consent form [Consent Form for the imaging assessment: UK Biobank]	Jan 2014	29 January 2014
Participant information sheet (PIS) [Participant Information Leaflet]	21/04/2010	21 April 2010
Participant information sheet (PIS) [Biobank Imaging Information Leaflet]	Dec 2015	01 December 2015
Participant information sheet (PIS) [Repeat Assessment Participant Information Leaflet]	26/03/2012	26 March 2012
Participant information sheet (PIS) [Further Information Leaflet]	001	08 April 2016
Participant information sheet (PIS) [Biobank Imaging Information Leaflet including ECG monitoring]	2.0	26 November 2014
Protocol for management of the tissue bank [UK Biobank Access Procedures]	1.0	01 November 2011
REC Application Form [RTB_Form_24032016]		24 March 2016

Licence from the Human Tissue Authority

Thank you for providing a copy of the above licence.

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet. Dr Tim Sprosen helped set up UK Biobank and was a member of the Scientific Steering Committee. It was agreed that Dr Sprosen would leave the room during the discussion and take no part in the discussion or decision making. Dr Valerie Siddall, Alternate Vice-Chair, would chair

that portion of the meeting.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached standard conditions give detailed guidance on reporting requirements for research tissue banks with a favourable opinion, including:

- Notifying substantial amendments
- Submitting Annual Progress reports

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at
http://www.hra.nhs.uk/hra-training/

16/NW/0274	**Please quote this number on all correspondence**

Yours sincerely

Dr Tim S Sprosen
Chair

E-mail: nrescommittee.northwest-haydock@nhs.net

M – University Research Ethics Committee approval

University of Glasgow College of Medical, Veterinary & Life Sciences

12th October 2015

Dear Dr Cullen

MVLS College Ethics Committee

Project Title: Cognitive outcomes in people with behavioural and brain disorders within UK Biobank
Project No: 200150023

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Project end date: 31 August 2017
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or with the groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.

Yours sincerely

Dorothy McKeegan

Dr Dorothy McKeegan
College Ethics Officer

Dr Dorothy McKeegan

Senior Lecturer

R303 Level 3
Institute of Biodiversity Animal Health and Comparative Medicine
Jarrett Building
Glasgow G61 1QH Tel: 0141 330 5712
E-mail: Dorothy.McKeegan@glasgow.ac.uk

University of Glasgow College of Medical, Veterinary & Life Sciences

1st August 2017

Dear Dr Cullen

MVLS College Ethics Committee

Project Title: Cognitive outcomes in people with behavioural and brain disorders within UK Biobank
Project No: 200150023

The College Ethics Committee has reviewed your application of August 2017 for a minor amendment to the above project and has granted this in full. Specifically, the final project end date is amended to 31 December 2017.

The research should be carried out only on the sites, and/or with the groups defined in the application.

- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.

Yours sincerely

Dorothy McKeegan

Dr Dorothy McKeegan
College Ethics Officer

Dr Dorothy McKeegan

Senior Lecturer

R303 Level 3
Institute of Biodiversity Animal Health and Comparative Medicine
Jarrett Building
Glasgow G61 1QH Tel: 0141 330 5712
E-mail: Dorothy.McKeegan@glasgow.ac.uk

N – NHS R&D acknowledgement

Breda Cullen

From: Reid, Lorraine <Lorraine.Reid2@ggc.scot.nhs.uk>
Sent: 28 January 2015 15:28
To: Breda Cullen
Subject: R&D Ref: GN14NE132 - R&D Acknowledgement of a Tissue Bank

Dear Dr Cullen

R&D Ref: GN14NE132
Ethics Ref: 11/NW/0382
Investigator: Dr Breda Cullen
Title: Cognitive outcomes in psychiatric and neurological populations within UK Biobank

I am please to inform you that NHS Greater Glasgow & Clyde R&D Department have reviewed the Tissue Bank documentation which included:

- Protocol ID: UKBB-PROT-09-06
- Protocol Addendum dated 09/04/09
- Ethics Favourable Opinion letter

Although R&D approval is not required in this case, we would like to acknowledge receipt of the above paperwork and wish you every success.

Please inform the R&D office of any amendments.

Kind regards
Lorraine

Lorraine Reid
Senior Research Administrator
R&D Proportionate Team
R&D Management Office
1st Floor, Tennent Institute
Western Infirmary
Glasgow
G11 6NT

Tel: 0141 211 1743
Email: Lorraine.Reid2@ggc.scot.nhs.uk
Generic: RandD.PRTeam@ggc.scot.nhs.uk

Live in Scotland? Join SHARE and help us improve Scottish Health: http://www.registerforshare.org/
SHARE is an important initiative to establish a register of people interested in participating in health research across Scotland, and it is very important that we advertise it as much as possible. If you could also access the website and register your details it would be very helpful.

"Please note that from the **27th May 2013** R&D will be operating an electronic record system. Please submit your study documents via e-mail or IRAS from this date."

: www.nhsggc.org.uk/r&d

****

O – Touchscreen mood questionnaire classifications in UK Biobank

Criteria for lifetime experience of features of bipolar disorder

Bipolar disorder, type I
Ever ‘manic or hyper’ for 2 days OR ever ‘irritable/argumentative’ for 2 days; plus at least 3 features from ‘more active’, ‘more talkative’, ‘needed less sleep’ and ‘more creative/more ideas’; plus duration of a week or more; plus ‘needed treatment or caused problems at work’.

Bipolar disorder, type II
Ever ‘manic or hyper’ for 2 days OR ever ‘irritable/argumentative’ for 2 days; plus at least 3 features from ‘more active’, ‘more talkative’, ‘needed less sleep’ and ‘more creative/more ideas’; plus duration of a week or more.

Criteria for lifetime experience of features of major depression

Probable recurrent major depression (severe)
Ever depressed/down for a whole week; plus at least two weeks duration; plus at least two episodes; plus ever seen a psychiatrist for ‘nerves, anxiety, tension, depression’ OR ever anhedonic (unenthusiasm/uninterest) for a whole week; plus at least two weeks duration; plus at least two episodes; plus ever seen a psychiatrist for ‘nerves, anxiety, tension, depression’.

Probable recurrent major depression (mild-moderate)
Ever depressed/down for a whole week; plus at least two weeks duration; plus at least two episodes; plus ever seen a GP (but not a psychiatrist) for ‘nerves, anxiety, tension, depression’ OR ever anhedonic (unenthusiasm/uninterest) for a whole week; plus at least two weeks duration; plus at least two episodes; plus ever seen a GP (but not a psychiatrist) for ‘nerves, anxiety, tension, depression’.

Single probable episode of major depression
Ever depressed/down for a whole week; plus at least two weeks duration; plus only one episode; plus ever seen a GP or a psychiatrist for ‘nerves, anxiety, tension, depression’ OR ever anhedonic (unenthusiasm/uninterest) for a whole week; plus at least two weeks duration; plus only one episode; plus ever seen a GP or a psychiatrist for ‘nerves, anxiety, tension, depression’.

P – Psychiatric and neurological diagnoses in linked hospital records

ICD-10 code	ICD-10 description
A8x.x	Viral infections of the central nervous system
B22.0	HIV disease resulting in encephalopathy
B90.0	Sequelae of central nervous system tuberculosis
B94.1	Sequelae of viral encephalitis
C70.0	Malignant neoplasm of meninges (cerebral)
C71.x	Malignant neoplasm of brain
C72.8	Overlapping lesion of brain and other parts of central nervous system
C75.1	Malignant neoplasm of pituitary gland
C75.3	Malignant neoplasm of pineal gland
C79.3	Secondary malignant neoplasm of brain and cerebral meninges
D32.0	Benign neoplasm of meninges (cerebral)
D33.0	Benign neoplasm of brain, supratentorial
D33.1	Benign neoplasm of brain, infratentorial
D33.2	Benign neoplasm of brain, unspecified
D35.2	Benign neoplasm of pituitary gland
D35.4	Benign neoplasm of pineal gland
D42.0	Neoplasm of uncertain or unknown behaviour of meninges (cerebral)
D43.0	Neoplasm of uncertain or unknown behaviour of brain, supratentorial
D43.1	Neoplasm of uncertain or unknown behaviour of brain, infratentorial
D43.2	Neoplasm of uncertain or unknown behaviour of brain, unspecified
D44.3	Neoplasm of uncertain or unknown behaviour of pituitary gland
D44.5	Neoplasm of uncertain or unknown behaviour of pineal gland
Fxx.x	Mental and behavioural disorders
G0x.x	Inflammatory diseases of the central nervous system
G10	Huntington disease
G11.x	Hereditary ataxia
G12.2	Motor neuron disease
G13.1	Other systemic atrophy primarily affecting central nervous system in neoplastic disease (Paraneoplastic limbic encephalopathy)
G2x.x	Extrapyramidal and movement disorders
G30.x	Alzheimer disease
G31.x	Other degenerative diseases of nervous system, not elsewhere classified
G32.8	Other specified degenerative disorders of nervous system in diseases classified elsewhere
G35	Multiple sclerosis
G36.x	Other acute disseminated demyelination
G37.x	Other demyelinating diseases of central nervous system
G4x.x	Episodic and paroxysmal disorders
G8x.x	Cerebral palsy and other paralytic syndromes
G90.3	Multi-system degeneration
G91.x	Hydrocephalus
G92	Toxic encephalopathy

ICD-10 code	ICD-10 description
G93.x	Other disorders of brain
G94.x	Other disorders of brain in diseases classified elsewhere
G96.x	Other disorders of central nervous system
G97.x	Postprocedural disorders of nervous system, not elsewhere classified
G98	Other disorders of nervous system, not elsewhere classified
H47.6	Disorders of visual cortex
I6x.x	Cerebrovascular diseases
Q0x.x	Congenital malformations of the nervous system
Q28.2	Arteriovenous malformation of cerebral vessels
Q28.3	Other malformations of cerebral vessels
Q9x.x	Chromosomal abnormalities, not elsewhere classified
R41.x	Other symptoms and signs involving cognitive functions and awareness
R90.0	Intracranial space-occupying lesion
R94.0	Abnormal results of function studies of central nervous system
S02.0x	Fracture of vault of skull
S02.1x	Fracture of base of skull
S06.x	Intracranial injury
S07.1	Crushing injury of skull
S09.7	Multiple injuries of head
T02.0x	Fractures involving head with neck
T04.0	Crushing injuries involving head with neck
T06.0	Injuries of brain and cranial nerves with injuries of nerves and spinal cord at neck level
T40.x	Poisoning by narcotics and psychodysleptics [hallucinogens]
T42.x	Poisoning by antiepileptic, sedative-hypnotic and antiparkinsonism drugs
T43.x	Poisoning by psychotropic drugs, not elsewhere classified
T51.x	Toxic effect of alcohol
T58	Toxic effect of carbon monoxide
T90.2	Sequelae of fracture of skull and facial bones
T90.5	Sequelae of intracranial injury

Q – Self-reported psychiatric and neurological diagnoses in UK Biobank

UK Biobank data field	Diagnosis
6150 (touchscreen - vascular)	Stroke
20001 (interview - cancer)	Brain cancer/primary malignant tumour
“	Meningeal cancer/malignant meningioma
20002 (interview - non-cancer)	Alcohol dependency
“	Anorexia/bulimia/other eating disorder
“	Anxiety/panic attacks
“	Benign/essential tremor
“	Brain haemorrhage
“	Brain/intracranial abscess
“	Cerebral aneurysm
“	Cerebral palsy
“	Chronic/degenerative neurological problem
“	Deliberate self-harm/suicide attempt
“	Dementia/Alzheimer's/cognitive impairment
“	Encephalitis
“	Epilepsy
“	Fracture skull/head
“	Head injury
“	Headaches (not migraine)
“	Infection of nervous system
“	Insomnia
“	Ischaemic stroke
“	Meningioma benign
“	Meningitis
“	Migraine
“	Motor neurone disease
“	Nervous breakdown
“	Neurological injury/trauma
“	Neuroma benign
“	Obsessive compulsive disorder (OCD)
“	Opioid dependency
“	Other demyelinating condition
“	Other neurological problem
“	Other substance abuse/dependency
“	Post-traumatic stress disorder
“	Psychological/psychiatric problem
“	Spina bifida
“	Stress
“	Stroke
“	Subarachnoid haemorrhage
“	Subdural haematoma
“	Transient ischaemic attack

R – Self-reported medications in UK Biobank

Table R.1 Psychotropic medications from UK Biobank field 20003

Mood stabilisers	Selective serotonin reuptake inhibitors	Other antidepressants	Traditional antipsychotics	Second generation antipsychotics	Sedatives & hypnotics
lithium product	paroxetine	mirtazapine	chlorpromazine	quetiapine	diazepam
Priadel (lithium)	Seroxat (paroxetine)	Zispin (mirtazapine)	cpz - chlorpromazine	Seroquel (quetiapine)	diazepam product
Camcolit (lithium)	fluoxetine	duloxetine	Largactil (chlorpromazine)	risperidone	Valium tablet (diazepam)
sodium valproate	Prozac (fluoxetine)	Cymbalta (duloxetine)	haloperidol	Risperdal (risperidone)	Valium syrup (diazepam)
Epilim (sodium valproate)	citalopram	Yentreve (duloxetine)	Haldol (haloperidol)	olanzapine	Valium supp (diazepam)
Depakote (semisodium valproate)	Cipramil (citalopram)	venlafaxine	Serenace (haloperidol)	Zyprexa (olanzapine)	temazepam
valproic acid	escitalopram	Efexor (venlafaxine)	fluphenazine decanoate	aripiprazole	Normison (temazepam)
carbamazepine product	Cipralex (escitalopram)	amitriptyline	fluphenazine	Abilify (aripiprazole)	Euhypnos (temazepam)
carbamazepine	sertraline	Elavil (amitriptyline)	Modecate (fluphenazine)	amisulpride	zopiclone
Tegretol (carbamazepine)	Lustral (sertraline)	Tryptizol (amitriptyline)	Moditen tablet (fluphenazine)	Solian (amisulpride)	Zimovane (zopiclone)
Teril (carbamazepine)	fluvoxamine	Lentizol (amitriptyline)	Moditen enanthate (fluphenazine)	clozapine	zaleplon
Teril retard (carbamazepine)		amitriptyline+perphenazine	flupentixol	Clozaril (clozapine)	Sonata (zaleplon)
Timonil retard (carbamazepine)		Triptafen (amitriptyline+perphenazine)	Flupenthixol (flupentixol)		zolpidem
Epimaz (carbamazepine)		amitriptyline+chlordiazepoxide	Depixol (flupentixol)		Stilnoct (zolpidem)
lamotrigine		Limbitrol 10 (amitriptyline+chlordiazepoxide)	Fluanxol (flupentixol)		nitrazepam
Lamictal (lamotrigine)		Limbitrol-5 (amitriptyline+chlordiazepoxide)	zuclopenthixol		Mogadon (nitrazepam)
		phenelzine	Clopixol (zuclopenthixol)		Nitrados (nitrazepam)

Mood stabilisers	Selective serotonin reuptake inhibitors	Other antidepressants	Traditional antipsychotics	Second generation antipsychotics	Sedatives & hypnotics
		maoi - phenelzine	loxapine		Remnos (nitrazepam)
		Nardil (phenelzine)	Loxapac (loxapine)		Somnite (nitrazepam)
		moclobemide	droperidol		Noctesed (nitrazepam)
		Manerix (moclobemide)	Droleptan (droperidol)		Surem (nitrazepam)
		imipramine	trifluoperazine		Unisomnia (nitrazepam)
		Tofranil (imipramine)	Stelazine (trifluoperazine)		flunitrazepam
		trimipramine	thioridazine		Rohypnol (flunitrazepam)
		Surmontil (trimipramine)	Melleril (thioridazine)		triazolam
		dothiepin			Halcion (triazolam)
		dosulepin			
		Prothiaden (dosulepin)			
		Thaden (dosulepin)			
		clomipramine			
		Anafranil (clomipramine)			
		lofepramine			
		Gamanil (lofepramine)			
		Lomont (lofepramine)			
		mianserin			
		Bolvidon (mianserin)			
		Norval (mianserin)			

Table R.2 Multiple sclerosis and Parkinson's disease medications from UK Biobank field 20003

Multiple sclerosis disease modifying medications	**Parkinson's disease medications**
interferon beta	apomorphine
interferon beta-1a	Uprima (apomorphine)
Rebif 12million iu/0.5ml prefilled syringe (interferon beta-1a)	bromocriptine
Rebif 6million iu/0.5ml prefilled syringe (interferon beta-1a)	Parlodel (bromocriptine)
Rebif 22micrograms/0.5ml prefilled syringe (interferon beta-1a)	cabergoline
Rebif 44micrograms/0.5ml prefilled syringe (interferon beta-1a)	Cabaser (cabergoline)
interferon beta-1b	pergolide
interferon beta-1b product	pramipexole
Avonex 6million iu injection (pdr for recon)+solvent (interferon beta-1b)	Mirapexin (pramipexole)
Avonex 6million iu/0.5ml prefilled syringe (interferon beta-1b)	ropinirole
Avonex 30micrograms/0.5ml prefilled syringe (interferon beta-1b)	Requip (ropinirole)
Betaferon 9.6 million iu injection (pdr for recon)+diluent (interferon beta-1b)	levodopa
Betaferon 300micrograms injection (pdr for recon)+diluent (interferon beta-1b)	l-dopa - levodopa
glatiramer	levodopa+careldopa+entacapone
Copaxone 20mg injection (pdr for recon) (glatiramer acetate)	Stalevo (carbidopa/levodopa/entacapone)
	co-careldopa
	Sinemet (co-careldopa)
	half-sinemet (co-careldopa)
	Sinemet-62.5 (co-careldopa)
	co-beneldopa
	co-beneldopa product
	Madopar (co-beneldopa)
	co-careldopa
	selegiline
	Eldepryl (selegiline)

Multiple sclerosis disease modifying medications	Parkinson's disease medications
	Zelapar(selegiline)
	entacapone
	tolcapone
	Tasmar (tolcapone)
	amantadine
	Symmetrel (amantadine)
	orphenadrine
	Norgesic (orphenadrine)
	Norflex (orphenadrine)
	procyclidine
	Kemadrin (procyclidine)
	Arpicolin (procyclidine)
	trihexyphenidyl
	Artane (trihexyphenidyl)
	benzhexol

S – Characteristics of the depression sub-groups in UK Biobank

Table S.1 Severe recurrent depression group characteristics

	Information source		Definition	
	NHS hospital records	**Touchscreen mood questionnaire**	**Broad**	**Narrow**
n	137	8,905	9,016	26
Age, M (SD)	56.8 (8.6)	55.6 (8.1)	55.6 (8.1)	54.8 (8.2)
Female, *n* (%)[a]	96 (70.1)	5,144 (57.8)	5,225 (58.0)	15 (57.7)
Ethnic group, *n* (%)[a]				
White	130 (95.6)	8,231 (92.8)	8,337 (92.8)	24 (92.3)
Asian/Asian British	4 (2.9)	242 (2.7)	245 (2.7)	1 (3.9)
Black/black British	0 (0.0)	189 (2.1)	189 (2.1)	0 (0.0)
Other	2 (1.5)	208 (2.3)	209 (2.3)	1 (3.9)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	18 (13.2)	1,159 (13.0)	1,175 (13.1)	2 (7.7)
Qu2	18 (13.2)	1,417 (15.9)	1,433 (15.9)	2 (7.7)
Qu3	32 (23.5)	1,668 (18.8)	1,694 (18.8)	6 (23.1)
Qu4	30 (22.1)	2,086 (23.5)	2,110 (23.5)	6 (23.1)
Qu5 (most deprived)	38 (27.9)	2,558 (28.8)	2,586 (28.7)	10 (38.5)
Has a degree, *n* (%)[a]	41 (30.4)	3,204 (36.2)	3,235 (36.2)	10 (38.5)
Smoking status, *n* (%)[a]				
Never	75 (55.6)	4,065 (45.8)	4,127 (45.9)	13 (50.0)
Former	38 (28.2)	3,314 (37.3)	3,345 (37.2)	7 (26.9)
Current	22 (16.3)	1,501 (16.9)	1,517 (16.9)	6 (23.1)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	25 (18.4)	1,773 (19.9)	1,795 (19.9)	3 (11.5)
3-4 times per week	19 (14.0)	1,660 (18.7)	1,676 (18.6)	3 (11.5)
1-2 times per week	26 (19.1)	1,916 (21.5)	1,937 (21.5)	5 (19.2)
1-3 times per month	13 (9.6)	1,108 (12.5)	1,121 (12.5)	0 (0.0)
Special occasions only	27 (19.9)	1,308 (14.7)	1,330 (14.8)	5 (19.2)
Never (former drinker)	18 (13.2)	689 (7.7)	699 (7.8)	8 (30.8)
Never (not former drinker)	8 (5.9)	443 (5.0)	449 (5.0)	2 (7.7)
Any psychotropic medication, *n* (%)[a]	107 (80.5)	3,072 (34.9)	3,154 (35.4)	25 (96.2)
Lithium, *n* (%)[a]	13 (10.2)	95 (1.1)	105 (1.2)	3 (12.0)
Other mood stabiliser, *n* (%)[a]	10 (7.8)	189 (2.2)	195 (2.2)	4 (16.0)
SSRI antidepressant, *n* (%)[a]	42 (32.3)	1,741 (19.9)	1,770 (20.0)	13 (52.0)
Other antidepressant, *n* (%)[a]	59 (45.0)	1,095 (12.5)	1,142 (12.9)	12 (48.0)
Neuroticism score, M (SD)	7.2 (3.6)	6.8 (3.4)	6.8 (3.4)	7.3 (3.8)
Current depressive symptom score, M (SD)	4.2 (4.2)	3.3 (3.0)	3.3 (3.1)	4.0 (4.3)

M, mean; Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table S.2 Mild-moderate recurrent depression group characteristics

	Information source		Definition	
	NHS hospital records	**Touchscreen mood questionnaire**	**Broad**	**Narrow**
n	312	15,012	15,320	4
Age, M (SD)	54.7 (8.1)	55.4 (7.9)	55.4 (7.9)	55.3 (2.8)
Female, *n* (%)[a]	178 (57.1)	10,324 (68.8)	10,499 (68.5)	3 (75.0)
Ethnic group, *n* (%)[a]				
White	298 (96.8)	14,175 (94.7)	14,470 (94.8)	3 (75.0)
Asian/Asian British	2 (0.7)	247 (1.7)	249 (1.6)	0 (0.0)
Black/black British	4 (1.3)	244 (1.6)	247 (1.6)	1 (25.0)
Other	4 (1.3)	302 (2.0)	306 (2.0)	0 (0.0)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	27 (8.7)	2,384 (15.9)	2,411 (15.8)	0 (0.0)
Qu2	33 (10.6)	2,873 (19.2)	2,906 (19.0)	0 (0.0)
Qu3	51 (16.4)	3,091 (20.6)	3,141 (20.5)	1 (25.0)
Qu4	69 (22.2)	3,425 (22.9)	3,493 (22.8)	1 (25.0)
Qu5 (most deprived)	131 (42.1)	3,210 (21.4)	3,339 (21.8)	2 (50.0)
Has a degree, *n* (%)[a]	82 (27.1)	5,211 (34.9)	5,293 (34.8)	0 (0.0)
Smoking status, *n* (%)[a]				
Never	131 (42.3)	7,785 (52.0)	7,915 (51.8)	1 (25.0)
Former	95 (30.7)	5,450 (36.4)	5,543 (36.3)	2 (50.0)
Current	84 (27.1)	1,748 (11.7)	1,831 (12.0)	1 (25.0)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	55 (18.0)	2,896 (19.3)	2,951 (19.3)	0 (0.0)
3-4 times per week	38 (12.4)	3,204 (21.4)	3,241 (21.2)	1 (25.0)
1-2 times per week	71 (23.2)	3,734 (24.9)	3,804 (24.9)	1 (25.0)
1-3 times per month	28 (9.2)	1,992 (13.3)	2,019 (13.2)	1 (25.0)
Special occasions only	57 (18.6)	2,001 (13.3)	2,058 (13.4)	0 (0.0)
Never (former drinker)	43 (14.1)	640 (4.3)	682 (4.5)	1 (25.0)
Never (not former drinker)	14 (4.6)	538 (3.6)	552 (3.6)	0 (0.0)
Any psychotropic medication, *n* (%)[a]	231 (74.0)	2,772 (18.7)	3,001 (19.9)	2 (50.0)
Lithium, *n* (%)[a]	27 (9.0)	1 (0.01)	28 (0.2)	0 (0.0)
Other mood stabiliser, *n* (%)[a]	42 (14.0)	73 (0.5)	115 (0.8)	0 (0.0)
SSRI antidepressant, *n* (%)[a]	87 (28.6)	1,835 (12.4)	1,922 (12.8)	0 (0.0)
Other antidepressant, *n* (%)[a]	117 (38.1)	842 (5.7)	957 (6.4)	2 (50.0)
Neuroticism score, M (SD)	8.2 (3.3)	5.8 (3.2)	5.8 (3.2)	6.3 (3.5)
Current depressive symptom score, M (SD)	5.4 (3.6)	2.5 (2.4)	2.5 (2.5)	-

M, mean; Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

Table S.3 Single episode depression group characteristics

	Information source		Definition	
	NHS hospital records	**Touchscreen mood questionnaire**	**Broad**	**Narrow**
n	3830	7,927	11,686	71
Age, M (SD)	55.6 (8.0)	56.3 (8.0)	56.1 (8.0)	55.3 (8.5)
Female, *n* (%)[a]	2,246 (58.6)	5,042 (63.6)	7,246 (62.0)	42 (59.2)
Ethnic group, *n* (%)[a]				
White	3,593 (94.5)	7,575 (96.0)	11,106 (95.5)	62 (87.3)
Asian/Asian British	74 (2.0)	97 (1.2)	169 (1.5)	2 (2.8)
Black/black British	54 (1.4)	109 (1.4)	159 (1.4)	4 (5.6)
Other	81 (2.1)	114 (1.5)	192 (1.7)	3 (4.2)
Townsend quintile[b], *n* (%)[a]				
Qu1 (least deprived)	430 (11.2)	1,353 (17.1)	1,778 (15.3)	5 (7.0)
Qu2	545 (14.3)	1,621 (20.5)	2,161 (18.5)	5 (7.0)
Qu3	560 (14.6)	1,681 (21.3)	2,224 (19.1)	17 (23.9)
Qu4	811 (21.2)	1,846 (23.3)	2,639 (22.6)	18 (25.4)
Qu5 (most deprived)	1,479 (38.7)	1,407 (17.8)	2,860 (24.5)	26 (36.6)
Has a degree, *n* (%)[a]	805 (21.6)	2,893 (36.7)	3,681 (31.9)	17 (23.9)
Smoking status, *n* (%)[a]				
Never	1,605 (42.3)	4,125 (52.2)	5,697 (49.0)	33 (46.5)
Former	1,251 (33.0)	2,957 (37.4)	4,188 (36.0)	20 (28.2)
Current	941 (24.8)	825 (10.4)	1,748 (15.0)	18 (25.4)
Alcohol frequency, *n* (%)[a]				
Daily/almost daily	607 (16.0)	1,616 (20.4)	2,211 (19.0)	12 (17.1)
3-4 times per week	514 (13.6)	1,862 (23.5)	2,359 (20.3)	17 (24.3)
1-2 times per week	765 (20.2)	2,010 (25.4)	2,764 (23.7)	11 (15.7)
1-3 times per month	441 (11.6)	957 (12.1)	1,387 (11.9)	11 (15.7)
Special occasions only	689 (18.2)	911 (11.5)	1,592 (13.7)	8 (11.4)
Never (former drinker)	508 (13.4)	302 (3.8)	807 (6.9)	3 (4.3)
Never (not former drinker)	269 (7.1)	261 (3.3)	522 (4.5)	8 (11.4)
Any psychotropic medication, *n* (%)[a]	2,325 (61.9)	658 (8.4)	2,949 (25.7)	34 (48.6)
Lithium, *n* (%)[a]	76 (2.1)	4 (0.05)	79 (0.7)	1 (1.5)
Other mood stabiliser, *n* (%)[a]	182 (5.0)	43 (0.6)	221 (2.0)	4 (5.8)
SSRI antidepressant, *n* (%)[a]	1,294 (34.9)	343 (4.4)	1,617 (14.2)	20 (29.0)
Other antidepressant, *n* (%)[a]	922 (25.1)	240 (3.1)	1,150 (10.1)	12 (17.1)
Neuroticism score, M (SD)	7.5 (3.3)	4.2 (3.0)	5.2 (3.5)	5.3 (3.1)
Current depressive symptom score, M (SD)	4.4 (3.5)	1.3 (1.7)	2.3 (2.8)	2.1 (2.2)

M, mean; Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
a. Missing excluded from denominator.
b. Quintiles generated from the whole cohort.

T – Additional characteristics of the exposed and unexposed groups in UK Biobank

Additional characteristics of the exposed and unexposed groups analysed in the prevalence study in Chapter 5

	Unexposed comparison	Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow
n	104,410	3,020	607	56,425	7,583	850	259	1,905	931	916	323
Ethnic group											
n (%) missing	398 (0.4)	31 (1.0)	4 (0.7)	279 (0.5)	20 (0.3)	21 (2.5)	4 (1.5)	11 (0.6)	8 (0.9)	6 (0.7)	1 (0.3)
White, *n* (%)[a]	94,798 (91.1)	2,727 (91.2)	563 (93.4)	53,515 (95.3)	7,171 (94.8)	719 (86.7)	221 (86.7)	1,859 (98.2)	899 (97.4)	876 (96.3)	310 (96.3)
Asian/Asian British	3,643 (3.5)	84 (2.8)	6 (1.0)	887 (1.6)	144 (1.9)	24 (2.9)	8 (3.1)	5 (0.3)	4 (0.4)	15 (1.7)	6 (1.9)
Black/black British	3,121 (3.0)	84 (2.8)	13 (2.2)	718 (1.3)	108 (1.4)	49 (5.9)	17 (6.7)	11 (0.6)	5 (0.5)	8 (0.9)	1 (0.3)
Other	2,450 (2.4)	94 (3.1)	21 (3.5)	1,026 (1.8)	140 (1.9)	37 (4.5)	9 (3.5)	19 (1.0)	15 (1.6)	11 (1.2)	5 (1.6)
Townsend quintile[b]											
n (%) missing	157 (0.2)	3 (0.1)	1 (0.2)	101 (0.2)	19 (0.3)	3 (0.4)	1 (0.4)	4 (0.2)	2 (0.2)	2 (0.2)	1 (0.3)
Qu1 (least deprived), *n* (%)[a]	18,130 (17.4)	356 (11.8)	71 (11.7)	8,872 (15.8)	1,015 (13.4)	33 (3.9)	8 (3.1)	382 (20.1)	181 (19.5)	222 (24.3)	80 (24.8)
Qu2	21,340 (20.5)	405 (13.4)	78 (12.9)	10,071 (17.9)	1,304 (17.2)	48 (5.7)	11 (4.3)	382 (20.1)	173 (18.6)	188 (20.6)	66 (20.5)
Qu3	21,782 (20.9)	492 (16.3)	99 (16.3)	10,836 (19.2)	1,388 (18.4)	75 (8.9)	19 (7.4)	389 (20.5)	190 (20.5)	177 (19.4)	59 (18.3)
Qu4	23,337 (22.4)	676 (22.4)	133 (22.0)	12,380 (22.0)	1,663 (22.0)	165 (19.5)	44 (17.1)	373 (19.6)	187 (20.1)	167 (18.3)	53 (16.5)
Qu5 (most deprived)	19,664 (18.9)	1,088 (36.1)	225 (37.1)	14,165 (25.2)	2,194 (29.0)	526 (62.1)	176 (68.2)	375 (19.7)	198 (21.3)	160 (17.5)	64 (19.9)
Smoking status											
n (%) missing	372 (0.4)	26 (0.9)	5 (0.8)	236 (0.4)	21 (0.3)	16 (1.9)	4 (1.5)	17 (0.9)	11 (1.2)	10 (1.1)	5 (1.6)
Never, *n* (%)[a]	60,061 (57.7)	1,281 (42.8)	262 (43.5)	27,577 (49.1)	3,611 (47.8)	299 (35.9)	89 (34.9)	897 (47.5)	419 (45.5)	555 (61.3)	201 (63.2)
Former	35,126 (33.8)	1,024 (34.2)	196 (32.6)	20,210 (36.0)	2,559 (33.8)	237 (28.4)	74 (29.0)	690 (36.6)	337 (36.6)	293 (32.3)	98 (30.8)
Current	8,851 (8.5)	689 (23.0)	144 (23.9)	8,402 (15.0)	1,392 (18.4)	298 (35.7)	92 (36.1)	301 (15.9)	164 (17.8)	58 (6.4)	19 (6.0)
Alcohol frequency											
n (%) missing	75 (0.1)	22 (0.7)	4 (0.7)	157 (0.3)	28 (0.4)	13 (1.5)	2 (0.8)	5 (0.3)	4 (0.4)	4 (0.4)	1 (0.3)
Daily/almost daily, *n* (%)[a]	22,026 (21.1)	561 (18.7)	99 (16.4)	10,801 (19.2)	1,361 (18.0)	125 (14.9)	34 (13.2)	368 (19.4)	169 (18.2)	178 (19.5)	62 (19.3)
3-4 times per week	24,920 (23.9)	492 (16.4)	92 (15.3)	11,020 (19.6)	1,281 (17.0)	84 (10.0)	24 (9.3)	342 (18.0)	149 (16.1)	167 (18.3)	50 (15.5)
1-2 times per week	26,928 (25.8)	641 (21.4)	126 (20.9)	13,097 (23.3)	1,634 (21.6)	161 (19.2)	53 (20.6)	447 (23.5)	214 (23.1)	221 (24.2)	75 (23.3)
1-3 times per month	11,266 (10.8)	355 (11.8)	75 (12.4)	7,031 (12.5)	938 (12.4)	79 (9.4)	22 (8.6)	232 (12.2)	120 (12.9)	86 (9.4)	32 (9.9)
Special occasions only	11,449 (11.0)	473 (15.8)	99 (16.4)	8,153 (14.5)	1,207 (16.0)	157 (18.8)	51 (19.8)	295 (15.5)	158 (17.0)	124 (13.6)	55 (17.1)
Never (former drinker)	2,783 (2.7)	296 (9.9)	74 (12.3)	3,609 (6.4)	699 (9.3)	151 (18.0)	46 (17.9)	128 (6.7)	71 (7.7)	79 (8.7)	27 (8.4)
Never (not former drinker)	4,963 (4.8)	180 (6.0)	38 (6.3)	2,557 (4.5)	435 (5.8)	80 (9.6)	27 (10.5)	88 (4.6)	46 (5.0)	57 (6.3)	21 (6.5)
Any psychotropic medication								-	-	-	-
n (%) missing	1,180 (1.1)	40 (1.3)	3 (0.5)	949 (1.7)	58 (0.8)	23 (2.7)	7 (2.7)				
n (%)[a]	1,424 (1.4)	1,663 (55.8)	542 (89.7)	23,280 (42.0)	5,530 (73.5)	688 (83.2)	231 (91.7)				
Lithium	-					-	-	-	-	-	-
n (%) missing		90 (3.0)	15 (2.5)	1,965 (3.5)	245 (3.2)						
n (%)[a]		501 (17.1)	209 (35.3)	373 (0.7)	117 (1.6)						
Other mood stabiliser	-					-	-	-	-	-	-
n (%) missing		75 (2.5)	10 (1.7)	1,930 (3.4)	237 (3.1)						
n (%)[a]		531 (18.0)	267 (44.7)	768 (1.4)	196 (2.7)						
SSRI antidepressant	-					-	-	-	-	-	-
n (%) missing		81 (2.7)	15 (2.5)	1,316 (2.3)	123 (1.6)						
n (%)[a]		508 (17.3)	102 (17.2)	15,529 (28.2)	3,792 (50.8)						

	Unexposed comparison	Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow
Other antidepressant	-					-	-	-	-	-	-
n (%) missing		90 (3.0)	16 (2.6)	1,625 (2.9)	180 (2.4)						
n (%)[a]		428 (14.6)	128 (21.7)	7,239 (13.2)	1,737 (23.5)						
Traditional antipsychotic	-			-	-			-	-	-	-
n (%) missing		97 (3.2)	18 (3.0)			57 (6.7)	17 (6.6)				
n (%)[a]		131 (4.5)	39 (6.6)			174 (21.9)	62 (25.6)				
Second generation antipsychotic	-			-	-			-	-	-	-
n (%) missing		82 (2.7)	12 (2.0)			43 (5.1)	10 (3.9)				
n (%)[a]		470 (16.0)	205 (34.5)			439 (54.4)	169 (67.9)				
Multiple sclerosis disease-modifying medication	-	-	-	-	-	-	-			-	-
n (%) missing								153 (8.0)	94 (10.1)		
n (%)[a]								175 (10.0)	120 (14.3)		
Parkinson's disease medication	-	-	-	-	-	-	-	-	-		
n (%) missing										58 (6.3)	11 (3.4)
n (%)[a]										721 (84.0)	300 (96.2)
Neuroticism score											
n (%) missing	17,114 (16.4)	665 (22.0)	130 (21.4)	11,146 (19.8)	1,392 (18.4)	223 (26.2)	67 (25.9)	442 (23.2)	219 (23.5)	200 (21.8)	72 (22.3)
Mean (SD)	3.3 (2.9)	6.8 (3.6)	7.1 (3.6)	6.4 (3.4)	7.4 (3.2)	6.8 (3.6)	6.9 (3.5)	4.6 (3.4)	4.5 (3.3)	4.1 (3.3)	4.4 (3.4)
Current depressive symptom score											
n (%) missing	8,543 (8.2)	311 (10.3)	53 (8.7)	5,585 (9.9)	666 (8.8)	155 (18.2)	45 (17.4)	243 (12.8)	149 (16.0)	113 (12.3)	39 (12.1)
Mean (SD)	1.1 (1.6)	3.6 (3.3)	3.4 (3.2)	3.1 (3.0)	4.0 (3.4)	3.7 (3.3)	3.9 (3.3)	2.6 (2.4)	2.9 (2.5)	2.3 (2.4)	2.6 (2.5)

Qu, quintile; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.

a. Missing excluded from denominator.

b. Quintiles generated from the whole UK Biobank cohort.

U – Population attributable prevalence results in UK Biobank

Highest population attributable prevalence in each exposed group

	Impaired *n*	Unimpaired *n*	Total *n*	Prevalence[a]	Population attributable fraction	Population attributable prevalence per 100,000
Mania/BD						
Point estimate						
Exposed *n*	89	911	1,000	0.0890[b]	0.0102	45.2
Unexposed *n*	4,336	94,664	99,000	0.0438		
Total *n*	4,425	95,575	100,000	0.0443		
CI lower						
Exposed *n*	67	933	1,000	0.0670	0.0053	23.2
Unexposed *n*	4,336	94,664	99,000	0.0438		
Total *n*	4,403	95,597	100,000	0.0440		
CI upper						
Exposed *n*	112	888	1,000	0.1120	0.0153	68.2
Unexposed *n*	4,336	94,664	99,000	0.0438		
Total *n*	4,448	95,552	100,000	0.0445		
Major depression						
Point estimate						
Exposed *n*	778	9,222	10,000	0.0778[c]	0.0468	256.0
Unexposed *n*	4,698	85,302	90,000	0.0522		
Total *n*	5,476	94,524	100,000	0.0548		
CI lower						
Exposed *n*	652	9,348	10,000	0.0652	0.0243	130.0
Unexposed *n*	4,698	85,302	90,000	0.0522		
Total *n*	5,350	94,650	100,000	0.0535		
CI upper						
Exposed *n*	903	9,097	10,000	0.0903	0.0680	381.0
Unexposed *n*	4,698	85,302	90,000	0.0522		
Total *n*	5,601	94,399	100,000	0.0560		
Schizophrenia						
Point estimate						
Exposed *n*	204	796	1,000	0.2040[d]	0.0283	151.8
Unexposed *n*	5,167	93,833	99,000	0.0522		
Total *n*	5,371	94,629	100,000	0.0537		
CI lower						
Exposed *n*	105	895	1,000	0.1050	0.0100	52.8
Unexposed *n*	5,167	93,833	99,000	0.0522		
Total *n*	5,272	94,728	100,000	0.0527		
CI upper						
Exposed *n*	304	696	1,000	0.3040	0.0460	251.8
Unexposed *n*	5,167	93,833	99,000	0.0522		
Total *n*	5,471	94,529	100,000	0.0547		
Multiple sclerosis						
Point estimate						
Exposed *n*	37.64	162.36	200	0.1882[e]	0.0055	27.7
Unexposed *n*	4,970.00	94,830.00	99,800	0.0498		
Total *n*	5,007.64	94,992.36	100,000	0.0501		
CI lower						
Exposed *n*	32.56	167.44	200	0.1628	0.0045	22.6
Unexposed *n*	4,970.00	94,830.00	99,800	0.0498		
Total *n*	5,002.56	94,997.44	100,000	0.0500		
CI upper						
Exposed *n*	42.74	157.26	200	0.2137	0.0065	32.8
Unexposed *n*	4,970.00	94,830.00	99,800	0.0498		
Total *n*	5,012.74	94,987.26	100,000	0.0501		

	Impaired *n*	Unimpaired *n*	Total *n*	Prevalence[a]	Population attributable fraction	Population attributable prevalence per 100,000
Parkinson's disease						
Point estimate						
Exposed *n*	37.38	162.62	200	0.1869[f]	0.0051	26.9
Unexposed *n*	5,209.56	94,590.44	99,800	0.0522		
Total *n*	5,246.94	94,753.06	100,000	0.0525		
CI lower						
Exposed *n*	11.54	188.46	200	0.0577	0.0002	1.1
Unexposed *n*	5,209.56	94,590.44	99,800	0.0522		
Total *n*	5,221.10	94,753.06	100,000	0.0522		
CI upper						
Exposed *n*	63.20	136.80	200	0.3160	0.0100	52.8
Unexposed *n*	5,209.56	94,590.44	99,800	0.0522		
Total *n*	5,272.76	94,727.24	100,000	0.0527		

BD, bipolar disorder; CI, confidence interval.
a. All standardised prevalence estimates taken from Table 5.2 in Chapter 5.
b. Visuospatial memory estimates in narrowly-defined group.
c. Numeric memory estimates in narrowly-defined group.
d. Numeric memory estimates in broadly-defined group.
e. Reaction time estimates in narrowly-defined group.
f. Numeric memory estimates in narrowly-defined group.

V – Sensitivity analysis results for prevalence study in UK Biobank

Table V.1 Crude prevalence results, excluding exposed participants with comorbid psychiatric or neurological conditions

Impairment threshold		**Mania/bipolar**		**Major depression**		**Schizophrenia**		**Multiple sclerosis**		**Parkinson's disease**	
		Broad	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**
Reasoning	*n*	1,343	191	28,271	4,163	26	7	268	108	159	60
≤ unexposed 4th percentile	Crude P %	6.78	3.66	3.76	4.32	7.69	0	1.49	0.93	6.29	10.0
score	95% CI	5.44, 8.12	1.00, 6.32	3.54, 3.98	3.70, 4.94	0.00, 17.93	-	0.04, 2.94	0.00, 2.74	2.52, 10.06	2.41, 17.59
(Unexposed prevalence	Crude PR	1.63*	0.88	0.90*	1.04	1.85	-	0.36*	0.22	1.51	2.40*
4.16%)	95% CI	1.33, 1.99	0.43, 1.83	0.85, 0.97	0.90, 1.20	0.49, 7.01	-	0.14, 0.95	0.03, 1.57	0.83, 2.76	1.13, 5.14
Reaction time	*n*	1,974	342	42,791	4,840	28	8	276	114	169	64
> unexposed 95th	Crude P %	5.98	6.73	4.72	5.64	17.86	12.50	10.14	12.28	6.51	6.25
percentile score	95% CI	4.93, 7.03	4.07, 9.39	4.52, 4.92	4.99, 6.29	3.67, 32.05	0.00, 35.42	6.58, 13.70	6.26, 18.30	2.79, 10.23	0.32, 12.18
(Unexposed prevalence	Crude PR	1.20*	1.35	0.95*	1.13*	3.59*	2.51	2.04*	2.47*	1.31	1.26
4.98%)	95% CI	1.01, 1.43	0.91, 2.01	0.90, 0.99	1.01, 1.28	1.62, 7.94	0.40, 15.71	1.43, 2.90	1.51, 4.03	0.74, 2.32	0.49, 3.24
Numeric memory	*n*	370	50	8,602	1,196	7	3	82	36	48	17
≤ unexposed 5th percentile	Crude P %	8.65	6.00	5.16	7.11	0	0	7.32	11.11	6.25	11.76
score	95% CI	5.79, 11.51	0.00, 12.58	4.69, 5.63	5.65, 8.57	-	-	1.68, 12.96	0.84, 21.38	0.00, 13.10	0.00, 27.07
(Unexposed prevalence	Crude PR	1.66*	1.15	0.99	1.36*	-	-	1.40	2.13	1.20	2.25
5.22%)	95% CI	1.19, 2.32	0.38, 3.45	0.89, 1.10	1.10, 1.68	-	-	0.65, 3.03	0.84, 5.37	0.40, 3.59	0.61, 8.29
Visuospatial memory	*n*	1,996	346	43,128	4,875	30	8	279	116	172	65
> unexposed 95th	Crude P %	5.91	6.65	4.55	4.45	3.33	0	3.23	3.45	7.56	7.69
percentile score	95% CI	4.88, 6.94	4.02, 9.28	4.35, 4.75	3.87, 5.03	0.00, 9.75	-	1.15, 5.30	0.13, 6.77	3.61, 11.51	1.21, 14.17
(Unexposed prevalence	Crude PR	1.35*	1.52*	1.04	1.02	0.76	-	0.74	0.79	1.73*	1.76
4.38%)	95% CI	1.13, 1.61	1.02, 2.26	0.99, 1.10	0.89, 1.16	0.11, 5.23	-	0.39, 1.40	0.30, 2.06	1.02, 2.91	0.76, 4.08
Prospective memory	*n*	1,390	199	28,976	4,286	29	8	277	114	169	64
Incorrect score	Crude P %	26.12	32.16	21.26	22.84	48.28	62.50	23.83	28.07	25.44	34.38
(Unexposed prevalence	95% CI	23.81, 28.43	25.67, 38.65	20.79, 21.73	21.58, 24.10	30.09, 66.47	28.95, 96.05	18.81, 28.85	19.82, 36.32	18.87, 32.01	22.74, 46.02
22.82%)	Crude PR	1.14*	1.41*	0.93*	1.00	2.12*	2.74*	1.04	1.23	1.11	1.51*
	95% CI	1.05, 1.25	1.15, 1.72	0.91, 0.95	0.95, 1.06	1.45, 3.08	1.60, 4.68	0.85, 1.29	0.92, 1.65	0.86, 1.44	1.07, 2.11

CI, confidence interval; P, prevalence; PR, prevalence ratio.
* $p < 0.05$

Table V.2 Prevalence of cognitive impairment across groups, standardised for age group, gender and educational attainment

Impairment threshold		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**
Reasoning	*n*	1,866	318	35,211	6,052	274	86	572	272	283	115
≤ unexposed 4th percentile	Standardised P %	7.32	7.24	4.33	5.45	10.86	-[b]	3.29	-[b]	5.95	9.07
score	95% CI	6.14, 8.50	4.39, 10.09	4.11, 4.54	4.88, 6.02	7.17, 14.54	-	1.83, 4.75	-	3.19, 8.70	3.82, 14.32
(Unexposed prevalence	Standardised PR	1.76*	1.74*	1.04[a]	1.31*	2.61*	-	0.79	-	1.43	2.18*
4.16%)	95% CI	1.48, 2.11	1.13, 2.70	0.98, 1.10	1.16, 1.48	1.78, 3.82	-	0.44, 1.42	-	0.73, 2.81	1.01, 4.69
Reaction time	*n*	2,955	593	55,773	7,484	798	243	1,866	906	897	316
> unexposed 95th	Standardised P %	7.22	8.72	5.53	7.02	13.50	11.45	15.34	19.27	6.27	5.98
percentile score	95% CI	6.29, 8.15	6.44, 10.99	5.34, 5.72	6.44, 7.60	11.13, 15.87	7.45, 15.46	13.70, 16.97	16.70, 21.84	4.69, 7.86	3.36, 8.59
(Unexposed prevalence	Standardised PR	1.45*	1.75*	1.11*[c]	1.41*[d]	2.71*	2.30*	3.08*[e]	3.87*[f]	1.26	1.20
4.98%)	95% CI	1.26, 1.68	1.30, 2.35	1.06, 1.17	1.28, 1.55	2.21, 3.32	1.52, 3.49	2.70, 3.52	3.30, 4.54	0.92, 1.71	0.73, 1.98
Numeric memory	*n*	505	75	10,808	1,746	63	20	193	103	94	35
≤ unexposed 5th percentile	Standardised P %	10.07	-[b]	5.85	7.88	19.42	-[b]	7.20	-[b]	-[b]	-[b]
score	95% CI	7.45, 12.70	-	5.40, 6.29	6.62, 9.15	9.65, 29.19	-	3.56, 10.85	-	-	-
(Unexposed prevalence	Standardised PR	1.93*	-	1.12*	1.51*	3.72*	-	1.38	-	-	-
5.22%)	95% CI	1.45, 2.58	-	1.01, 1.23	1.26, 1.82	2.20, 6.28	-	0.69, 2.75	-	-	-
Visuospatial memory	*n*	3,006	606	56,308	7,571	833	256	1,887	920	907	321
> unexposed 95th	Standardised P %	6.48	9.15	5.12	5.65	9.68	9.72	6.53	7.97	6.57	7.05
percentile score	95% CI	5.60, 7.36	6.86, 11.45	4.94, 5.31	5.13, 6.17	7.67, 11.69	6.09, 13.35	5.41, 7.64	6.22, 9.72	4.96, 8.18	4.25, 9.85
(Unexposed prevalence	Standardised PR	1.48*	2.09*	1.17*	1.29*	2.21*	2.22*[g]	1.49*	1.82*	1.50*	1.61
4.38%)	95% CI	1.28, 1.72	1.59, 2.76	1.12, 1.23	1.16, 1.44	1.74, 2.82	1.38, 3.58	1.21, 1.83	1.39, 2.37	1.14, 1.98	0.98, 2.65
Prospective memory	*n*	1,959	335	36,237	6,267	325	103	608	291	306	124
Incorrect score	Standardised P %	29.89	36.74	22.82	25.56	44.50	42.22	26.24	29.89	27.84	35.60
(Unexposed prevalence	95% CI	27.87, 31.92	31.58, 41.90	22.39, 23.25	24.48, 26.64	39.10, 49.90	32.68, 51.76	22.75, 29.74	24.63, 35.15	22.82, 32.86	27.17, 44.03
22.82%)	Standardised PR	1.31*[h]	1.61*	1.00[i]	1.12*[j]	1.95*	1.85*	1.15	1.31*	1.22	1.56*
	95% CI	1.22, 1.41	1.38, 1.86	0.98, 1.03	1.07, 1.17	1.71, 2.23	1.44, 2.37	0.99, 1.34	1.07, 1.60	0.98, 1.53	1.16, 2.09

CI, confidence interval; P, prevalence; PR, prevalence ratio.
Standardised estimates are directly standardised by age, gender and education with reference to the unexposed comparison group.
* $p < 0.05$
a. Significant interaction with gender: women PR = 0.95 (CI 0.88, 1.03); men PR = 1.13 (CI 1.03, 1.24).
b. Estimates not reported because more than 3 of 8 strata contained no exposed participants with impairment.
c. Significant interaction with gender: women PR = 1.04 (CI 0.98, 1.10); men PR = 1.21 (CI 1.12, 1.30).
d. Significant interaction with gender: women PR = 1.28 (CI 1.14, 1.44); men PR = 1.57 (CI 1.34, 1.83).
e. Significant interaction with age and gender: <60 years PR = 4.29 (CI 3.63, 5.09); ≥60 years PR = 2.43 (CI 1.99, 2.97); women PR = 2.53 (CI 2.20, 2.92); men PR = 3.78 (CI 3.06, 4.66).
f. Significant interaction with age and gender: <60 years PR = 5.96 (CI 4.91, 7.24); ≥60 years PR = 2.75 (CI 2.12, 3.56); women PR = 2.96 (CI 2.46, 3.57); men PR = 5.01 (CI 3.94, 6.37).
g. Significant interaction with education: no degree PR = 1.46 (CI 0.76, 2.82); has degree PR = 3.88 (CI 1.96, 7.68).
h. Significant interaction with age: <60 years PR = 1.41 (CI 1.28, 1.54); ≥60 years PR = 1.24 (CI 1.11, 1.28).
i. Significant interaction with gender and education: women PR = 0.97 (CI 0.94, 0.99); men PR = 1.04 (CI 1.01, 1.08); no degree PR = 1.03 (CI 0.99, 1.05); has degree = PR 0.94 (CI 0.89, 0.98).
j. Significant interaction with gender: women PR = 1.06 (CI 0.99, 1.12); men PR = 1.19 (CI 1.10, 1.28).

Table V.3 Crude prevalence results, excluding participants with missing educational attainment data

Impairment threshold		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
		Broad	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**	**Broad**	**Narrow**
Reasoning	*n*	1,855	318	35,011	6,029	272	86	569	270	282	115
≤ unexposed 4th percentile	Crude P %	6.90	6.29	4.14	5.19	11.03	6.98	2.81	1.85	4.96	6.69
score	95% CI	5.74, 8.05	3.62, 8.96	3.93, 4.35	4.63, 5.75	7.31, 14.75	1.59, 12.37	1.45, 4.17	0.24, 3.46	2.43, 7.49	2.12, 11.26
(Unexposed prevalence	Crude PR	1.70*	1.55*	1.02	1.28*	2.72*	1.72	0.69	0.46	1.23	1.72
4.05%)	95% CI	1.44, 2.02	1.02, 2.38	0.96, 1.08	1.15, 1.43	1.94, 3.82	0.80, 3.73	0.43, 1.13	0.19, 1.09	0.74, 2.05	0.88, 3.36
Reaction time	*n*	2,925	589	55,263	7,433	778	236	1,831	883	881	311
> unexposed 95th	Crude P %	6.63	7.64	5.28	6.54	12.72	10.17	14.2	17.78	6.81	7.40
percentile score	95% CI	5.73, 7.53	5.49, 9.79	5.09, 5.47	5.98, 7.10	10.38, 15.06	6.31, 14.03	12.60, 15.80	15.26, 20.30	5.15, 8.47	4.49, 10.31
(Unexposed prevalence	Crude PR	1.34*	1.54*	1.07*	1.32*	2.57*	2.06*	2.87*	3.59*	1.38*	1.49*
4.95%)	95% CI	1.17, 1.54	1.16, 2.05	1.02, 1.12	1.21, 1.45	2.14, 3.10	1.41, 3.01	2.56, 3.22	3.11, 4.15	1.08, 1.76	1.01, 2.22
Numeric memory	*n*	501	74	10,748	1,735	63	20	192	103	93	35
≤ unexposed 5th percentile	Crude P %	9.98	5.41	5.59	7.55	20.63	15.00	6.25	8.74	8.6	14.29
score	95% CI	7.36, 12.60	0.26, 10.56	5.16, 6.02	6.31, 8.79	10.64, 30.62	0.00, 30.65	2.83, 9.67	3.29, 14.19	2.90, 14.30	2.70, 25.88
(Unexposed prevalence	Crude PR	1.95*	1.06	1.09	1.48*	4.03*	2.93*	1.22	1.71	1.68	2.79*
5.12%)	95% CI	1.49, 2.55	0.41, 2.74	0.99, 1.20	1.24, 1.75	2.48, 6.56	1.03, 8.33	0.70, 2.12	0.91, 3.19	0.86, 3.27	1.24, 6.29
Visuospatial memory	*n*	2,980	603	55,843	7,524	819	249	1,866	905	896	317
> unexposed 95th	Crude P %	6.24	8.46	4.74	5.14	9.04	8.03	6.06	7.07	8.04	7.57
percentile score	95% CI	5.37, 7.11	6.24, 10.68	4.56, 4.92	4.64, 5.64	7.08, 11.00	4.65, 11.41	4.98, 7.14	5.40, 8.74	6.26, 9.82	4.66, 10.48
(Unexposed prevalence	Crude PR	1.44*	1.95*	1.09*	1.18*	2.08*	1.85*	1.39*	1.63*	1.85*	1.74*
4.35%)	95% CI	1.25, 1.65	1.49, 2.53	1.04, 1.14	1.07, 1.31	1.67, 2.59	1.21, 2.81	1.16, 1.67	1.28, 2.06	1.48, 2.31	1.18, 2.56
Prospective memory	*n*	1,944	334	35,988	6,232	323	103	602	287	304	124
Incorrect score	Crude P %	28.81	34.73	22.16	24.49	44.58	43.69	26.25	28.92	28.95	36.29
(Unexposed prevalence	95% CI	26.80, 30.82	29.62, 39.84	21.73, 22.59	23.42, 25.26	39.16, 50.00	34.11, 53.27	22.74, 29.76	23.67, 34.17	23.85, 34.05	27.83, 44.75
22.58%)	Crude PR	1.28*	1.54*	0.98	1.08*	1.97*	1.93*	1.16*	1.28*	1.28*	1.61*
	95% CI	1.19, 1.37	1.33, 1.78	0.96, 1.00	1.04, 1.13	1.75, 2.23	1.55, 2.41	1.02, 1.33	1.07, 1.54	1.07, 1.53	1.27, 2.03

CI, confidence interval; P, prevalence; PR, prevalence ratio.
* $p < 0.05$

Table V.4 Prevalence of cognitive impairment in alternative versions of mood disorder groups[a]

Impairment threshold		Mania/bipolar		Major depression	
		Broad	**Narrow**	**Broad**	**Narrow**
Reasoning	*n*	572	152	9,925	816
≤ unexposed 4th percentile score	Crude prevalence %	7.69	10.53	5.95	8.82
(Unexposed prevalence 4.16%)	95% CI	5.51, 9.87	5.65, 15.41	5.48, 6.42	6.87, 10.77
	Crude prevalence ratio	1.85*	2.53*	1.43*	2.12*
	95% CI	1.39, 2.46	1.59, 4.03	1.32, 1.56	1.70, 2.65
	Standardised prevalence %	7.61	11.15	6.07	9.24
	95% CI	5.44, 9.79	6.15, 16.15	5.60, 6.54	7.25, 11.22
	Standardised prevalence ratio	1.83*	2.68*	1.46*[b]	2.22*[c]
	95% CI	1.37, 2.45	1.64, 4.37	1.34, 1.60	1.75, 2.80
Reaction time	*n*	1,621	424	30,051	2,123
> unexposed 95th percentile score	Crude prevalence %	8.27	8.25	5.80	7.68
(Unexposed prevalence 4.98%)	95% CI	6.93, 9.61	5.63, 10.87	5.54, 6.06	6.55, 8.81
	Crude prevalence ratio	1.66*	1.66*	1.16*	1.54*
	95% CI	1.41, 1.96	1.21, 2.28	1.10, 1.23	1.33, 1.79
	Standardised prevalence %	8.62	9.51	6.13	8.12
	95% CI	7.25, 9.98	6.72, 12.30	5.85, 6.40	6.96, 9.28
	Standardised prevalence ratio	1.73*	1.91*	1.23*[d]	1.63*
	95% CI	1.46, 2.05	1.37, 2.66	1.17, 1.30	1.40, 1.91
Numeric memory	*n*	152	37	2,910	233
≤ unexposed 5th percentile score	Crude prevalence %	12.50	0.00	7.49	11.16
(Unexposed prevalence 5.22%)	95% CI	7.24, 17.76	-	6.53, 8.45	7.12, 15.20
	Crude prevalence ratio	2.40*	-	1.44*	2.14*
	95% CI	1.57, 3.66	-	1.25, 1.65	1.48, 3.08
	Standardised prevalence %	12.95	-	7.51	11.07
	95% CI	7.61, 18.28	-	6.56, 8.47	7.04, 15.09
	Standardised prevalence ratio	2.48*	-	1.44*	2.12*
	95% CI	1.61, 3.81	-	1.24, 1.66	1.40, 3.21
Visuospatial memory	*n*	1,651	434	30,365	2,151
> unexposed 95th percentile score	Crude prevalence %	7.63	7.37	5.37	6.51
(Unexposed prevalence 4.38%)	95% CI	6.35, 8.91	4.91, 9.83	5.12, 5.62	5.47, 7.55
	Crude prevalence ratio	1.74*	1.68*	1.23*	1.49*
	95% CI	1.47, 2.07	1.20, 2.35	1.16, 1.30	1.26, 1.75
	Standardised prevalence %	7.80	8.23	5.87	7.05
	95% CI	6.50, 9.09	5.65, 10.82	5.60, 6.13	5.97, 8.13
	Standardised prevalence ratio	1.78*[e]	1.88*	1.34*	1.61*
	95% CI	1.50, 2.12	1.33, 2.67	1.26, 1.42	1.36, 1.90
Prospective memory	*n*	614	165	10,395	877
Incorrect score	Crude prevalence %	36.64	41.21	26.35	31.58
(Unexposed prevalence 22.82%)	95% CI	32.83, 40.45	33.70, 48.72	25.50, 27.20	28.50, 34.66
	Crude prevalence ratio	1.61*	1.81*	1.15*	1.38*
	95% CI	1.45, 1.78	1.50, 2.17	1.12, 1.19	1.25, 1.53
	Standardised prevalence %	36.74	42.67	27.16	33.09
	95% CI	32.93, 40.55	35.13, 50.22	26.30, 28.01	29.97, 36.20
	Standardised prevalence ratio	1.61*[f]	1.87*	1.19*[g]	1.45*[h]
	95% CI	1.45, 1.79	1.54, 2.26	1.15, 1.23	1.32, 1.61

CI, confidence interval.
Standardised estimates are directly standardised by age and gender with reference to the unexposed comparison group.
* $p < 0.05$
a. Exposure groups formed without reference to the touchscreen mood questionnaire data.
b. Significant interaction with gender: women ratio = 1.31 (CI 1.18, 1.46); men ratio = 1.62 (CI 1.41, 1.87).
c. Significant interaction with gender: women ratio = 1.65 (CI 1.20, 2.29); men ratio = 2.84 (CI 2.05, 3.92).
d. Significant interaction with age and gender: <60 years ratio = 1.42 (CI 1.31, 1.54); ≥60 years ratio = 1.13 (CI 1.05, 1.22); women ratio = 1.13 (CI 1.05, 1.21); men ratio = 1.36 (CI 1.25, 1.49).
e. Significant interaction with age: <60 years ratio = 2.17 (CI 1.74, 2.72); ≥60 years ratio = 1.52 (CI 1.16, 1.98).
f. Significant interaction with age: <60 years ratio = 1.79 (CI 1.55, 2.07); ≥60 years ratio = 1.46 (CI 1.25, 1.71).
g. Significant interaction with age and gender: <60 years ratio = 1.25 (CI 1.19, 1.32); ≥60 years ratio = 1.13 (CI 1.08, 1.19); women ratio = 1.12 (CI 1.07, 1.17); men ratio = 1.26 (CI 1.19, 1.34).
h. Significant interaction with gender: women ratio = 1.31 (CI 1.15, 1.49); men ratio = 1.61 (CI 1.39, 1.88).

Table V.5 Comparison of multiple sclerosis and Parkinson's disease groups identified via linked hospital records, including and excluding data from Scotland

	Multiple sclerosis		Parkinson's disease	
	Including Scotland	**Excluding Scotland**	**Including Scotland**	**Excluding Scotland**
n	1,059	981	381	362
Age (years) Mean (SD)	55.3 (7.5)	55.4 (7.4)	62.3 (5.6)	62.2 (5.6)
Female %	72.2	72.3	41.7	40.6
Has a degree %	30.8	29.6	25.5	26.6
On medication[a] %	13.0	13.6	84.9	85.0
Reasoning score Mean (SD)	5.7 (2.0)	5.7 (2.0)	5.7 (2.1)	5.7 (2.1)
Reaction time (ms) Median (Q1, Q3)	598 (518, 715)	598 (520, 719)	578 (516, 653)	575 (512, 653)
Numeric memory score Mean (SD)	6.4 (1.3)	6.4 (1.3)	6.5 (1.4)	6.5 (1.4)
Visuospatial memory errors Median (Q1, Q3)	3 (2, 6)	3 (2, 6)	4 (2, 6)	4 (2, 6)
Prospective memory % correct	71.1	71.1	65.5	65.5

Q, quartile; SD, standard deviation.

a. Multiple sclerosis disease modifying medication or Parkinson's disease medication, respectively.

Table V.6a Characteristics of exposed and unexposed groups[a], with or without reasoning data

	Unexposed comparison		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
	Has data	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**
n	100,927	1,595	1,866	40	35,211	485	274	17	572	17	283	6
Age (years) Mean (SD)	56.9 (8.2)	59.6 (7.8)	54.7 (8.1)	58.6 (7.5)	55.7 (8.0)	58.8 (7.7)	54.4 (8.0)	54.6 (8.8)	55.7 (7.7)	59.8 (6.8)	62.1 (5.7)	60.7 (3.1)
Female %	50.0	51.0	49.8	50.0	65.0	56.9	33.9	35.3	74.5	52.9	37.5	50.0
Has a degree %	35.5	15.5	37.0	21.6	34.9	14.0	31.3	11.8	36.2	6.3	34.4	0.0
No known comorbidity %	-	-	72.0	50.0	80.3	19.7	9.5	5.9	46.9	29.4	56.2	50.0

SD, standard deviation.
a. Broad group definitions.

Table V.6b Characteristics of exposed and unexposed groups[a], with or without reaction time data

	Unexposed comparison		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
	Has data	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**
n	103,303	1,107	2,955	65	55,773	652	798	52	1,866	39	897	19
Age (years) Mean (SD)	57.0 (8.2)	57.2 (8.7)	54.9 (8.0)	55.2 (8.1)	55.6 (7.9)	57.1 (8.2)	53.7 (8.0)	56.8 (8.1)	55.3 (7.5)	57.4 (6.2)	62.2 (5.6)	63.8 (4.3)
Female %	50.0	50.5	52.5	58.5	64.9	58.0	34.0	42.3	73.4	76.9	37.8	36.8
Has a degree %	35.0	22.9	37.9	14.6	32.7	16.2	26.9	7.3	34.9	8.6	31.0	20.0
No known comorbidity %	-	-	67.3	46.2	76.9	63.2	3.5	3.9	14.8	7.7	18.8	15.8

SD, standard deviation.
a. Broad group definitions.

Table V.6c Characteristics of exposed and unexposed groups[a], with or without numeric memory data

	Unexposed comparison		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
	Has data	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**
n	30,351	749	505	22	10,808	301	63	9	193	9	94	3
Age (years) Mean (SD)	56.8 (8.3)	56.7 (8.7)	54.7 (8.1)	55.2 (7.6)	55.3 (8.1)	55.2 (8.6)	54.3 (8.9)	52.9 (5.9)	56.1 (7.8)	54.4 (8.8)	61.2 (6.2)	66.7 (2.1)
Female %	49.8	43.8	48.9	59.1	65.5	53.8	39.7	22.2	75.7	55.6	39.4	33.3
Has a degree %	32.4	16.1	35.5	23.8	33.0	13.4	27.0	11.1	31.3	37.5	30.1	0.0
No known comorbidity %	-	-	73.3	50.0	79.6	76.1	11.1	11.1	42.5	33.3	51.1	100.0

SD, standard deviation.
a. Broad group definitions.

Table V.6d Characteristics of exposed and unexposed groups[a], with or without visuospatial memory data

	Unexposed comparison		Mania/bipolar		Major depression		Schizophrenia		Multiple sclerosis		Parkinson's disease	
	Has data	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**	**Has data**	**Missing data**
n	104,410	0	3,006	14	56,308	117	833	17	1,887	18	907	9
Age (years) Mean (SD)	57.0 (8.2)	-	54.9 (8.0)	51.5 (8.4)	55.6 (7.9)	55.9 (7.9)	53.9 (8.1)	54.5 (7.5)	55.4 (7.5)	55.8 (5.8)	62.2 (5.5)	58.3 (7.3)
Female %	50.0	-	52.5	71.4	64.9	47.9	34.8	17.7	73.6	66.7	37.6	55.6
Has a degree %	34.9	-	37.5	-	32.6	-	25.9	-	34.4	-	30.8	-
No known comorbidity %	-	-	66.9	57.1	76.8	64.1	3.6	0.0	14.8	0.0	19.0	0.0

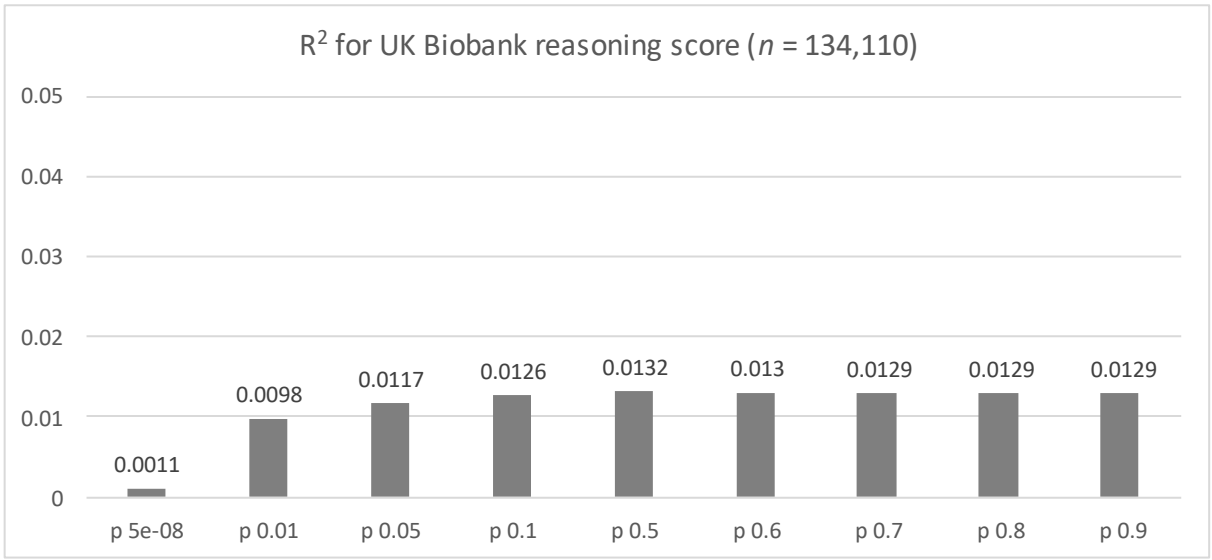
SD, standard deviation.
a. Broad group definitions.

W – Genome-wide polygenic score results

GPS were created at a range of *p* value thresholds using summary statistics from previous GWAS of education, bipolar disorder, major depression and schizophrenia, as described in Chapters 7 and 8. The results below show the model R^2 at each *p* threshold for the target phenotype in UK Biobank, and the association between deciles of the optimum GPS and the target phenotype. Analyses were restricted to unrelated participants of white British genetic ancestry, and the GPS were residualised for genotyping array and batch, UK Biobank assessment centre, and the first 20 genetic principal components.

Education

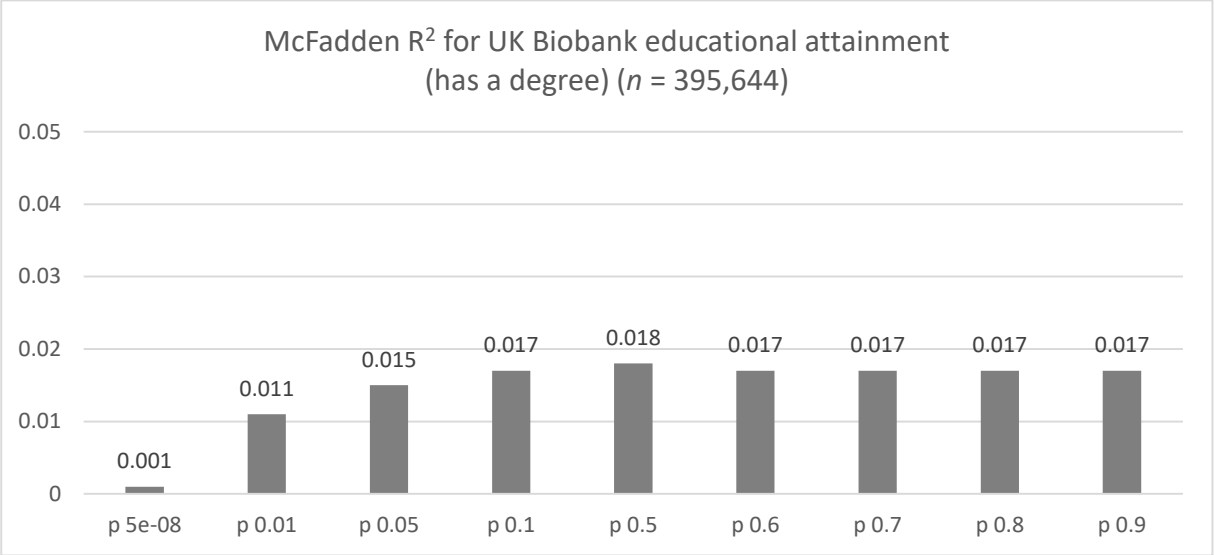
Using continuous GPS to predict UK Biobank reasoning score:



Using deciles of GPS as predictor, at *p* 0.5 (highest R^2 in graph above):

Decile	*Coefficient*	*95% CI lower*	*95% CI upper*
D1	(reference)		
D2	.1529	.1010	.2049
D3	.2585	.2067	.3104
D4	.3130	.2613	.3647
D5	.3963	.3446	.4480
D6	.4513	.3996	.5030
D7	.5101	.4585	.5617
D8	.5817	.5301	.6332
D9	.6819	.6303	.7336
D10	.8866	.8349	.9382

Using continuous GPS to predict UK Biobank educational attainment:

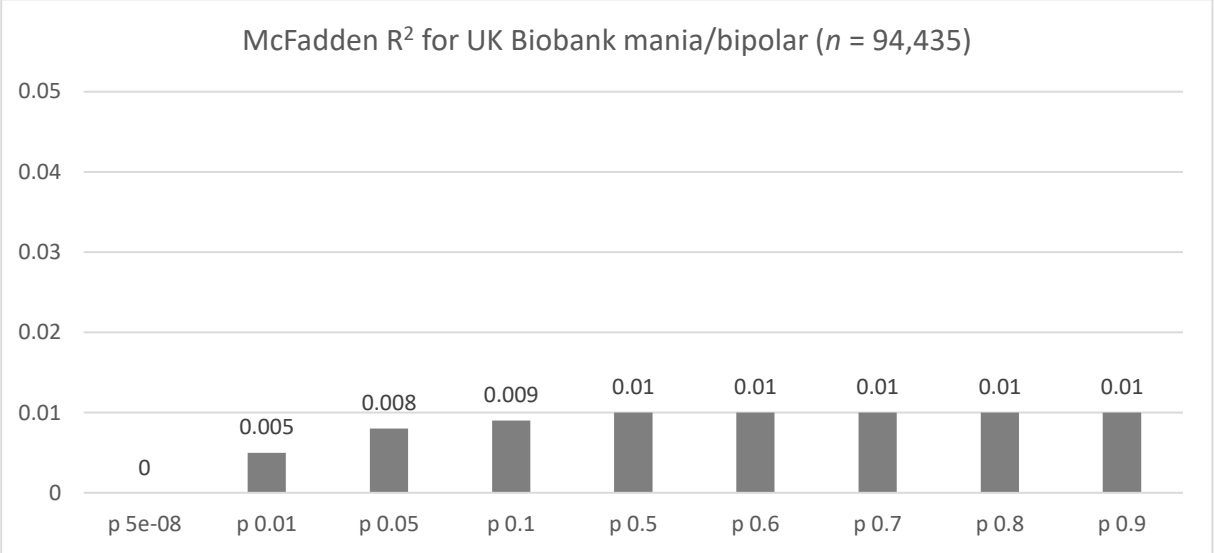


Using deciles of GPS as predictor, at *p* 0.5 (highest R^2 in graph above):

Decile	*Odds ratio*	*95% CI lower*	*95% CI upper*
D1	(reference)		
D2	1.2340	1.1946	1.2747
D3	1.3952	1.3512	1.4406
D4	1.4956	1.4488	1.5440
D5	1.6369	1.5860	1.6893
D6	1.7798	1.7248	1.8366
D7	1.9698	1.9093	2.0323
D8	2.1644	2.0983	2.2327
D9	2.4101	2.3368	2.4858
D10	3.0417	2.9497	3.1366

Bipolar disorder

Using continuous GPS to predict UK Biobank mania/bipolar disorder (broad):

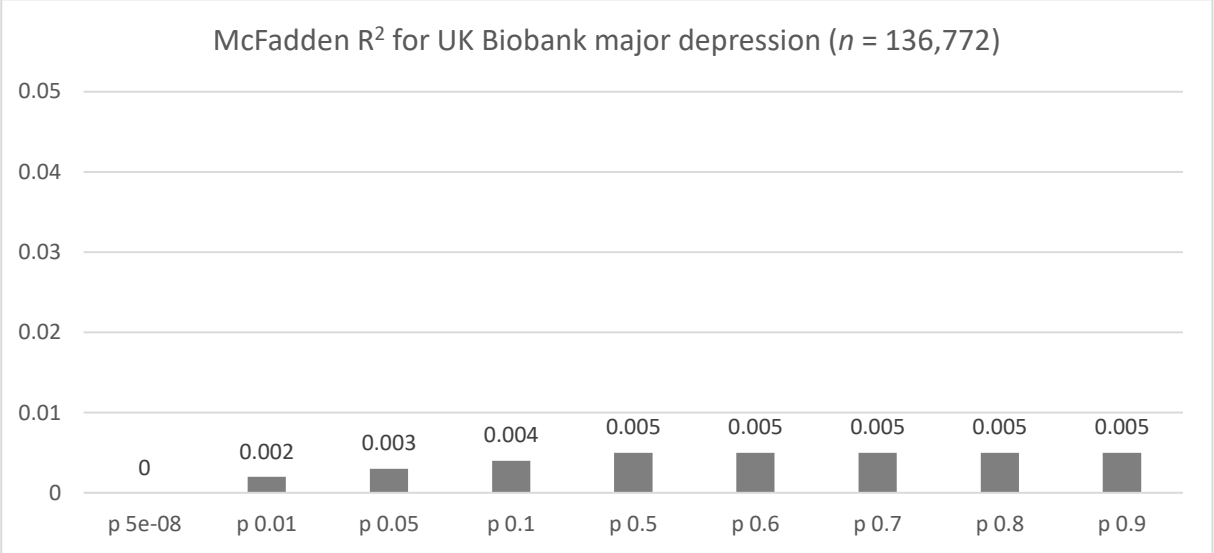


Using deciles of GPS as predictor, at *p* 0.5 (highest R^2 in graph above):

Decile	*Odds ratio*	*95% CI lower*	*95% CI upper*
D1	(reference)		
D2	1.2430	.9936	1.5549
D3	1.3469	1.0807	1.6787
D4	1.4606	1.1754	1.8150
D5	1.5537	1.2547	1.9239
D6	1.5615	1.2605	1.9343
D7	1.6603	1.3427	2.0529
D8	1.9207	1.5622	2.3615
D9	2.2258	1.8181	2.7249
D10	2.8981	2.3829	3.5247

Major depression

Using continuous GPS to predict UK Biobank major depression (broad):

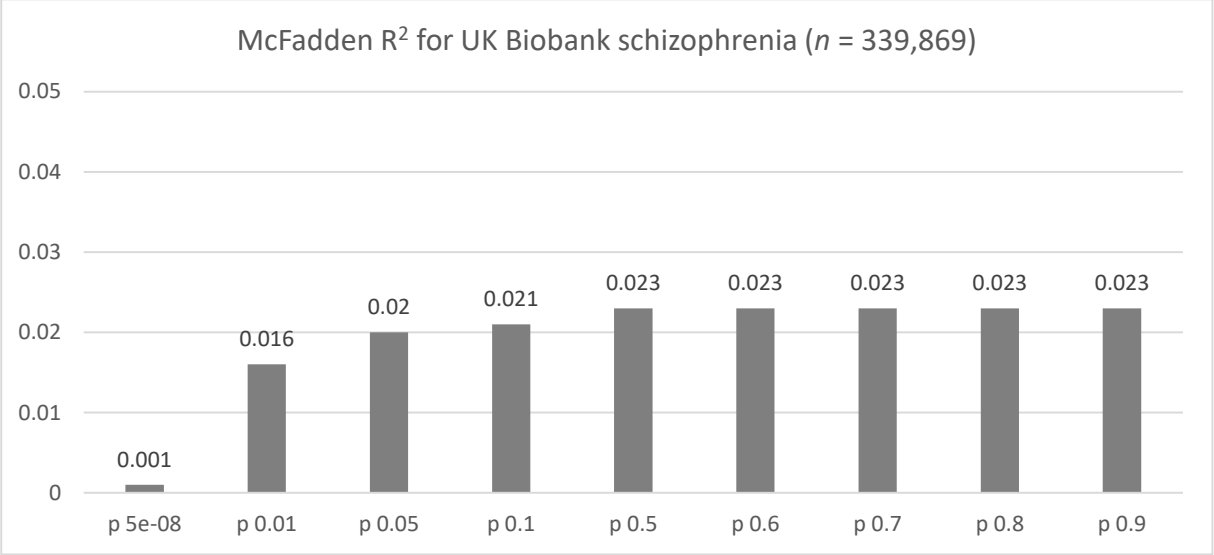


Using deciles of GPS as predictor, at *p* 0.5 (highest R^2 in graph above):

Decile	*Odds ratio*	*95% CI lower*	*95% CI upper*
D1	(reference)		
D2	1.1760	1.1148	1.2406
D3	1.2209	1.1578	1.2875
D4	1.2407	1.1769	1.3080
D5	1.3218	1.2539	1.3933
D6	1.3621	1.2923	1.4357
D7	1.4215	1.3491	1.4977
D8	1.5444	1.4663	1.6266
D9	1.5965	1.5159	1.6814
D10	1.8576	1.7645	1.9556

Schizophrenia

Using continuous GPS to predict UK Biobank schizophrenia (broad):



Using deciles of GPS as predictor, at *p* 0.5 (highest R^2 in graph above):

Decile	*Odds ratio*	*95% CI lower*	*95% CI upper*
D1	(reference)		
D2	1.4074	.6251	3.1690
D3	1.4114	.6268	3.1779
D4	2.2286	1.0552	4.7069
D5	1.8212	.8406	3.9458
D6	3.6847	1.8284	7.4258
D7	3.6812	1.8266	7.4188
D8	4.3097	2.1620	8.5906
D9	5.3652	2.7264	10.5579
D10	7.4185	3.8259	14.3846

X – Testable independency results for mania/bipolar DAG

Summary of residual partial correlations from alternative versions of the mania/bipolar DAG:

DAG description	**Number of implied independencies**	***n* (%) of results in each correlation coefficient range**							**Largest correlation coefficient**
		|0.00| to |0.09|	**|0.10| to |0.19|**	**|0.20| to |0.29|**	**|0.30| to |0.39|**	**|0.40| to |0.49|**	**|0.50| to |0.59|**	**|0.60| to |0.69|**	
Original DAG (edu→bipolar)	142	105 (73.9)	19 (13.4)	3 (2.1)	7 (4.9)	2 (1.4)	3 (2.1)	3 (2.1)	-0.67 curr_psych_meds ⊥ famhx_pd | adult_SES bipolar bipolar_geno past_mental_state
Original DAG with education arrow reversed (edu←bipolar)	142	102 (71.8)	18 (12.7)	5 (3.5)	8 (5.6)	3 (2.1)	3 (2.1)	3 (2.1)	-0.67 curr_psych_meds ⊥ famhx_pd | adult_SES bipolar bipolar_geno past_mental_state
Modified DAG with new node representing other neuro/psy conditions (edu→bipolar) (edu→other_neupsy)	138	110 (79.7)	19 (13.8)	4 (2.9)	4 (2.9)	1 (0.7)	0 (0)	0 (0)	0.43 curr_psych_meds ⊥ famhx_pd | adult_SES age bipolar bipolar_geno child_adversity edu_attain gender other_neupsy
Modified DAG with new node representing other neuro/psy conditions (edu←bipolar) (edu←other_neupsy)	138	109 (79.0)	19 (13.8)	4 (2.9)	5 (3.6)	1 (0.7)	0 (0)	0 (0)	0.43 curr_psych_meds ⊥ famhx_pd | adult_SES bipolar bipolar_geno edu_attain gender other_neupsy past_mental_state

Optimal model fit is indicated by smaller correlation coefficients. Coefficients ≥ |0.10| may indicate model misspecification, measurement error, selection bias etc.
The third DAG above was taken forward into the causal analyses.

Results of additional regression models to obtain more details on the highest residual correlations indicated above (≥ |0.30|):

(1) curr_psych_meds ⊥ famhx_pd | adult_SES age bipolar bipolar_geno child_adversity edu_attain gender other_neupsy
Odds ratio 0.89 (95% CI 0.62, 1.28)

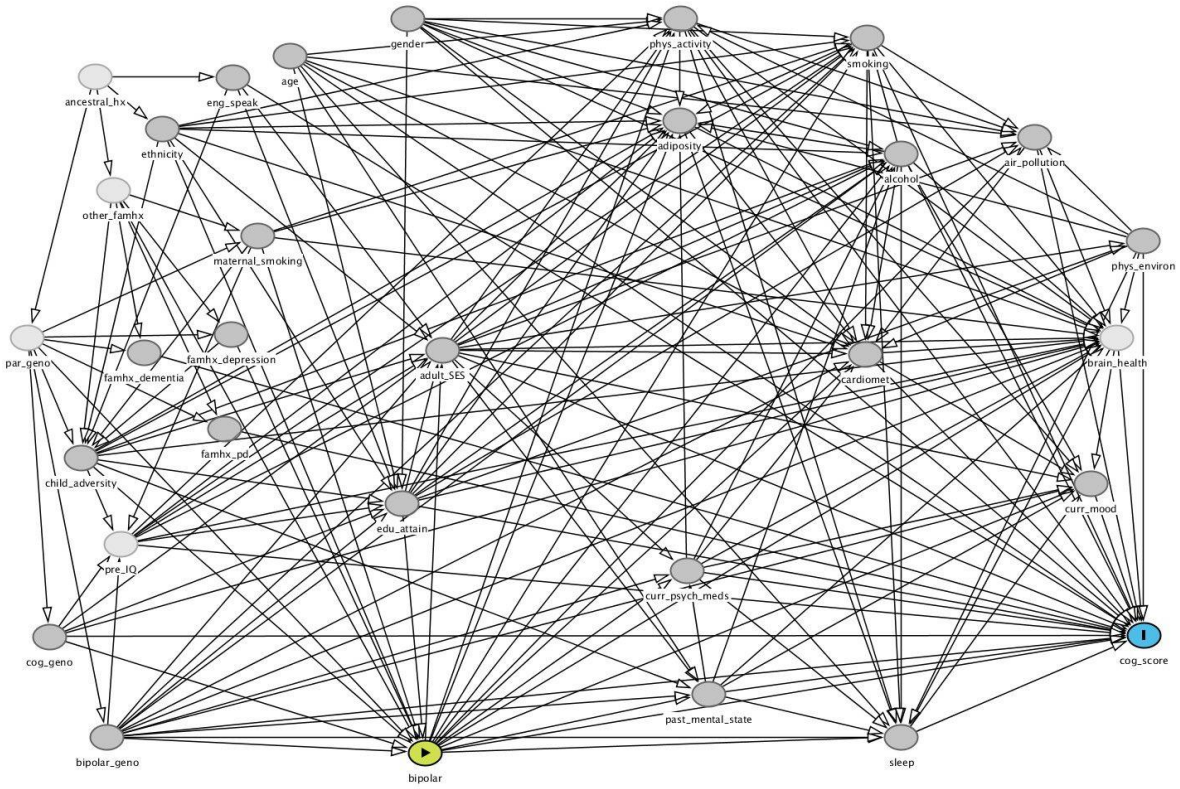
(2) curr_psych_meds ⊥ famhx_depression | adult_SES bipolar bipolar_geno edu_attain gender other_neupsy past_mental_state
Odds ratio 1.25 (95% CI 1.11, 1.40)

(3) curr_psych_meds ⊥ famhx_pd | adult_SES bipolar bipolar_geno edu_attain gender other_neupsy past_mental_state
Odds ratio 1.02 (95% CI 0.83, 1.25)

(4) famhx_dementia ⊥ phys_environ | bipolar bipolar_geno child_adversity cog_geno edu_attain eng_speak ethnicity gender maternal_smoking other_neupsy
Odds ratio 0.94 (95% CI 0.86, 1.02)

(5) famhx_pd ⊥ phys_environ | bipolar bipolar_geno child_adversity cog_geno edu_attain eng_speak ethnicity gender maternal_smoking other_neupsy
Odds ratio 1.09 (95% CI 0.94, 1.27)

Y – Explanation of DAG used in causal analyses



Original DAG used in the mania/bipolar disorder analyses reported in Chapter 7

The above DAG (shown as Fig. 7.2 in Chapter 7) represents the original assumptions made about the causal relationship between mania/BD and cognitive function. As explained in Chapter 6 (Box 6.1), the presence of an arrow represents the weak assumption of a causal relationship for at least one member of the population, and the absence of an arrow represents the strong assumption of no causal relationship for any member of the population. All nodes that are causally antecedent to a given node are known as ancestors; shared ancestors of the exposure and outcome are potential confounders, because they lie on back-door paths. Intermediate nodes were included in the DAG where these were of interest in the mediation analyses or were required as common parents of other pairs of nodes; these were not exhaustive, and so each individual arrow could in principle be shown in more detail as a chain of intermediate nodes that, for the purposes of the present research, are omitted or unknown. The node names, the constructs they are intended to represent, and the corresponding measures in UK Biobank, are listed in the table below.

Node name	Construct represented	Measurement in UK Biobank
Exposure		
bipolar	Lifetime history of mania or bipolar disorder (prior to cognitive assessment)	Disorder exposure status (versus comparison group)
Outcome		
cog_score	Performance on cognitive assessment	• Reasoning • Reaction time • Numeric memory • Visuospatial memory • Prospective memory
Shared ancestors of exposure and outcome		
age	Age	Age in years
ancestral_hx	Various ancestral/migration factors that determine ethnicity, country of origin and family history (genetic and non-genetic)	*Unmeasured*
bipolar_geno	Genotype associated with bipolar disorder	Genome-wide polygenic score
child_adversity	Adverse experiences in childhood	Childhood abuse and neglect (self-reported as 'never true' to 'very often true'): When I was growing up… a) I felt loved b) People in my family hit me so hard that it left me with bruises or marks c) I felt that someone in my family hated me d) Someone molested me (sexually) e) There was someone to take me to the doctor if I needed it
cog_geno	Genotype associated with cognitive function	Genome-wide polygenic score (using education GWAS as proxy)
edu_attain	Educational attainment	Has a degree or not (self-reported)
eng_speak	Born in an English-speaking country	Born in an English-speaking country (self-reported)
ethnicity	Ethnic background	• Ethnic category (self-reported) • Genetically-identified white British ancestry
famhx_depression	Parent/sibling with depression	Self-report of biological parent or sibling with 'severe depression'
gender	Gender	Self-reported male or female
maternal_smoking	Mother smoked around time of participant's birth	Participant's response to "Did your mother smoke regularly around the time when you were born?"
other_famhx	Other aspects of family history (non-genetic) and circumstances/environment	*Unmeasured*
par_geno	Genotype of parents	*Unmeasured*
pre_IQ	Premorbid intellectual ability	*Unmeasured*
Intermediates between exposure and outcome		
adiposity	Body fat	Body mass index
adult_SES	Socioeconomic status or deprivation in adulthood	Townsend index score
air_pollution	Airborne toxic particles/gases	Neighbourhood measures of: • Particulate matter • Nitrogen dioxide
alcohol	Frequency/amount of alcohol consumption	Self-reported frequency of intake
brain_health	Structural/functional brain state	*Unmeasured (except for small subgroup)*
cardiomet	History of cardiometabolic disease	• Self-reported history of angina, hypertension or diabetes (non-gestational) • Adjudicated history of myocardial infarction or stroke

Node name	Construct represented	Measurement in UK Biobank
curr_mood	Mood state at time of cognitive assessment	Patient Health Questionnaire (four self-reported items)
curr_psych_meds	Psychotropic medication at time of cognitive assessment	On any psychotropic medication (self-reported)
past_mental_state	Past psychiatric symptoms/illness course/duration/severity, over and above history of simply having exposure of interest or not	Number of depressed/unenthusiastic episodes (self-reported on touchscreen or web)
phys_activity	Level of physical activity	International Physical Activity Questionnaire (self-reported)
phys_environ	Physical aspects of the local environment	• Inverse distance to nearest major road • Home area population density
sleep	Sleep pattern/quality/duration	Self-reported sleeplessness/insomnia (never/rarely; sometimes; usually)
smoking	Tobacco smoking history	Self-reported smoking status (never; former; current)
Other ancestors of outcome (not descended from exposure)		
famhx_dementia	Parent/sibling with dementia	Self-report of biological parent or sibling with 'Alzheimer's/dementia'
famhx_pd	Parent/sibling with Parkinson's disease	Self-report of biological parent or sibling with 'Parkinson's disease'

Shared ancestors of exposure and outcome

Older age increases the risk of cognitive impairment, although the trajectory and mechanisms are not fully understood (Deary et al., 2009). Age was also assumed to have an indirect effect on mania/BD status, via educational attainment. Gender differences in average cognitive performance have often been reported, although again the causal mechanisms are not well understood (Miller & Halpern, 2014). Gender was assumed not to affect mania/BD status (Blanco et al., 2017; Lloyd et al., 2005), except through educational attainment. As described in Chapter 7, the direction of the relationship between educational attainment and mania/BD was uncertain, and the model was tested with the arrow as shown above and with it reversed. Given the temporal order of the measures in UK Biobank, educational attainment (past) was assumed to influence cognitive performance (current). Premorbid intellectual ability was not measured in UK Biobank, but was depicted as an antecedent of current cognitive performance and of other nodes such as educational attainment, socioeconomic status and health-related behaviours (e.g. smoking). No arrow was drawn between premorbid ability and mania/BD status, because it was assumed that any statistical association between them would be accounted for by their shared genetic and early life antecedents (see below), or by the indirect causal path through educational attainment.

Genotypes associated with bipolar disorder and with cognitive function were assumed to have shared effects on those respective phenotypes and on other outcomes (Hagenaars et al., 2016; Hill, Davies, et al., 2016). These genotypes were depicted as descending from parental genotype (unmeasured). Parental genotype and other aspects of family history (unmeasured) were also assumed to affect parental behaviour (maternal smoking measure), childhood adversity, and family history of psychiatric and neurological conditions. A distal node representing ancestral history was conceptualised as giving rise to parental genotype and other aspects of family history, as well as ethnicity and English-speaking status. Ethnicity was assumed to be a possible antecedent of mania/BD (Lloyd et al., 2005), and (through other nodes such as socioeconomic status) of cognitive performance. Being from a non-English-speaking country was assumed to influence educational attainment and childhood adversity (e.g. among individuals who had migrated at a young age), in turn influencing mania/BD status, and it was also assumed to affect cognitive performance.

Maternal smoking was assumed to affect mania/BD (Talati et al., 2013), and to affect cognitive function indirectly through other nodes (e.g. brain health). Childhood adversity (conceptualised broadly when constructing the graph, although measured in UK Biobank solely as abuse and neglect history) was assumed to be a possible cause of mania/BD (Aas et al., 2016), and of cognitive function via nodes such as educational attainment, socioeconomic status and health-related behaviours. Finally, family history of depression was depicted as a cause of mania/BD (Mortensen, Pedersen, Melbye, Mors, & Ewald, 2003), and of cognitive function via childhood adversity.

Intermediates between exposure and outcome

Lifetime history of mania/BD was assumed to affect multiple behaviour-related measures, namely physical activity, adiposity, alcohol consumption and smoking (Scott et al., 2015), and these in turn were assumed to influence cognitive function (Baumgart et al., 2015), including via their effects on brain structure. Cardiometabolic disease was also assumed to be influenced by mania/BD status (Fiedorowicz et al., 2011; Kemp & Fan, 2012) and to affect cognitive outcome (Knopman et al., 2001; Qiu & Fratiglioni, 2015). It was assumed that mania/BD might affect socioeconomic status in adulthood (e.g. via impact on occupational functioning) (Gitlin & Miklowitz, 2017); this in turn might influence performance on cognitive tests (Lyu & Burr, 2016), although this relationship is complicated by shared genetic factors (Marioni et al., 2014), as shown in the graph. Note that early life socioeconomic status was not measured in UK Biobank, but this can be

conceptualised as part of the childhood adversity node, and can thus be assumed to influence both mania/BD and adult socioeconomic status. With regard to local environment variables, mania/BD was assumed to influence exposure to air pollution, including via socioeconomic status, smoking and physical activity; the relationship between mania/BD and other aspects of the physical environment was assumed to arise indirectly via socioeconomic status. Physical environment exposures, including pollution, were assumed to affect cognitive function (Killin, Starr, Shiue, & Russ, 2016; Power, Adar, Yanosky, & Weuve, 2016).

It was assumed that mood state around the time of the cognitive assessment would be influenced by mania/BD status, and would in turn affect cognitive performance (Bourne et al., 2013; Ganguli, Du, Dodge, Ratcliff, & Chang, 2006). Mood state was here measured by items assessing current depressive symptoms only; manic mood state (unmeasured in UK Biobank) was not included as a separate node in the graph, as previous research had reported no association between residual mania and cognitive performance (Bourne et al., 2013). Sleeplessness was also depicted as an intermediate between mania/BD status and cognitive performance. Although sleep disturbance can be a trigger for relapse in BD, it was placed temporally downstream of mania/BD status in this graph because it represented recent sleep patterns (in the four-week period preceding the UK Biobank cognitive assessment), whereas the mania/BD node represented lifetime status. Sleeplessness is an ongoing problem for many people with BD (A. G. Harvey, Schmidt, Scarna, Semler, & Goodwin, 2005), and it may affect cognitive performance (Fortier-Brochu & Morin, 2014; Goldman-Mellor et al., 2015).

A node representing past mental state was included as an intermediate between mania/BD status and cognitive score. Although the past mental state node and the mania/BD exposure node both represent past states (i.e. lifetime experiences up to the time of the cognitive assessment), mania/BD was placed first in the temporal order depicted in the graph, on the grounds that having mania/BD influences the severity of the illness experience over time (measured here as number of depressed episodes). It was not a requirement in the exposure definition used here that a participant had to have experienced more than one affective episode to be classified in the mania/BD group, and so it was deemed more plausible that mania/BD status would influence the number of episodes, rather than the reverse. Similarly, current psychotropic medications at the time of the cognitive assessment were assumed to be a consequence of lifetime mania/BD status, and being on such medications did not contribute to the exposure classification used here. The temporal order of the past

mental state and current psychotropic medication nodes was depicted such that the former influenced the latter, although it is likely that there is a reciprocal relationship between these variables over time (i.e. a graph for a longitudinal analysis might show mental state at time 1 influencing medication at time 2, and medication at time 2 influencing mental state at time 3). Given the nature of the present cross-sectional analysis, however, it was considered plausible that cumulative lifetime affective episodes (past mental state) would be an antecedent of current medication status. Both past mental state and current psychotropic medications were assumed to affect cognitive performance (Balanzá-Martínez et al., 2010; Bourne et al., 2013; Robinson & Ferrier, 2006).

A node representing brain health (i.e. as potentially measured by structural volume and integrity, and functional activation and connectivity) was depicted as intermediate between mania/BD status and cognitive performance. This assumes that changes occur in the brain as a consequence of mania/BD, as shown in longitudinal studies (Lim et al., 2013), although it may also be the case that some brain changes (e.g. reduced white matter tract integrity) are a marker of BD vulnerability that precedes illness onset, given that similar findings are evident in unaffected relatives of people with BD (Miskowiak et al., 2017). Structural and functional brain changes were assumed to cause cognitive impairment, although—as shown in the graph—premorbid cognitive ability and shared genetic antecedents likely contribute to this relationship (Deary et al., 2009). It was not assumed that every causal path leading to cognitive outcome went through brain health, however, as other factors may be at play (such as confidence or test experience) that would affect performance on cognitive tests but are not necessarily mediated by brain structure or function. Since neuroimaging data were available only for a relatively small and non-representative sub-group (~2%) of the UK Biobank cohort, this node was tagged as unmeasured when planning the analyses in this thesis.

Other ancestors of outcome

Family history of neurodegenerative disease was assumed to be a potential additional cause of cognitive impairment, given that participants with this background may be at higher risk of cognitive decline arising from disease processes not necessarily captured by their own medical history data (e.g. individuals with unrecognised early-stage disease). This was represented in the graph by separate nodes for family history of dementia and of Parkinson’s disease, although this could have been depicted equivalently using one node, because they were considered to share the same antecedents and consequences (separate

nodes were included originally because PD-specific analyses were undertaken for prevalence elsewhere in the thesis, although causal analyses were not attempted for PD owing to insufficient background clinical data). Family history of neurodegenerative disease was assumed not to be a causal antecedent of mania/BD status.

Z – Supplementary results for mania/bipolar mediation analyses

Table Z.1 Tests of interactions between exposure and mediators

	Coefficient for mania/BD * mediator	95% CI	*p*
Mediator: Cardiometabolic disease			
Reasoning[a]	0.029	-0.207, 0.264	0.812
Reaction time[a]	0.149	-0.047, 0.345	0.137
Numeric memory[a]	-0.124	-0.652, 0.403	0.644
Visuospatial memory[a]	0.239	-0.019, 0.497	0.070
Prospective memory[b]	0.991	0.475, 2.071	0.982
Mediator: Psychotropic medication			
Reasoning[a]	0.005	-0.255, 0.264	0.973
Reaction time[a]	0.095	-0.110, 0.301	0.364
Numeric memory[a]	0.311	-0.294, 0.916	0.313
Visuospatial memory[a]	-0.045	-0.295, 0.205	0.725
Prospective memory[b]	0.883	0.435, 1.796	0.732

BD, bipolar disorder; CI, confidence interval.
All models included mania/BD, the mediator and their product, as well as all the covariates entered into the `gformula` mediation models.
a. Linear regression model with outcome measured in z-score units.
b. Logistic regression model with outcome measured as correct or not; estimate expressed as odds ratio.

Table Z.2 Mediation of the effect of mania/bipolar disorder on cognitive outcome via cardiometabolic disease, with missing data imputation

	n	Estimate	95% CI[a]
Reasoning[b]	80,698		
TCE		-0.092	-0.158, -0.026
CDE		-0.087	-0.152, -0.022
NDE		-0.085	-0.151, -0.019
NIE		-0.007	-0.017, 0.003
Reaction time[b]	82,648		
TCE		-0.094	-0.150, -0.037
CDE		-0.091	-0.148, -0.034
NDE		-0.091	-0.148, -0.033
NIE		-0.003	-0.014, 0.008
Numeric memory[b]	26,248		
TCE		-0.142	-0.275, -0.008
CDE		-0.158	-0.291, -0.024
NDE		-0.150	-0.283, -0.018
NIE		0.008	-0.011, 0.027
Visuospatial memory[b]	81,773		
TCE		-0.206	-0.267, -0.145
CDE		-0.200	-0.261, -0.139
NDE		-0.210	-0.272, -0.149
NIE		0.005	-0.007, 0.016
Prospective memory[c]	82,194		
TCE		-0.041	-0.062, -0.021
CDE		-0.037	-0.057, -0.017
NDE		-0.039	-0.059, -0.018
NIE		-0.003	-0.006, 0.001

CDE, controlled direct effect when cardiometabolic disease = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; NO_2, nitrogen dioxide; PM_{10}, particulate matter of up to 10μm diameter; TCE, total causal effect.

Models were restricted to participants of white British genetic ancestry, and were adjusted for age, gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), body mass index, alcohol frequency, smoking status, physical activity, and psychotropic medication. Missing mediator and covariate data were imputed via a single stochastic imputation using chained equations.

a. Normal-based, from bootstrapped standard error (1000 replicates).

b. Estimate expressed as a standardised mean difference.

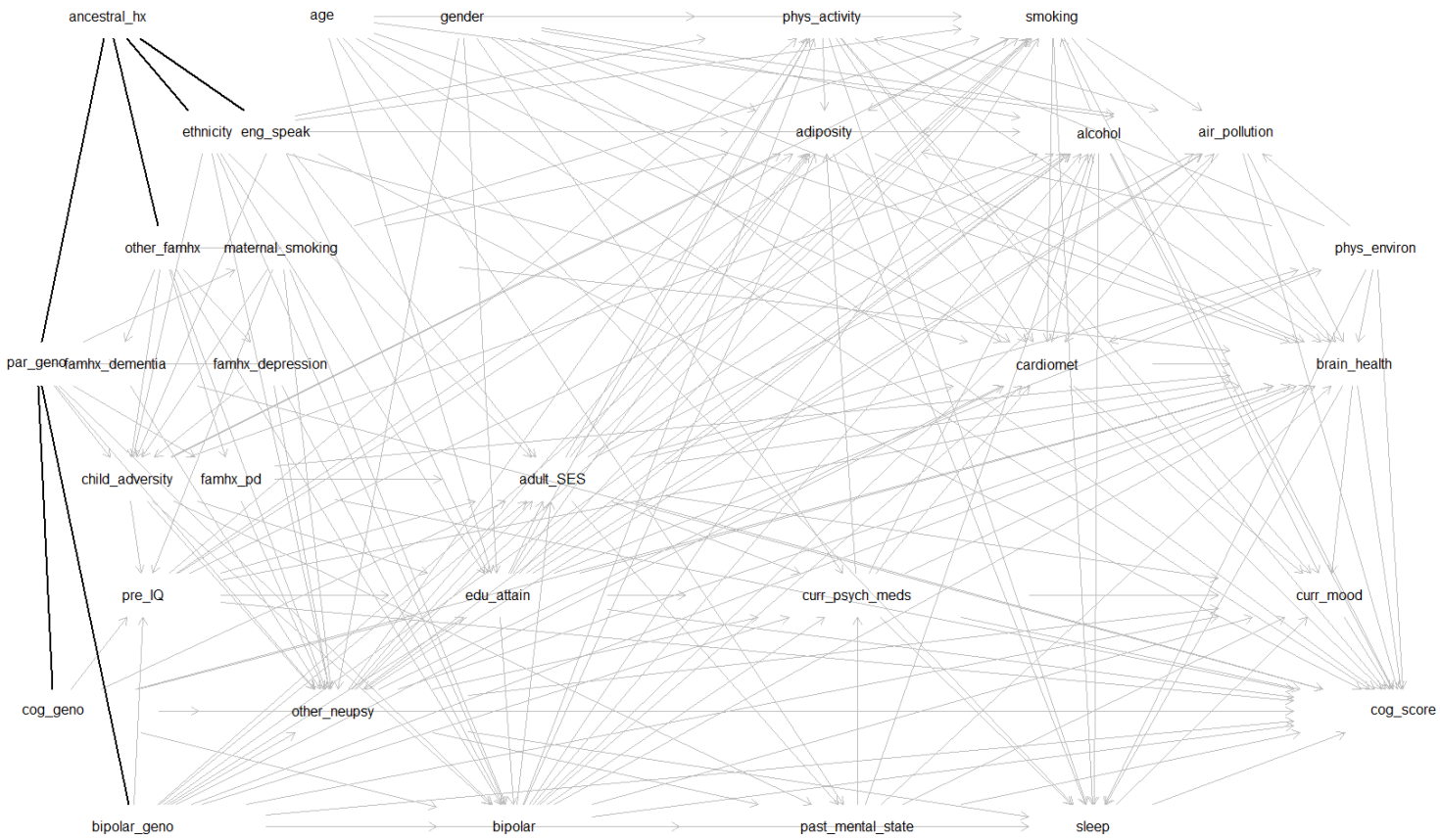
c. Estimate expressed as a risk difference for the probability of being correct.

Table Z.3 Mediation of the effect of mania/bipolar disorder on cognitive outcome via psychotropic medication, with missing data imputation

	n	Estimate	95% CI[a]
Reasoning[b]	80,698		
TCE		-0.095	-0.162, -0.027
CDE		-0.034	-0.103, 0.035
NDE		-0.034	-0.103, 0.035
NIE		-0.061	-0.078, -0.044
Reaction time[b]	82,648		
TCE		-0.080	-0.136, -0.024
CDE		-0.034	-0.095, 0.028
NDE		-0.024	-0.085, 0.037
NIE		-0.056	-0.078, -0.033
Numeric memory[b]	26,248		
TCE		-0.147	-0.276, -0.018
CDE		-0.079	-0.210, 0.051
NDE		-0.082	-0.212, 0.049
NIE		-0.065	-0.098, -0.033
Visuospatial memory[b]	81,773		
TCE		-0.196	-0.255, -0.137
CDE		-0.147	-0.210, -0.084
NDE		-0.146	-0.209, -0.083
NIE		-0.050	-0.076, -0.024
Prospective memory[c]	82,194		
TCE		-0.025	-0.045, -0.005
CDE		-0.010	-0.029, 0.010
NDE		-0.012	-0.031, 0.007
NIE		-0.014	-0.019, -0.008

CDE, controlled direct effect when psychotropic medication = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for gender, educational attainment, English-speaking birth country, education/cognition GPS, bipolar disorder GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia. Missing mediator and covariate data were imputed via a single stochastic imputation using chained equations.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

AA – Equivalence class for mania/bipolar DAG



Summary diagram for all DAGs that are structurally equivalent to the DAG used in the analyses

Bold lines indicate paths that can be reversed while still maintaining the same predicted independencies.

BB – Testable independency results for major depression DAG

Summary of residual partial correlations from alternative versions of the major depression DAG:

DAG description	**Number of implied independencies**	***n* (%) of results in each correlation coefficient range**							**Largest correlation coefficient**
		|0.00| to |0.09|	**|0.10| to |0.19|**	**|0.20| to |0.29|**	**|0.30| to |0.39|**	**|0.40| to |0.49|**	**|0.50| to |0.59|**	**|0.60| to |0.69|**	
DAG with education as antecedent (edu→major_depression) (edu→other_neupsy)	137	116 (84.7)	16 (11.7)	5 (3.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.2248 curr_psych_meds ⊥ famhx_depression | adult_SES age child_adversity edu_attain gender major_depression majordep_geno other_neupsy
DAG with education as consequence (edu←major_depression) (edu←other_neupsy)	137	115 (83.9)	17 (12.4)	5 (3.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.2663 curr_psych_meds ⊥ famhx_depression | child_adversity cog_geno eng_speak ethnicity gender major_depression majordep_geno maternal_smoking other_neupsy

Optimal model fit is indicated by smaller correlation coefficients. Coefficients ≥ |0.10| may indicate model misspecification, measurement error, selection bias etc.
The first DAG above was taken forward into the causal analyses.

Results of additional regression models to obtain more details on the highest residual correlations indicated above (≥ |0.20|):

(1) curr_psych_meds ⊥ famhx_depression | adult_SES age child_adversity edu_attain gender major_depression majordep_geno other_neupsy
Odds ratio 1.45 (95% CI 1.34, 1.56)

(2) famhx_depression ⊥ sleep | child_adversity cog_geno edu_attain eng_speak ethnicity gender major_depression majordep_geno maternal_smoking other_neupsy
Odds ratio 1.20 (95% CI 1.13, 1.27)

(3) curr_psych_meds ⊥ famhx_depression | child_adversity cog_geno edu_attain eng_speak ethnicity gender major_depression majordep_geno maternal_smoking other_neupsy
Odds ratio 1.48 (95% CI 1.37, 1.60)

(4) edu_attain ⊥ famhx_pd | child_adversity cog_geno eng_speak ethnicity majordep_geno maternal_smoking
Odds ratio 0.99 (95% CI 0.90, 1.09)

(5) famhx_pd ⊥ phys_environ | child_adversity cog_geno edu_attain eng_speak ethnicity gender major_depression majordep_geno maternal_smoking other_neupsy
Odds ratio 1.09 (95% CI 0.95, 1.24)

CC – Supplementary results for major depression mediation analyses

Table CC.1 Tests of interactions between exposure and mediators

	Coefficient for major depression * mediator	95% CI	*p*
Mediator: Cardiometabolic disease			
Reasoning[a]	0.029	-0.027, 0.085	0.304
Reaction time[a]	0.053	-0.001, 0.107	0.056
Numeric memory[a]	0.057	-0.052, 0.166	0.306
Visuospatial memory[a]	0.036	-0.024, 0.095	0.241
Prospective memory[b]	1.157	0.952, 1.407	0.142
Mediator: Psychotropic medication			
Reasoning[a]	0.244	0.135, 0.354	<0.001
Reaction time[a]	0.080	-0.025, 0.185	0.134
Numeric memory[a]	0.203	0.013, 0.393	0.036
Visuospatial memory[a]	0.082	-0.037, 0.200	0.176
Prospective memory[b]	1.454	1.057, 2.000	0.021

CI, confidence interval.
All models included major depression, the mediator and their product, as well as all the covariates entered into the `gformula` mediation models.
a. Linear regression model with outcome measured in z-score units.
b. Logistic regression model with outcome measured as correct or not; estimate expressed as odds ratio.

Table CC.2 Tests of interactions between exposure and intermediate confounders

	Coefficient for major depression * deprivation	**95% CI**	***p***	**Coefficient for major depression * lifetime number of episodes of depressed mood/anhedonia**	**95% CI**	***p***
Reasoning[a]	0.018	-0.030, 0.066	0.458	-0.020	-0.074, 0.035	0.481
Reaction time[a]	-0.017	-0.064, 0.030	0.475	0.033	-0.020, 0.085	0.224
Numeric memory[a]	0.041	-0.057, 0.138	0.413	-0.002	-0.106, 0.103	0.974
Visuospatial memory[a]	0.056	0.005, 0.108	0.033	-0.023	-0.081, 0.035	0.433
Prospective memory[b]	0.989	0.836, 1.171	0.900	1.000	0.834, 1.212	0.996

CI, confidence interval.
Deprivation was entered as a dichotomous indicator for the two most deprived Townsend quintiles versus the three least deprived quintiles. Lifetime number of episodes of depressed mood/anhedonia was entered as a dichotomous indicator for ≥2 episodes versus <2 episodes. All models included the product terms indicated in the table above, as well as major depression, deprivation, lifetime number of episodes of depressed mood/anhedonia, psychotropic medication, major depression * psychotropic medication, and all the other covariates entered into the `gformula` mediation models.
a. Linear regression model with outcome measured in z-score units.
b. Logistic regression model with outcome measured as correct or not; estimate expressed as odds ratio.

Table CC.3 Mediation of the effect of major depression on cognitive outcome via cardiometabolic disease, with missing data imputation

	n	Estimate	95% CI[a]
Reasoning[b]	102,643		
TCE		0.001	-0.017, 0.019
CDE		-0.005	-0.023, 0.013
NDE		-0.009	-0.026, 0.009
NIE		0.009	-0.023, 0.013
Reaction time[b]	119,830		
TCE		-0.024	-0.042, -0.006
CDE		-0.022	-0.039, -0.004
NDE		-0.032	-0.050, -0.014
NIE		0.008	-0.002, 0.019
Numeric memory[b]	33,250		
TCE		-0.006	-0.039, 0.027
CDE		-0.013	-0.047, 0.020
NDE		-0.006	-0.039, 0.027
NIE		0.000	-0.016, 0.017
Visuospatial memory[b]	118,363		
TCE		-0.072	-0.091, -0.054
CDE		-0.074	-0.093, -0.056
NDE		-0.072	-0.091 -0.053
NIE		-0.001	-0.011, 0.010
Prospective memory[c]	104,509		
TCE		-0.004	-0.009, 0.001
CDE		-0.002	-0.007, 0.003
NDE		-0.004	-0.009, 0.002
NIE		-0.000	-0.003, 0.002

CDE, controlled direct effect when cardiometabolic disease = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; NO_2, nitrogen dioxide; PM_{10}, particulate matter of up to 10μm diameter; TCE, total causal effect.

Models were restricted to participants of white British genetic ancestry, and were adjusted for age, gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, population density, road proximity, air pollution (PM_{10} and NO_2), body mass index, alcohol frequency, smoking status, physical activity, and psychotropic medication. Missing mediator and covariate data were imputed via a single stochastic imputation using chained equations.

a. Normal-based, from bootstrapped standard error (1000 replicates).

b. Estimate expressed as a standardised mean difference.

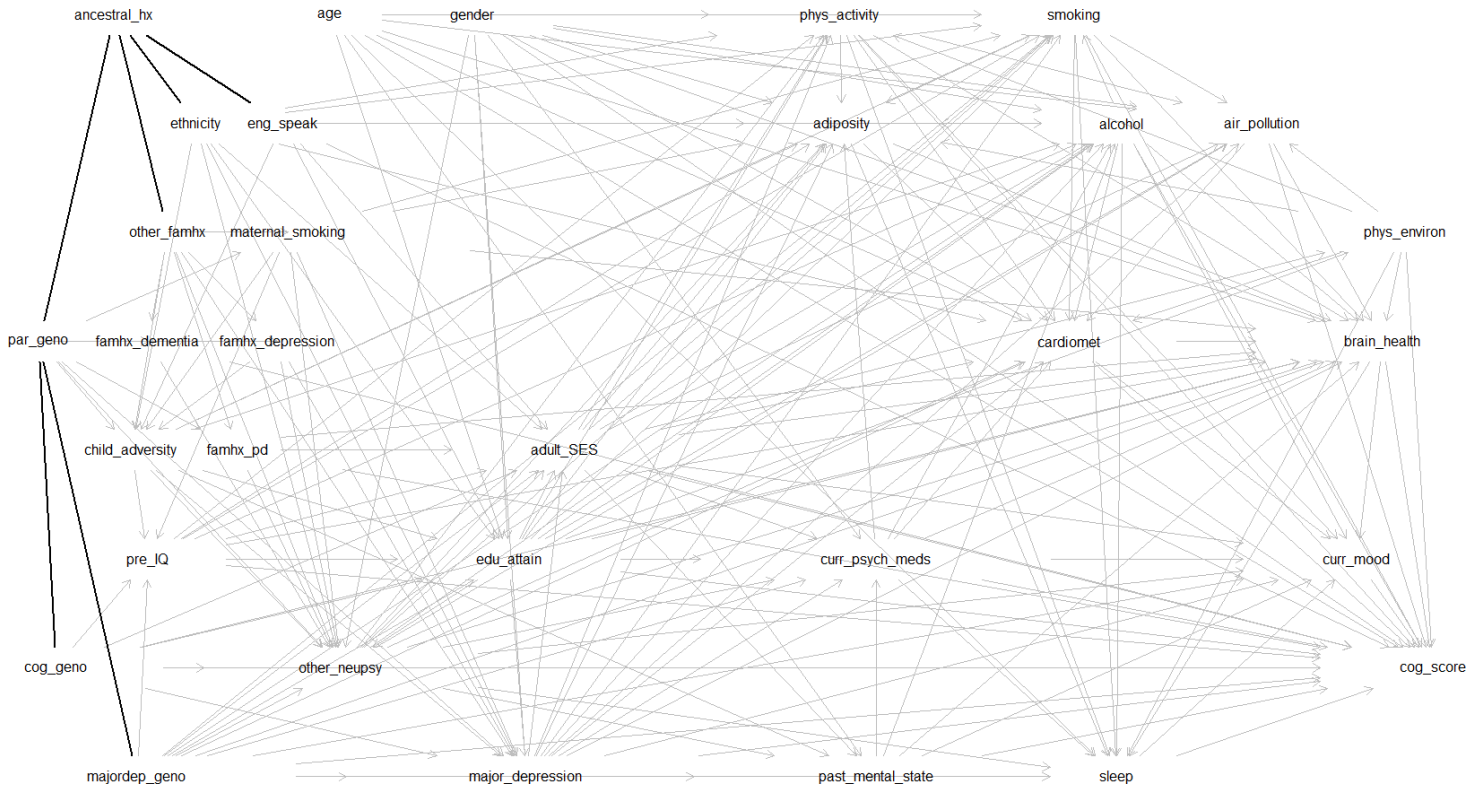
c. Estimate expressed as a risk difference for the probability of being correct.

Table CC.4 Mediation of the effect of major depression on cognitive outcome via psychotropic medication, with missing data imputation

	n	Estimate	95% CI[a]
Reasoning[b]	102,643		
TCE		-0.004	-0.022, 0.014
CDE		0.003	-0.016, 0.021
NDE		-0.010	-0.028, 0.008
NIE		0.006	-0.003, 0.015
Reaction time[b]	119,830		
TCE		-0.009	-0.026, 0.008
CDE		0.011	-0.009, 0.030
NDE		-0.008	-0.026, 0.009
NIE		-0.001	-0.011, 0.010
Numeric memory[b]	33,250		
TCE		-0.022	-0.056, 0.011
CDE		-0.001	-0.035, 0.034
NDE		-0.005	-0.040, 0.030
NIE		-0.017	-0.036, 0.002
Visuospatial memory[b]	118,363		
TCE		-0.068	-0.087, -0.049
CDE		-0.035	-0.057, -0.014
NDE		-0.056	-0.074, -0.037
NIE		-0.012	-0.023, -0.001
Prospective memory[c]	104,509		
TCE		0.002	-0.003, 0.008
CDE		0.007	0.001, 0.013
NDE		0.001	-0.004, 0.006
NIE		0.001	-0.002, 0.004

CDE, controlled direct effect when psychotropic medication = 0; CI, confidence interval; GPS, genome-wide polygenic score; NDE, natural direct effect; NIE, natural indirect effect; TCE, total causal effect.
Models were restricted to participants of white British genetic ancestry, and were adjusted for gender, educational attainment, English-speaking birth country, education/cognition GPS, major depression GPS, family history of dementia, family history of Parkinson's disease, maternal smoking around birth, childhood trauma, other psychiatric/neurological conditions, deprivation, and lifetime number of episodes of depressed mood/anhedonia. All models included a product term for major depression * psychotropic medication. Missing mediator and covariate data were imputed via a single stochastic imputation using chained equations.
a. Normal-based, from bootstrapped standard error (1000 replicates).
b. Estimate expressed as a standardised mean difference.
c. Estimate expressed as a risk difference for the probability of being correct.

DD – Equivalence class for major depression DAG



Summary diagram for all DAGs that are structurally equivalent to the DAG used in the analyses

Bold lines indicate paths that can be reversed while still maintaining the same predicted independencies.

EE – Testable independency results for schizophrenia DAG

Summary of residual partial correlations from alternative versions of the schizophrenia DAG:

DAG description	**Number of implied independencies**	***n* (%) of results in each correlation coefficient range**							**Largest correlation coefficient**
		|0.00| to |0.09|	**|0.10| to |0.19|**	**|0.20| to |0.29|**	**|0.30| to |0.39|**	**|0.40| to |0.49|**	**|0.50| to |0.59|**	**|0.60| to |0.69|**	
Original DAG with education as antecedent (edu→schizophrenia) (edu→other_neupsy)	137	109 (79.6)	18 (13.1)	4 (2.9)	4 (2.9)	2 (1.5)	0 (0.0)	0 (0.0)	0.4743 curr_psych_meds ⊥ famhx_pd | child_adversity cog_geno edu_attain eng_speak ethnicity gender schizophrenia scz_geno maternal_smoking other_neupsy
Original DAG with education as consequence (edu←schizophrenia) (edu←other_neupsy)	137	106 (77.4)	18 (13.1)	7 (5.1)	3 (2.2)	3 (2.2)	0 (0.0)	0 (0.0)	0.4743 curr_psych_meds ⊥ famhx_pd | child_adversity cog_geno eng_speak ethnicity gender schizophrenia scz_geno maternal_smoking other_neupsy
Modified DAG including cannabis node (education as per row 1 above)[a] (cannabis→schizophrenia) (cannabis→other_neupsy)	141	107 (75.9)	20 (14.2)	6 (4.3)	7 (5.0)	0 (0.0)	1 (0.7)	0 (0.0)	0.5511 curr_psych_meds ⊥ famhx_pd | adult_SES cannabis edu_attain gender other_neupsy past_mental_state schizophrenia scz_geno
Modified DAG including cannabis node (education as per row 1 above) (cannabis←schizophrenia) (cannabis←other_neupsy)	141	109 (77.3)	22 (15.6)	4 (2.8)	5 (3.5)	0 (0.0)	1 (0.7)	0 (0.0)	0.5511 curr_psych_meds ⊥ famhx_pd | adult_SES cannabis edu_attain gender other_neupsy past_mental_state schizophrenia scz_geno

Optimal model fit is indicated by smaller correlation coefficients. Coefficients ≥ |0.10| may indicate model misspecification, measurement error, selection bias etc.
The first DAG above was taken forward into the causal analyses.
a. Graph with cannabis added is shown overleaf.

Results of additional regression models to obtain more details on the highest residual correlations indicated above (≥ |0.30|):

(1) curr_psych_meds ⊥ famhx_pd | child_adversity cog_geno edu_attain eng_speak ethnicity gender schizophrenia scz_geno maternal_smoking other_neupsy
Odds ratio 1.17 (95% CI 0.97, 1.40)

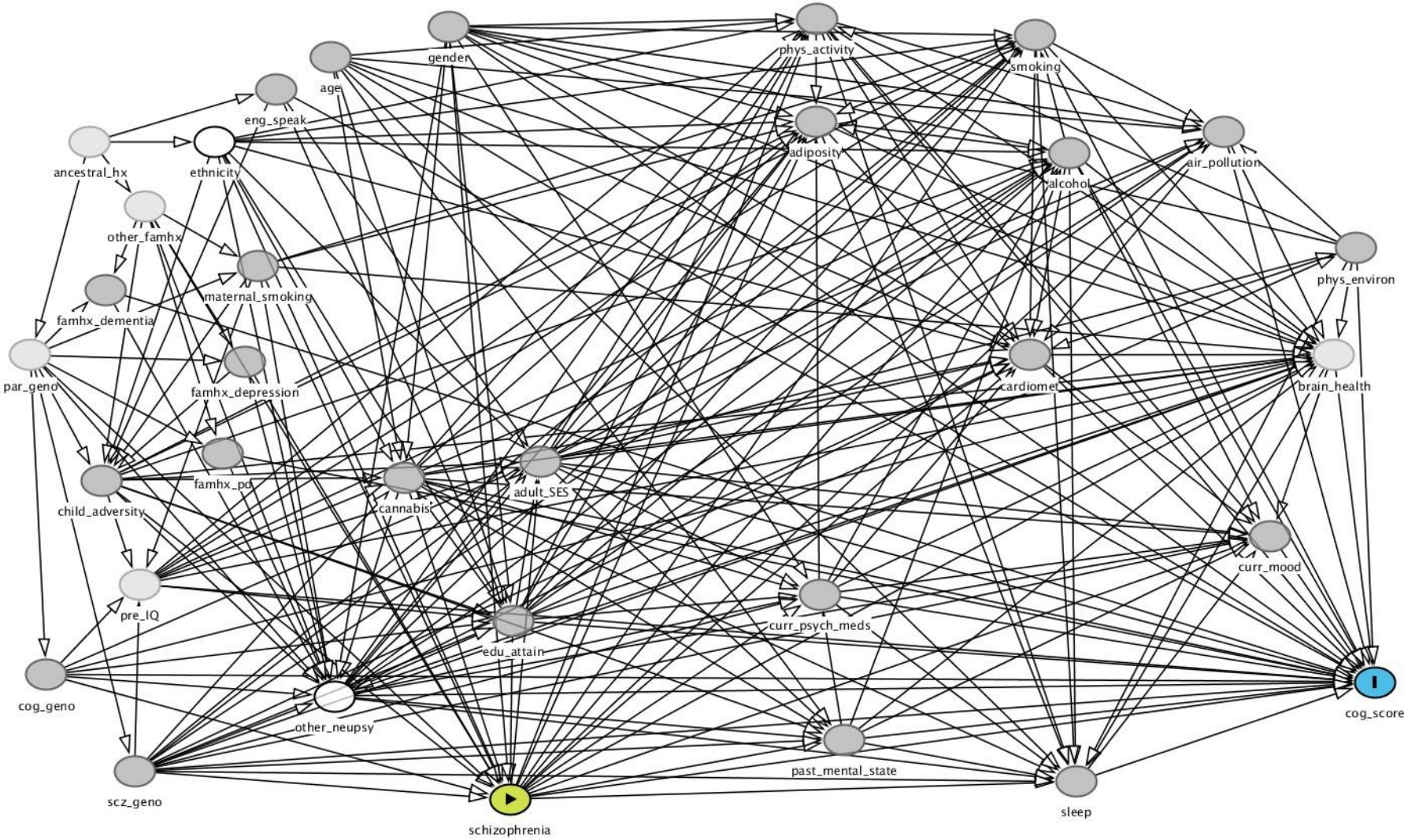
(2) curr_psych_meds ⊥ famhx_depression | child_adversity cog_geno edu_attain eng_speak ethnicity gender schizophrenia scz_geno maternal_smoking other_neupsy
Odds ratio 1.33 (95% CI 1.20, 1.48)

(3) curr_psych_meds ⊥ famhx_dementia | child_adversity cog_geno edu_attain eng_speak ethnicity gender schizophrenia scz_geno maternal_smoking other_neupsy
Odds ratio 1.17 (95% CI 1.05, 1.32)

(4) curr_psych_meds ⊥ famhx_depression | adult_SES age child_adversity edu_attain gender schizophrenia scz_geno other_neupsy
Odds ratio 1.37 (95% CI 1.24, 1.51)

(5) curr_psych_meds ⊥ famhx_depression | adult_SES edu_attain gender schizophrenia scz_geno other_neupsy past_mental_state
Odds ratio 1.26 (95% CI 1.13, 1.39)

(6) curr_psych_meds ⊥ famhx_pd | adult_SES edu_attain gender schizophrenia scz_geno other_neupsy past_mental_state
Odds ratio 1.15 (95% CI 0.97, 1.38)



Modified DAG for schizophrenia, with cannabis included

References

Aarsland, D., Zaccai, J., & Brayne, C. (2005). A systematic review of prevalence studies of dementia in Parkinson's disease. *Movement Disorders, 20*(10), 1255-1263. doi:10.1002/mds.20527

Aas, M., Henry, C., Andreassen, O. A., Bellivier, F., Melle, I., & Etain, B. (2016). The role of childhood trauma in bipolar disorders. *International Journal of Bipolar Disorders, 4*, 2. doi:10.1186/s40345-015-0042-0

Altshuler, L. L., Ventura, J., van Gorp, W. G., Green, M. F., Theberge, D. C., & Mintz, J. (2004). Neurocognitive function in clinically stable men with bipolar I disorder or schizophrenia and normal control subjects. *Biological Psychiatry, 56*(8), 560-569. doi:10.1016/j.biopsych.2004.08.002

Aminoff, S. R., Tesli, M., Bettella, F., Aas, M., Lagerberg, T. V., Djurovic, S., . . . Melle, I. (2015). Polygenic risk scores in bipolar disorder subgroups. *Journal of Affective Disorders, 183*, 310-314. doi:10.1016/j.jad.2015.05.021

Andrade, L., Caraveo-Anduaga, J. J., Berglund, P., Bijl, R. V., De Graaf, R., Vollebergh, W., . . . Wittchen, H. U. (2003). The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *International Journal of Methods in Psychiatric Research, 12*(1), 3-21. doi:10.1002/mpr.138

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455. doi:10.2307/2291629

Arslan, F. C., Tiryaki, A., & Ozkorumak, E. (2014). A comparison of euthymic bipolar patients with unaffected first-degree relatives and healthy controls in terms of neuropsychological functions. *International Journal of Psychiatry in Clinical Practice, 18*(3), 208-214. doi:10.3109/13651501.2013.859706

Arts, B., Jabben, N., Krabbendam, L., & van Os, J. (2008). Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychological Medicine, 38*(6), 771-785. doi:10.1017/s0033291707001675

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi:10.1080/00273171.2011.568786

Balanzá-Martínez, V., Selva, G., Martinez-Aran, A., Prickaerts, J., Salazar, J., Gonzalez-Pinto, A., . . . Tabares-Seisdedos, R. (2010). Neurocognition in bipolar disorders-A closer look at comorbidities and medications. *European Journal of Pharmacology, 626*(1), 87-96. doi:10.1016/j.ejphar.2009.10.018

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics, 61*(4), 962-972. doi:10.1111/j.1541-0420.2005.00377.x

Baron, R. M., & Kenny, D. A. (1986). The Moderator Mediator Variable Distinction in Social Psychological-Research - Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173

Barrera, A., Vazquez, G., Tannenhaus, L., Lolich, M., & Herbst, L. (2013). Theory of mind and functionality in bipolar patients with symptomatic remission. *Revista De Psiquiatria Y Salud Mental, 6*(2), 67-74. doi:10.1016/j.rpsm.2012.07.004

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., & Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimers & Dementia, 11*(6), 718-726. doi:10.1016/j.jalz.2015.05.016

Baune, B. T., & Malhi, G. S. (2015). A review on the impact of cognitive dysfunction on social, occupational, and general functional outcomes in bipolar disorder. *Bipolar Disorders, 17*, 41-55. doi:10.1111/bdi.12341

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., . . . Thomas, S. M. (2008). The mental wealth of nations. *Nature, 455*(7216), 1057-1060. doi:10.1038/4551057a

Benton, A. L., Hamsher, K. d., & Sivan, A. B. (1994). *Multilingual Aphasia Examination* (3rd ed.). Belford: Ann Arbor Publishers.

Bergamaschi, R. (2007). Prognostic factors in multiple sclerosis. *International Review of Neurobiology, 79*, 423-447. doi:10.1016/S0074-7742(07)79019-0

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., . . . Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect, 27*(2), 169-190. doi:10.1016/S0145-2134(02)00541-0

Bhati, M. T. (2013). Defining Psychosis: The Evolution of DSM-5 Schizophrenia Spectrum Disorders. *Current Psychiatry Reports, 15*(11). doi:10.1007/s11920-013-0409-9

Blakely, T., Lynch, J., & Bentley, R. (2016). Commentary: DAGs and the restricted potential outcomes approach are tools, not theories of causation. *International Journal of Epidemiology, 45*(6), 1835-1837. doi:10.1093/ije/dyw228

Blanco, C., Compton, W. M., Saha, T. D., Goldstein, B. I., Ruan, W. J., Huang, B. J., & Grant, B. F. (2017). Epidemiology of DSM-5 bipolar I disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions - III. *Journal of Psychiatric Research, 84*, 310-317. doi:10.1016/j.jpsychires.2016.10.003

Bobholz, J. A., & Rao, S. M. (2003). Cognitive dysfunction in multiple sclerosis: a review of recent developments. *Current Opinion in Neurology, 16*(3), 283-288. doi:10.1097/01.wco.0000073928.19076.84

Booth, M. (2000). Assessment of physical activity: An international perspective. *Research Quarterly for Exercise and Sport, 71*(2), S114-S120.

Bora, E. (2016). Differences in cognitive impairment between schizophrenia and bipolar disorder: Considering the role of heterogeneity. *Psychiatry and Clinical Neurosciences, 70*(10), 424-433. doi:10.1111/pcn.12410

Bora, E., Akdede, B. B., & Alptekin, K. (2017). Neurocognitive impairment in deficit and non-deficit schizophrenia: a meta-analysis. *Psychological Medicine, 47*(14), 2401-2413. doi:10.1017/S0033291717000952

Bora, E., Harrison, B. J., Yucel, M., & Pantelis, C. (2013). Cognitive impairment in euthymic major depressive disorder: a meta-analysis. *Psychological Medicine, 43*(10), 2017-2026. doi:10.1017/s0033291712002085

Bora, E., & Ozerdem, A. (2017). Meta-analysis of longitudinal studies of cognition in bipolar disorder: comparison with healthy controls and schizophrenia. *Psychological Medicine, 47*(16), 2753-2766. doi:10.1017/s0033291717001490

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders, 113*(1-2), 1-20. doi:10.1016/j.jad.2008.06.009

Bora, E., Yucel, M., Pantelis, C., & Berk, M. (2011). Meta-analytic review of neurocognition in bipolar II disorder. *Acta Psychiatrica Scandinavica, 123*(3), 165-174. doi:10.1111/j.1600-0447.2010.01638.x

Bortolato, B., Miskowiak, K. W., Koehler, C. A., Vieta, E., & Carvalho, A. F. (2015). Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatric Disease and Treatment, 11*, 3111-3125. doi:10.2147/ndt.s76700

Bortolato, B., Miskowiak, K. W., Kohler, C. A., Maes, M., Fernandes, B. S., Berk, M., & Carvalho, A. F. (2016). Cognitive remission: a novel objective for the treatment of major depression? *BMC Medicine, 14*(9). doi:10.1186/S12916-016-0560-3

Bostock, E. C. S., Kirkby, K. C., Garry, M. I., & Taylor, B. V. M. (2017). Systematic Review of Cognitive Function in euthymic Bipolar Disorder and Pre-Surgical Temporal Lobe epilepsy. *Frontiers in Psychiatry, 8*, 133. doi:10.3389/fpsyt.2017.00133

Bourne, C., Aydemir, O., Balanza-Martinez, V., Bora, E., Brissos, S., Cavanagh, J. T. O., . . . Goodwin, G. M. (2013). Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. *Acta Psychiatrica Scandinavica, 128*(3), 149-162. doi:10.1111/acps.12133

Bowden, J., Smith, G. D., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology, 44*(2), 512-525. doi:10.1093/ije/dyv080

Bowden, J., Smith, G. D., Haycock, P. C., & Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology, 40*(4), 304-314. doi:10.1002/gepi.21965

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., . . . Davey Smith, G. (2013). Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology, 42*(1), 111-127. doi:10.1093/ije/dys064

Brito Ferreira, M. L. (2010). Cognitive deficits in multiple sclerosis A systematic review. *Arquivos De Neuro-Psiquiatria, 68*(4), 632-641. doi:10.1590/s0004-282x2010000400029

Broadbent, A., Vandenbroucke, J. P., & Pearce, N. (2016). Response: Formalism or pluralism? A reply to commentaries on 'Causality and causal inference in epidemiology'. *International Journal of Epidemiology, 45*(6), 1841-1851. doi:10.1093/ije/dyw298

Bromet, E. J., Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., Ruggero, C., & Chang, S. W. (2011). Diagnostic Shifts During the Decade Following First Admission for Psychosis. *American Journal of Psychiatry, 168*(11), 1186-1194. doi:10.1176/appi.ajp.2011.11010048

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Sturmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*(12), 1149-1156. doi:10.1093/aje/kwj149

Buist-Bouwman, M. A., Ormel, J., de Graaf, R., de Jonge, P., van Sonderen, E., Alonso, J., . . . Investigators, E. M. (2008). Mediators of the association between depression and role functioning. *Acta Psychiatrica Scandinavica, 118*(6), 451-458. doi:10.1111/j.1600-0447.2008.01285.x

Burdick, K. E., Russo, M., Frangou, S., Mahon, K., Braga, R. J., Shanahan, M., & Malhotra, A. K. (2014). Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. *Psychological Medicine, 44*(14), 3083-3096. doi:10.1017/S0033291714000439

Burgess, S., Small, D. S., & Thompson, S. G. (2015). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*. doi:10.1177/0962280215597579

Burrell, J. R., & Piguet, O. (2015). Lifting the veil: how to use clinical neuropsychology to assess dementia. *Journal of Neurology Neurosurgery and Psychiatry, 86*(11), 1216-1224. doi:10.1136/jnnp-2013-307483

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . Marchini, J. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRXiv*. doi:10.1101/166298

Cambridge Cognition. (2017). *CANTAB Cognitive assessment software*. Cambridge: Cambridge Cognition.

Camelo, E. V. M., Velasques, B., Ribeiro, P., Netto, T., & Cheniaux, E. (2013). Attention impairment in bipolar disorder: a systematic review. *Psychology & Neuroscience, 6*(3), 299-309. doi:10.3922/j.psns.2013.3.08

Cattaneo, M. D., Drukker, D. M., & Holland, A. D. (2013). Estimation of multivalued treatment effects under conditional independence. *Stata Journal, 13*(3), 407-450.

Cavanagh, J. T. O., Van Beck, M., Muir, W., & Blackwood, D. H. R. (2002). Case-control study of neurocognitive function in euthymic patients with bipolar disorder: an association with mania. *British Journal of Psychiatry, 180*, 320-326. doi:10.1192/bjp.180.4.320

Chen, W., Shi, J., Qian, L., & Azen, S. P. (2014). Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study. *BMC Medical Research Methodology, 14*, 82. doi:10.1186/1471-2288-14-82

Cheung, E. Y. W., Halari, R., Cheng, K. M., Leung, S. K., & Young, A. H. (2013). Cognitive performance is impaired in euthymic Chinese patients with Bipolar 1 Disorder. *Journal of Affective Disorders, 151*(1), 156-163. doi:10.1016/j.jad.2013.05.070

Collins, R. (2012). What makes UK Biobank special? *Lancet, 379*(9822), 1173-1174. doi:10.1016/S0140-6736(12)60404-8

Crawford, J. R. (2012). Quantitative aspects of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians* (2nd ed., pp. 129-158). Chichester: John Wiley & Sons Ltd.

Cullen, B., Nicholl, B. I., Mackay, D. F., Martin, D., Ul-Haq, Z., McIntosh, A., . . . Smith, D. J. (2015). Cognitive function and lifetime features of depression and bipolar disorder in a large population sample: Cross-sectional study of 143,828 UK Biobank participants. *European Psychiatry, 30*(8), 950-958. doi:10.1016/j.eurpsy.2015.08.006

Daban, C., Mathieu, F., Raust, A., Cochet, B., Scott, J., Etain, B., . . . Bellivier, F. (2012). Is processing speed a valid cognitive endophenotype for bipolar disorder? *Journal of Affective Disorders, 139*(1), 98-101. doi:10.1016/j.jad.2012.02.028

Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine, 32*(9), 1584-1618. doi:10.1002/sim.5686

Daniel, R. M., De Stavola, B. L., & Cousens, S. N. (2011). gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *Stata Journal, 11*(4), 479-517.

Daniel, R. M., De Stavola, B. L., & Vansteelandt, S. (2016). Commentary: The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *International Journal of Epidemiology, 45*(6), 1817-1829. doi:10.1093/ije/dyw227

Daniel, R. M., Kenward, M. G., Cousens, S. N., & De Stavola, B. L. (2012). Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research, 21*(3), 243-256. doi:10.1177/0962280210394469

Davis, K. A., Coleman, J. R. I., Adams, M., Allen, N. E., Breen, G., Cullen, B., . . . Hotopf, M. (2018). Mental Health in UK Biobank – development, implementation and results from an online questionnaire completed by 157,366 participants. *BJPsych Open, in press*.

Davis, K. A., Sudlow, C. L., & Hotopf, M. (2016). Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of

routinely collected diagnoses. *BMC Psychiatry, 16*, 263. doi:10.1186/s12888-016-0963-x

de Leon, J., & Diaz, F. J. (2005). A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophrenia Research, 76*(2-3), 135-157. doi:10.1016/j.schres.2005.02.010

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B., & Micali, N. (2015). Mediation Analysis With Intermediate Confounding: Structural Equation Modeling Viewed Through the Causal Inference Lens. *American Journal of Epidemiology, 181*(1), 64-80. doi:10.1093/aje/kwu239

de Zeeuw, E. L., van Beijsterveldt, C. E. M., Glasner, T. J., Bartels, M., Ehli, E. A., Davies, G. E., . . . Social Science Genetic Association. (2014). Polygenic Scores Associated With Educational Attainment in Adults Predict Educational Achievement and ADHD Symptoms in Children. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 165*(6), 510-520. doi:10.1002/ajmg.b.32254

Deary, I. J. (2009). Introduction to the special issue on cognitive epidemiology. *Intelligence, 37*(6), 517-519. doi:10.1016/j.intell.2009.05.001

Deary, I. J. (2013). Intelligence. *Current Biology, 23*(16), R673-R676. doi:10.1016/j.cub.2013.07.021

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., . . . Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin, 92*(1), 135-152. doi:10.1093/bmb/ldp033

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research Methods, 43*(1), 258-268. doi:10.3758/s13428-010-0024-1

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*(3), 201-211. doi:10.1038/nrn2793

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. Oxford: Pearson Assessment.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test - Second UK Edition (CVLT-II UK)*. Oxford: Pearson Assessment.

Dendrou, C. A., Fugger, L., & Friese, M. A. (2015). Immunopathology of multiple sclerosis. *Nature Reviews Immunology, 15*(9), 545-558. doi:10.1038/nri3871

Dickinson, T., Becerra, R., & Coombes, J. (2017). Executive functioning deficits among adults with Bipolar Disorder (types I and II): A systematic review and meta-analysis. *Journal of Affective Disorders, 218*, 407-427. doi:10.1016/j.jad.2017.04.010

Diniz, B. S., Teixeira, A. L., Cao, F., Gildengers, A., Soares, J. C., Butters, M. A., & Reynolds, C. F. (2017). History of Bipolar Disorder and the Risk of Dementia: A Systematic Review and Meta-Analysis. *American Journal of Geriatric Psychiatry, 25*(4), 357-362. doi:10.1016/j.jagp.2016.11.014

Doganavsargil-Baysal, O., Cinemre, B., Aksoy, U. M., Akbas, H., Metin, O., Fettahoglu, C., . . . Davran, F. (2013). Levels of TNF-, soluble TNF receptors (sTNFR1, sTNFR2), and cognition in bipolar disorder. *Human Psychopharmacology-Clinical and Experimental, 28*(2), 160-167. doi:10.1002/hup.2301

Duarte, W., Becerra, R., & Cruise, K. (2016). The Relationship Between Neurocognitive Functioning and Occupational Functioning in Bipolar Disorder: A Literature Review. *Europe's Journal of Psychology, 12*(4), 659-678. doi:10.5964/ejop.v12i4.909

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics, 9*(3), e1003348. doi:10.1371/journal.pgen.1003348

Ebrahim, S., & Smith, G. D. (2013). Commentary: Should we always deliberately be non-representative? *International Journal of Epidemiology, 42*(4), 1022-1026. doi:10.1093/ije/dyt105

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634.

Elshahawi, H. H., Essawi, H., Rabie, M. A., Mansour, M., Beshry, Z. A., & Mansour, A. N. (2011). Cognitive functions among euthymic bipolar I patients after a single manic episode versus recurrent episodes. *Journal of Affective Disorders, 130*(1-2), 180-191. doi:10.1016/j.jad.2010.10.027

Elwert, F. (2013). Graphical causal models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 245-273). New York: Springer.

Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research, 19*(3), 237-270. doi:10.1177/0962280209105014

Emsley, R., & Liu, H. (2013). PARAMED: Stata module to perform causal mediation analysis using parametric regression models. Boston: Statistical Software Components.

Evans, V. C., Iverson, G. L., Yatham, L. N., & Lam, R. W. (2014). The Relationship Between Neurocognitive and Psychosocial Functioning in Major Depressive Disorder: A Systematic Review. *Journal of Clinical Psychiatry, 75*(12), 1359-1370. doi:10.4088/JCP.13r08939

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences, 6*(1), 21-29. doi:10.1016/0191-8869(85)90026-1

Fajutrao, L., Locklear, J., Priaulx, J., & Heyes, A. (2009). A systematic review of the evidence of the burden of bipolar disorder in Europe. *Clinical Practice and Epidemiology in Mental Health, 5*, 3. doi:10.1186/1745-0179-5-3

Fakhry, H., El Ghonemy, S. H., & Salem, A. (2013). Cognitive functions and cognitive styles in young euthymic patients with bipolar I disorder. *Journal of Affective Disorders, 151*(1), 369-377. doi:10.1016/j.jad.2013.05.095

Ferrier, I. N., Stanton, B. R., Kelly, T. P., & Scott, J. (1999). Neuropsychological function in euthymic patients with bipolar disorder. *British Journal of Psychiatry, 175*, 246-251. doi:10.1192/bjp.175.3.246

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-d.-G., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 35*(3), 573-588. doi:10.1016/j.neubiorev.2010.07.001

Fiedorowicz, J. G., He, J., & Merikangas, K. R. (2011). The association between mood and anxiety disorders with vascular diseases and risk factors in a nationally representative sample. *Journal of Psychosomatic Research, 70*(2), 145-154. doi:10.1016/j.jpsychores.2010.07.010

Fioravanti, M., Bianchi, V., & Cinti, M. E. (2012). Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence. *BMC Psychiatry, 12*. doi:10.1186/1471-244x-12-64

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition*. New York, NY: Biometrics Research, New York State Psychiatric Institute.

Fischer, M., Kunkel, A., Bublak, P., Faiss, J. H., Hoffmann, F., Sailer, M., . . . Koehler, W. (2014). How reliable is the classification of cognitive impairment across different criteria in early and late stages of multiple sclerosis? *Journal of the Neurological Sciences, 343*(1-2), 91-99. doi:10.1016/j.jns.2014.05.042

Fortier-Brochu, E., & Morin, C. M. (2014). Cognitive Impairment in Individuals with Insomnia: Clinical Significance and Correlates. *Sleep, 37*(11), 1789-1800. doi:10.5665/sleep.4172

Frangou, S., Donaldson, S., Hadjulis, M., Landau, S., & Goldstein, L. H. (2005). The Maudsley Bipolar Disorder Project: Executive dysfunction in bipolar disorder I and its clinical correlates. *Biological Psychiatry, 58*(11), 859-864. doi:10.1016/j.biopsych.2005.04.056

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189-1232. doi:10.1214/aos/1013203451

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., . . . Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with the General Population. *American Journal of Epidemiology, 186*(9), 1026-1034. doi:10.1093/aje/kwx246

Gage, S. H., Jones, H. J., Burgess, S., Bowden, J., Davey Smith, G., Zammit, S., & Munafo, M. R. (2017). Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychological Medicine, 47*(5), 971-980. doi:10.1017/S0033291716003172

Gajwani, P. (2017). Bipolar disorder in adults. *BMJ Best Practice.* Retrieved from https://bestpractice.bmj.com/topics/en-gb/488/pdf/488.pdf

Ganguli, M., Du, Y. C., Dodge, H. H., Ratcliff, G. G., & Chang, C. C. H. (2006). Depressive symptoms and cognitive decline in late life - A prospective epidemiological study. *Archives of General Psychiatry, 63*(2), 153-160. doi:10.1001/archpsyc.63.2.153

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for Constructing and Assessing Propensity Scores. *Health Services Research, 49*(5), 1701-1720. doi:10.1111/1475-6773.12182

Gelfand, L. A., Mensinger, J. L., & Tenhave, T. (2009). Mediation analysis: a retrospective snapshot of practice and more recent directions. *Journal of General Psychology, 136*(2), 153-176. doi:10.3200/GENP.136.2.153-178

Gitlin, M. J., & Miklowitz, D. J. (2017). The difficult lives of individuals with bipolar disorder: A review of functional outcomes and their implications for treatment. *Journal of Affective Disorders, 209*, 147-154. doi:10.1016/j.jad.2016.11.021

Glymour, M. M. (2006). Using causal diagrams to understand common problems in social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 387-422). San Francisco: Jossey-Bass.

Glymour, M. M., & Greenland, S. (2008). Causal diagrams. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.), *Modern Epidemiology* (3rd ed., pp. 183-209). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Golding, J. (1990). Children of the nineties. A longitudinal study of pregnancy and childhood based on the population of Avon (ALSPAC). *West of England Medical Journal, 105*(3), 80-82.

Goldman-Mellor, S., Caspi, A., Gregory, A. M., Harrington, H., Poulton, R., & Moffitt, T. E. (2015). Is Insomnia Associated with Deficits in Neuropsychological Functioning? Evidence from a Population-Based Study. *Sleep, 38*(4), 623-631. doi:10.5665/sleep.4584

Goldstein, L. H., & McNeil, J. E. (2013). *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians* (2nd ed.). Chichester, West Sussex ; Malden, MA: Wiley-Blackwell.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136*(3), 389-413. doi:10.1037/0096-3445.136.3.389

Gorman, E., Leyland, A. H., McCartney, G., Katikireddi, S. V., Rutherford, L., Graham, L., . . . Gray, L. (2017). Adjustment for survey non-representativeness using record-

linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction, 112*(7), 1270-1280. doi:10.1111/add.13797

Goswami, U., Sharma, A., Varma, A., Gulrajani, C., Ferrier, I. N., Young, A. H., . . . Moore, P. B. (2009). The neurocognitive performance of drug-free and medicated euthymic bipolar patients do not differ. *Acta Psychiatrica Scandinavica, 120*(6), 456-463. doi:10.1111/j.1600-0447.2009.01390.x

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*(4), 404-411.

Gray, L., McCartney, G., White, I. R., Katikireddi, S. V., Rutherford, L., Gorman, E., & Leyland, A. H. (2013). Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data a: a study protocol. *BMJ Open, 3*(3), e002647. doi:10.1136/bmjopen-2013-002647

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology, 10*(1), 37-48. doi:10.1097/00001648-199901000-00008

Gualtieri, C. T., & Morgan, D. W. (2008). The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: An unaccounted source of variance in clinical trials. *Journal of Clinical Psychiatry, 69*(7), 1122-1130.

Guimaraes, J., & Sa, M. J. (2012). Cognitive dysfunction in multiple sclerosis. *Frontiers in Neurology, 3*, 74-74. doi:10.3389/fneur.2012.00074

Guo, S., & Fraser, M. W. (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications.

Hafeman, D. M., & Schwartz, S. (2009). Opening the Black Box: a motivation for the assessment of mediation. *International Journal of Epidemiology, 38*(3), 838-845. doi:10.1093/ije/dyn372

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J., . . . Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry, 21*(11), 1624-1632. doi:10.1038/mp.2015.225

Harvey, A. G., Schmidt, D. A., Scarna, A., Semler, C. N., & Goodwin, G. M. (2005). Sleep-related functioning in euthymic patients with bipolar disorder, patients with insomnia, and subjects without sleep problems. *American Journal of Psychiatry, 162*(1), 50-57. doi:10.1176/appi.ajp.162.1.50

Harvey, P. D., Wingo, A. P., Burdick, K. E., & Baldessarini, R. J. (2010). Cognition and disability in bipolar disorder: lessons from schizophrenia research. *Bipolar Disorders, 12*(4), 364-375. doi:10.1111/j.1399-5618.2010.00831.x

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York: The Guilford Press.

Hernán, M. A. (2005). Invited commentary: Hypothetical interventions to define causal effects - Afterthought or prerequisite? *American Journal of Epidemiology, 162*(7), 618-620. doi:10.1093/aje/kwi255

Hernán, M. A. (2016). Does water kill? A call for less casual causal inferences. *Annals of Epidemiology, 26*(10), 674-680. doi:10.1016/j.annepidem.2016.08.016

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health, 60*(7), 578-586. doi:10.1136/jech.2004.029496

Hernán, M. A., & Robins, J. M. (2017). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming.

Hernán, M. A., & Taubman, S. L. (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity, 32*, S8-S14. doi:10.1038/ijo.2008.82

Hill, W. D., Davies, G., Liewald, D. C., McIntosh, A. M., Deary, I. J., & CHARGE Cognitive Working Group. (2016). Age-Dependent Pleiotropy Between General

Cognitive Function and Major Psychiatric Disorders. *Biological Psychiatry, 80*(4), 266-273. doi:10.1016/j.biopsych.2015.08.033

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., . . . Deary, I. J. (2016). Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Current Biology, 26*(22), 3083-3089. doi:10.1016/j.cub.2016.09.035

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., . . . Deary, I. J. (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry, [Epub ahead of print]*. doi:10.1038/s41380-017-0001-5

Hindle, J. V., Petrelli, A., Clare, L., & Kalbe, E. (2013). Nonpharmacological enhancement of cognitive function in Parkinson's disease: A systematic review. *Movement Disorders, 28*(8), 1034-1049. doi:10.1002/mds.25377

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960. doi:10.2307/2289064

Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ, 340*, c2289. doi:10.1136/bmj.c2289

Ibrahim, N., Rahman, A. H. A., & Shah, S. A. (2009). Verbal memory test performance in patients with bipolar I disorder attending a psychiatric clinic of a university hospital in Kuala Lumpur, Malaysia. *ASEAN Journal of Psychiatry, 10*(2), 157-168.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry, 167*(7), 748-751. doi:10.1176/appi.ajp.2010.09091379

Jamrozinski, K., Gruber, O., Kemmer, C., Falkai, P., & Scherk, H. (2009). Neurocognitive functions in euthymic bipolar patients. *Acta Psychiatrica Scandinavica, 119*(5), 365-374. doi:10.1111/j.1600-0447.2008.01320.x

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence, 32*(1), 95-107. doi:10.1016/S0160-2896(03)00062-X

Judd, C. M., & Kenny, D. A. (1981). Process Analysis - Estimating Mediation in Treatment Evaluations. *Evaluation Review, 5*(5), 602-619. doi:10.1177/0193841x8100500502

Juselius, S., Kieseppa, T., Kaprio, J., Lonnqvist, J., & Tuulio-Henriksson, A. (2009). Executive Functioning in Twins with Bipolar I Disorder and Healthy Co-Twins. *Archives of Clinical Neuropsychology, 24*(6), 599-606. doi:10.1093/arclin/acp047

Kahn, R. S., & Keefe, R. S. E. (2013). Schizophrenia Is a Cognitive Illness Time for a Change in Focus. *JAMA Psychiatry, 70*(10), 1107-1112. doi:10.1001/jamapsychiatry.2013.155

Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *Lancet, 386*(9996), 896-912. doi:10.1016/S0140-6736(14)61393-3

Kamm, C. P., Uitdehaag, B. M., & Polman, C. H. (2014). Multiple sclerosis: current knowledge and future outlook. *European Neurology, 72*(3-4), 132-141. doi:10.1159/000360528

Kaufman, J. S. (2016). There is no virtue in vagueness. Comment on: Causal Identification: A Charge of Epidemiology in Danger of Marginalization by Sharon Schwartz, Nicolle M. Gatto, and Ulka B. Campbell. *Annals of Epidemiology, 26*(10), 683-684. doi:10.1016/j.annepidem.2016.08.018

Kaufman, J. S., Kaufman, S., & Poole, C. (2003). Causal inference from randomized trials in social epidemiology. *Social Science & Medicine, 57*(12), 2397-2409. doi:10.1016/S0277-9536(03)00135-7

Keefe, R. S. E. (2008). Should cognitive impairment be included in the diagnostic criteria for schizophrenia? *World Psychiatry, 7*(1), 22-28.

Keefe, R. S. E., & Fenton, W. S. (2007). How should DSM-V criteria for schizophrenia include cognitive impairment? *Schizophrenia Bulletin, 33*(4), 912-920. doi:10.1093/schbul/sbm046

Kemp, D. E., & Fan, J. B. (2012). Cardiometabolic Health in Bipolar Disorder. *Psychiatric Annals, 42*(5), 179-183. doi:10.3928/00485713-20120507-05

Kendler, K. S. (2017). Causal Inference in Psychiatric Epidemiology. *Jama Psychiatry, 74*(6), 561-562. doi:10.1001/jamapsychiatry.2017.0502

Kessing, L. V. (1998). Cognitive impairment in the euthymic phase of affective disorder. *Psychological Medicine, 28*(5), 1027-1038. doi:10.1017/s0033291798006862

Keyes, K. M., & Galea, S. (2017). The Limits of Risk Factors Revisited. Is It Time for a Causal Architecture Approach? *Epidemiology, 28*(1), 1-5. doi:10.1097/Ede.0000000000000578

Kieseppa, T., Tuulio-Henriksson, A., Haukka, J., Van Erp, T., Glahn, D., Cannon, T. D., . . . Lonnqvist, J. (2005). Memory and verbal learning functions in twins with bipolar-I disorder, and the role of information-processing speed. *Psychological Medicine, 35*(2), 205-215. doi:10.1017/s0033291704003125

Killin, L. O. J., Starr, J. M., Shiue, I. J., & Russ, T. C. (2016). Environmental risk factors for dementia: a systematic review. *BMC Geriatrics, 16*, 175. doi:10.1186/s12877-016-0342-y

Kirkbride, J. B., Errazuriz, A., Croudace, T. J., Morgan, C., Jackson, D., Boydell, J., . . . Jones, P. B. (2012). Incidence of Schizophrenia and Other Psychoses in England, 1950-2009: A Systematic Review and Meta-Analyses. *PloS ONE, 7*(3), e31660. doi:10.1371/journal.pone.0031660

Klein, J. D. (1997). The National Longitudinal Study on Adolescent Health. Preliminary results: great expectations. *JAMA, 278*(10), 864-865.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: The Guilford Press.

Kluwe-Schiavon, B., Sanvicente-Vieira, B., Kristensen, C. H., & Grassi-Oliveira, R. (2013). Executive functions rehabilitation for schizophrenia: A critical systematic review. *Journal of Psychiatric Research, 47*(1), 91-104. doi:10.1016/j.jpsychires.2012.10.001

Kluwe-Schiavon, B., Viola, T. W., Levandowski, M. L., Bortolotto, V. R., Azevedo e Souza, L. S., Tractenberg, S. G., & Soares, T. (2015). A systematic review of cognitive rehabilitation for bipolar disorder. *Trends in Psychiatry and Psychotherapy, 37*(4), 194-201. doi:10.1590/2237-6089-2015-0006

Knopman, D., Boland, L. L., Mosley, T., Howard, G., Liao, D., Szklo, M., . . . ARIC Study Investigators. (2001). Cardiovascular risk factors and cognitive decline in middle-aged adults. *Neurology, 56*(1), 42-48. doi:10.1212/Wnl.56.1.42

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J. B., Rimfeld, K., von Stumm, S., . . . Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry, 21*(9), 1188-1193. doi:10.1038/mp.2015.126

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry, 21*(3), 437-443. doi:10.1038/mp.2015.2

Krieger, N., & Smith, G. D. (2016a). Response: FACEing reality: productive tensions between our epidemiological questions, methods and mission. *International Journal of Epidemiology, 45*(6), 1852-1865. doi:10.1093/ije/dyw330

Krieger, N., & Smith, G. D. (2016b). The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology, 45*(6), 1787-1808. doi:10.1093/ije/dyw114

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613.

Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer.

Laursen, T. M., Agerbo, E., & Pedersen, C. B. (2009). Bipolar Disorder, Schizoaffective Disorder, and Schizophrenia Overlap: A New Comorbidity Index. *Journal of Clinical Psychiatry, 70*(10), 1432-1438. doi:10.4088/JCP.08m04807

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Smith, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine, 27*(8), 1133-1163. doi:10.1002/sim.3034

Lawlor, D. A., Tilling, K., & Smith, G. D. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology, 45*(6), 1866-1886. doi:10.1093/ije/dyw314

Le Hellard, S., & Steen, V. M. (2014). Genetic architecture of cognitive traits. *Scandinavian Journal of Psychology, 55*(3), 255-262. doi:10.1111/sjop.12112

Lee, R. S. C., Hermens, D. F., Scott, J., Redoblado-Hodge, M. A., Naismith, S. L., Lagopoulos, J., . . . Hickie, I. B. (2014). A meta-analysis of neuropsychological functioning in first-episode bipolar disorders. *Journal of Psychiatric Research, 57*, 1-11. doi:10.1016/j.jpsychires.2014.06.019

Lepage, M., Bodnar, M., & Bowie, C. R. (2014). Neurocognition: Clinical and Functional Outcomes in Schizophrenia. *Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie, 59*(1), 5-12.

Leung, I. H. K., Walton, C. C., Hallock, H., Lewis, S. J. G., Valenzuela, M., & Lampit, A. (2015). Cognitive training in Parkinson disease A systematic review and meta-analysis. *Neurology, 85*(21), 1843-1851. doi:10.1212/wnl.0000000000002145

Lim, C. S., Baldessarini, R. J., Vieta, E., Yucel, M., Bora, E., & Sim, K. (2013). Longitudinal neuroimaging and neuropsychological changes in bipolar disorder patients: Review of the evidence. *Neuroscience and Biobehavioral Reviews, 37*(3), 418-435. doi:10.1016/j.neubiorev.2013.01.003

Linden, A., & Karlson, K. B. (2013). Using mediation analysis to identify causal mechanisms in disease management interventions. *Health Services and Outcomes Research Methodology, 13*, 86-108. doi:10.1007/s10742-013-0106-5

Litvan, I., Aarsland, D., Adler, C. H., Goldman, J. G., Kulisevsky, J., Mollenhauer, B., . . . Weintraub, D. (2011). MDS Task Force on Mild Cognitive Impairment in Parkinson's Disease: Critical Review of PD-MCI. *Movement Disorders, 26*(10), 1814-1824. doi:10.1002/mds.23823

Lloyd, T., Kennedy, N., Fearon, P., Kirkbride, J., Mallett, R., Leff, J., . . . AESOP Study Team. (2005). Incidence of bipolar affective disorder in three UK cities: results from the AESOP study. *British Journal of Psychiatry, 186*, 126-131. doi:10.1192/bjp.186.2.126

Lopera-Vasquez, J., Bell, V., & López-Jaramillo, C. (2011). What is the contribution of executive dysfunction to the cognitive profile of bipolar disorder? A well-controlled direct comparison study. *Revista Colombiana de Psiquiatría, 40*(Suppl 1), 64S-75S.

Lopez-Jaramillo, C., Lopera-Vasquez, J., Gallo, A., Ospina-Duque, J., Bell, V., Torrent, C., . . . Vieta, E. (2010). Effects of recurrence on the cognitive performance of patients with bipolar I disorder: implications for relapse prevention and treatment adherence. *Bipolar Disorders, 12*(5), 557-567. doi:10.1111/j.1399-5618.2010.00835.x

Lovera, J., & Kovner, B. (2012). Cognitive impairment in multiple sclerosis. *Current Neurology and Neuroscience Reports, 12*(5), 618-627. doi:10.1007/s11910-012-0294-3

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., . . . Pell, J. P. (2016). Cognitive Test Scores in UK Biobank: Data Reduction in 480,416

Participants and Longitudinal Stability in 20,346 Participants. *PLoS ONE, 11*(4), e0154222. doi:10.1371/journal.pone.0154222

Lyall, D. M., Lopez, L. M., Bastin, M. E., Maniega, S. M., Penke, L., Valdes Hernandez, M. d. C., . . . Deary, I. J. (2013). ADRB2, brain white matter integrity and cognitive ageing in the Lothian Birth Cohort 1936. *Behavior Genetics, 43*(1), 13-23. doi:10.1007/s10519-012-9570-x

Lyu, J., & Burr, J. A. (2016). Socioeconomic Status Across the Life Course and Cognitive Function Among Older Adults: An Examination of the Latency, Pathways, and Accumulation Hypotheses. *Journal of Aging and Health, 28*(1), 40-67. doi:10.1177/0898264315585504

MacDonald, B. K., Cockerell, O. C., Sander, J. W. A. S., & Shorvon, S. D. (2000). The incidence and lifetime prevalence of neurological disorders in a prospective community-based study in the UK. *Brain, 123*, 665-676. doi:10.1093/brain/123.4.665

MacKinnon, D. F. (2017). Depression in adults. *BMJ Best Practice.* Retrieved from http://bestpractice.bmj.com/topics/en-gb/55/pdf/55.pdf

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical Approaches for Enhancing Causal Interpretation of the M to Y Relation in Mediation Analysis. *Personality and Social Psychology Review, 19*(1), 30-43. doi:10.1177/1088868314542878

Mann-Wrobel, M. C., Carreno, J. T., & Dickinson, D. (2011). Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. *Bipolar Disorders, 13*(4), 334-342. doi:10.1111/j.1399-5618.2011.00935.x

Manolio, T. A., Weis, B. K., Cowie, C. C., Hoover, R. N., Hudson, K., Kramer, B. S., . . . Collins, F. S. (2012). New Models for Large Prospective Studies: Is There a Better Way? *American Journal of Epidemiology, 175*(9), 859-866. doi:10.1093/aje/kwr453

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., . . . Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence, 44*, 26-32. doi:10.1016/j.intell.2014.02.006

Martin, D. J., Ul-Haq, Z., Nicholl, B. I., Cullen, B., Evans, J., Gill, J. M. R., . . . Smith, D. J. (2016). Cardiometabolic disease and features of depression and bipolar disorder: population-based, cross-sectional study. *British Journal of Psychiatry, 208*(4), 343-351. doi:10.1192/bjp.bp.114.157784

Martino, D. J., Igoa, A., Marengo, E., Scapola, M., & Strejilevich, S. A. (2011a). Neurocognitive Impairments and Their Relationship With Psychosocial Functioning in Euthymic Bipolar II Disorder. *Journal of Nervous and Mental Disease, 199*(7), 459-464. doi:10.1097/NMD.0b013e3182214190

Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011b). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research, 189*(3), 379-384. doi:10.1016/j.psychres.2011.04.033

Martino, D. J., Strejilevich, S. A., Marengo, E., Ibanez, A., Scapola, M., & Igoa, A. (2014). Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. *Journal of Affective Disorders, 167*, 118-124. doi:10.1016/j.jad.2014.05.059

Martino, D. J., Strejilevich, S. A., Scapola, M., Igoa, A., Marengo, E., Ais, E. D., & Perinot, L. (2008). Heterogeneity in cognitive functioning among patients with bipolar disorder. *Journal of Affective Disorders, 109*(1-2), 149-156. doi:10.1016/j.jad.2007.12.232

Martino, D. J., Strejilevich, S. A., Torralva, T., & Manes, F. (2011c). Decision making in euthymic bipolar I and bipolar II disorders. *Psychological Medicine, 41*(6), 1319-1327. doi:10.1017/s0033291710001832

Matthews, P. M., Roncaroli, F., Waldman, A., Sormani, M. P., De Stefano, N., Giovannoni, G., & Reynolds, R. (2016). A practical review of the neuropathology and neuroimaging of multiple sclerosis. *Practical Neurology, 16*(4), 279-287. doi:10.1136/practneurol-2016-001381

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9*(4), 403-425. doi:10.1037/1082-989x.9.4.403

McDonald, S., Flanagan, S., & Rollins, J. (2002). *The Awareness of Social Inference Test (TASIT)*. Oxford: Pearson Assessment.

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews, 30*(1), 67-76. doi:10.1093/epirev/mxn001

McIntosh, A. M., Stewart, R., John, A., Smith, D. J., Davis, K., Sudlow, C., . . . Group, M. Q. D. S. (2016). Data science for mental health: a UK perspective on a global challenge. *Lancet Psychiatry, 3*(10), 993-998. doi:10.1016/S2215-0366(16)30089-X

McNamee, R. (2003). Confounding and confounders. *Occupational and Environmental Medicine, 60*(3), 227-234. doi:10.1136/oem.60.3.227

Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., . . . Zarkov, Z. (2011). Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Archives of General Psychiatry, 68*(3), 241-251. doi:10.1001/archgenpsychiatry.2011.12

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*(1), 37-45. doi:10.1016/j.tics.2013.10.011

Miskowiak, K. W., Carvalho, A. F., Vieta, E., & Kessing, L. V. (2016). Cognitive enhancement treatments for bipolar disorder: A systematic review and methodological recommendations. *European Neuropsychopharmacology, 26*(10), 1541-1561. doi:10.1016/j.euroneuro.2016.08.011

Miskowiak, K. W., Kjaerstad, H. L., Meluken, I., Petersen, J. Z., Maciel, B. R., Kohler, C. A., . . . Carvalho, A. F. (2017). The search for neuroimaging and cognitive endophenotypes: A critical systematic review of studies involving unaffected first-degree relatives of individuals with bipolar disorder. *Neuroscience and Biobehavioral Reviews, 73*, 1-22. doi:10.1016/j.neubiorev.2016.12.011

Miskowiak, K. W., Ott, C. V., Petersen, J. Z., & Kessing, L. V. (2016). Systematic review of randomized controlled trials of candidate treatments for cognitive impairment in depression and methodological challenges in the field. *European Neuropsychopharmacology, 26*(12), 1845-1867. doi:10.1016/j.euroneuro.2016.09.641

Mitchell, A. J., Kemp, S., Benito-Leon, J., & Reuber, M. (2010). The influence of cognitive impairment on health-related quality of life in neurological disease. *Acta Neuropsychiatrica, 22*(1), 2-13. doi:10.1111/j.1601-5215.2009.00439.x

Mitchell, K. J. (2012). What is complex about complex disorders? *Genome Biology, 13*(1), 237. doi:10.1186/gb-2012-13-1-237

Mitolo, M., Venneri, A., Wilkinson, I. D., & Sharrack, B. (2015). Cognitive rehabilitation in multiple sclerosis: A systematic review. *Journal of the Neurological Sciences, 354*(1-2), 1-9. doi:10.1016/j.jns.2015.05.004

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology, 62*(10), 1006-1012. doi:10.1016/j.jclinepi.2009.06.005

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1. doi:10.1186/2046-4053-4-1

Mortensen, P. B., Pedersen, C. B., Melbye, M., Mors, O., & Ewald, H. (2003). Individual and familial risk factors for bipolar affective disorders in Denmark. *Archives of General Psychiatry, 60*(12), 1209-1215. doi:10.1001/archpsyc.60.12.1209

MRC. (2014). *Maximising the value of UK population cohorts*. Swindon: Medical Research Council.

Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. E., & Smith, G. D. (2017). Collider scope: when selection bias can substantially influence observed associations. *International Journal of Epidemiology, 47*(1), 226-235. doi:10.1093/ije/dyx206

Munn, Z., Moola, S., Riitano, D., & Lisy, K. (2014). The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *International Journal of Health Policy Management, 3*(3), 123-128. doi:10.15171/ijhpm.2014.71

Mur, M., Portella, M. J., Martinez-Aran, A., Pifarre, J., & Vieta, E. (2007). Persistent neuropsychological deficit in euthymic bipolar patients: Executive function as a core deficit. *Journal of Clinical Psychiatry, 68*(7), 1078-1086.

Muthén, B. O., & Asparouhov, T. (2015). Causal Effects in Mediation Modeling: An Introduction With Applications to Latent Variables. *Structural Equation Modeling-a Multidisciplinary Journal, 22*(1), 12-23. doi:10.1080/10705511.2014.935843

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and Mediation Analysis using Mplus*. Los Angeles: Muthén & Muthén.

Neyman, J. S. (1990[1923]). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science, 5*, 465-480.

NICE. (2011). Common mental health problems: identification and pathways to care (CG123). Retrieved from https://www.nice.org.uk/guidance/cg123

Normala, I., Abdul, H. A. R., Azlin, B., Nik Ruzyanei, N. J., Hazli, Z., & Shah, S. A. (2010). Executive function and attention span in euthymic patients with bipolar 1 disorder. *The Medical Journal of Malaysia, 65*(3), 199-203.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . LifeLines Cohort Study. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542. doi:10.1038/nature17671

Osher, Y., Dobron, A., Belmaker, R. H., Bersudsky, Y., & Dwolatzky, T. (2011). Computerized Testing of Neurocognitive Function in Euthymic Bipolar Patients Compared to Those with Mild Cognitive Impairment and Cognitively Healthy Controls. *Psychotherapy and Psychosomatics, 80*(5), 298-303. doi:10.1159/000324508

Pagano, G., Rengo, G., Pasqualetti, G., Femminella, G. D., Monzani, F., Ferrara, N., & Tagliati, M. (2015). Cholinesterase inhibitors for Parkinson's disease: a systematic review and meta-analysis. *Journal of Neurology Neurosurgery and Psychiatry, 86*(7), 767-773. doi:10.1136/jnnp-2014-308764

Parker, G., & Brotchie, H. (2010). Gender differences in depression. *International Review of Psychiatry, 22*(5), 429-436. doi:10.3109/09540261.2010.492391

Patti, F., Nicoletti, A., Messina, S., Bruno, E., Lo Fermo, S., Quattrocchi, G., . . . Zappia, M. (2015). Prevalence and incidence of cognitive impairment in multiple sclerosis: a population-based survey in Catania, Sicily. *Journal of Neurology, 262*(4), 923-930. doi:10.1007/s00415-015-7661-3

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence, 29*(3), 241-288. doi:10.1016/0004-3702(86)90072-X

Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research, 27*(2), 226-284. doi:10.1177/0049124198027002004

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge: Cambridge University Press.

Pearl, J. (2012). The Causal Mediation Formula-A Guide to the Assessment of Pathways and Mechanisms. *Prevention Science, 13*(4), 426-436. doi:10.1007/s11121-011-0270-1

Pearl, J. (2014). Interpretation and Identification of Causal Mediation. *Psychological Methods, 19*(4), 459-481. doi:10.1037/a0036434

Pearl, J., & Bareinboim, E. (2014). External Validity: From Do-Calculus to Transportability Across Populations. *Statistical Science, 29*(4), 579-595. doi:10.1214/14-Sts486

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Chichester, West Sussex: Wiley.

Penke, L., Maniega, S. M., Bastin, M. E., Hernandez, M. C. V., Murray, C., Royle, N. A., . . . Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry, 17*(10), 1026-1030. doi:10.1038/mp.2012.66

Petersen, M. L., Sinisi, S. E., & van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology, 17*(3), 276-284. doi:10.1097/01.ede.0000208475.99429.2d

Picchioni, M. M., & Murray, R. M. (2007). Schizophrenia. *British Medical Journal, 335*(7610), 91-95. doi:10.1136/bmj.39227.616447.BE

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *British Journal of Psychiatry, 177*, 486-492. doi:10.1192/bjp.177.6.486

Pirkola, T., Tuulio-Henriksson, A., Glahn, D., Kieseppa, T., Haukka, J., Kaprio, J., . . . Cannon, T. D. (2005). Spatial working memory function in twins with schizophrenia and bipolar disorder. *Biological Psychiatry, 58*(12), 930-936. doi:10.1016/j.biopsych.2005.05.041

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20*(1), 98-108. doi:10.1038/mp.2014.105

Plomin, R., & Simpson, M. A. (2013). The future of genomics for developmentalists. *Development and Psychopathology, 25*(4 Pt 2), 1263-1278. doi:10.1017/S0954579413000606

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148-159. doi:10.1038/nrg.2017.104

Poewe, W. (2008). Non-motor symptoms in Parkinson's disease. *European Journal of Neurology, 15 Suppl 1*, 14-20. doi:10.1111/j.1468-1331.2008.02056.x

Power, M. C., Adar, S. D., Yanosky, J. D., & Weuve, J. (2016). Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: A systematic review of epidemiologic research. *Neurotoxicology, 56*, 235-253. doi:10.1016/j.neuro.2016.06.004

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904-909. doi:10.1038/ng1847

Prins, N. D., van Dijk, E. J., den Heijer, T., Vermeer, S. E., Jolles, J., Koudstaal, P. J., . . . Breteler, M. M. B. (2005). Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain, 128*, 2034-2041. doi:10.1093/brain/awh553

Qiu, C., & Fratiglioni, L. (2015). A major role for cardiovascular burden in age-related cognitive decline. *Nature Reviews Cardiology, 12*(5), 267-277. doi:10.1038/nrcardio.2014.223

Rait, G., Walters, K., Griffin, M., Buszewicz, M., Petersen, I., & Nazareth, I. (2009). Recent trends in the incidence of recorded depression in primary care. *British Journal of Psychiatry, 195*(6), 520-524. doi:10.1192/bjp.bp.108.058636

Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive Dysfunction in Multiple-Sclerosis. 1. Frequency, Patterns, and Prediction. *Neurology, 41*(5), 685-691.

Rao, S. M., Leo, G. J., Ellington, L., Nauertz, T., Bernardin, L., & Unverzagt, F. (1991). Cognitive Dysfunction in Multiple-Sclerosis. 2. Impact on Employment and Social Functioning. *Neurology, 41*(5), 692-696.

Rao, S. M., Martin, A. L., Huelin, R., Wissinger, E., Khankhel, Z., Kim, E., & Fahrbach, K. (2014). Correlations between MRI and Information Processing Speed in MS: A Meta-Analysis. *Multiple Sclerosis International, 2014*, 975803. doi:10.1155/2014/975803

Raucher-Chene, D., Achim, A. M., Kaladjian, A., & Besche-Richard, C. (2017). Verbal fluency in bipolar disorders: A systematic review and meta-analysis. *Journal of Affective Disorders, 207*, 359-366. doi:10.1016/j.jad.2016.09.039

Reichenberg, A., Harvey, P. D., Bowie, C. R., Mojtabai, R., Rabinowitz, J., Heaton, R. K., & Bromet, E. (2009). Neuropsychological Function and Dysfunction in Schizophrenia and Psychotic Affective Disorders. *Schizophrenia Bulletin, 35*(5), 1022-1029. doi:DOI 10.1093/schbul/sbn044

Revell, E. R., Neill, J. C., Harte, M., Khan, Z., & Drake, R. J. (2015). A systematic review and meta-analysis of cognitive remediation in early schizophrenia. *Schizophrenia Research, 168*(1-2), 213-222. doi:10.1016/j.schres.2015.08.017

Richards, B., Rewilak, D., Kaplan, E., Proulx, G. B., & Leach, L. (2000). *Kaplan Baycrest Neurocognitive Assessment (KBNA UK)*. Oxford: Pearson Assessment.

Richiardi, L., Bellocco, R., & Zugna, D. (2013). Mediation analysis in epidemiology: methods, interpretation and bias. *International Journal of Epidemiology, 42*(5), 1511-1519. doi:10.1093/ije/dyt127

Robertson, I. H., Nimmo-Smith, I., Ward, T., & Ridgeway, V. (1994). *Test of Everyday Attention (TEA)*. Oxford: Pearson Assessment.

Robins, J. M. (1986). A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period - Application to Control of the Healthy Worker Survivor Effect. *Mathematical Modelling, 7*(9-12), 1393-1512. doi:10.1016/0270-0255(86)90088-6

Robins, J. M., & Greenland, S. (1992). Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology, 3*(2), 143-155. doi:10.1097/00001648-199203000-00013

Robinson, L. J., & Ferrier, I. N. (2006). Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. *Bipolar Disorders, 8*(2), 103-116. doi:10.1111/j.1399-5618.2006.00277.x

Robinson, L. J., Thompson, J. M., Gallagher, P., Goswami, U., Young, A. H., Ferrier, I. N., & Moore, P. B. (2006). A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders, 93*(1-3), 105-115. doi:10.1016/j.jad.2006.02.016

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029-2040. doi:10.1017/S0033291713002535

Rosenbaum, P. R. (2002). *Observational Studies* (2nd ed.). New York: Springer.

Rosenblat, J. D., Kakar, R., & McIntyre, R. S. (2015). The Cognitive Effects of Antidepressants in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *The International Journal of Neuropsychopharmacology, 19*(2). doi:10.1093/ijnp/pyv082

Rossor, M., & Knapp, M. (2015). Can we model a cognitive footprint of interventions and policies to help to meet the global challenge of dementia? *Lancet, 386*(9997), 1008-1010. doi:10.1016/S0140-6736(15)60248-3

Rothman, K. J. (1976). Causes. *American Journal of Epidemiology, 104*(6), 587-592.

Rothman, K. J., Greenland, S., Poole, C., & Lash, T. L. (2008). Causation and causal inference. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.), *Modern*

Epidemiology (3rd ed., pp. 5-31). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701. doi:10.1037/h0037350

Rubin, D. B. (1986). Statistics and Causal Inference - Which Ifs Have Causal Answers. *Journal of the American Statistical Association, 81*(396), 961-962. doi:10.2307/2289065

Salagre, E., Sole, B., Tomioka, Y., Fernandes, B. S., Hidalgo-Mazzei, D., Garriga, M., . . . Grande, I. (2017). Treatment of neurocognitive symptoms in unipolar depression: A systematic review and future perspectives. *Journal of Affective Disorders, 221*, 205-221. doi:10.1016/j.jad.2017.06.034

Samamé, C., Martino, D. J., & Strejilevich, S. A. (2014). Longitudinal course of cognitive deficits in bipolar disorder: A meta-analytic study. *Journal of Affective Disorders, 164*, 130-138. doi:10.1016/j.jad.2014.04.028

Samamé, C., Martino, D. J., & Strejilevich, S. A. (2015). An individual task meta-analysis of social cognition in euthymic bipolar disorders. *Journal of Affective Disorders, 173*, 146-153. doi:10.1016/j.jad.2014.10.055

Sánchez-Morla, E. M., Barabash, A., Martinez-Vizcaino, V., Tabares-Seisdedos, R., Balanza-Martinez, V., Cabranes-Diaz, J. A., . . . Santos Gomez, J. L. (2009). Comparative study of neurocognitive function in euthymic bipolar patients and stabilized schizophrenic patients. *Psychiatry Research, 169*(3), 220-228. doi:10.1016/j.psychres.2008.06.032

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*(7510), 421-427. doi:10.1038/nature13595

Schoenberg, M. R., Lange, R. T., Marsh, P., & Saklofske, D. H. (2011). Premorbid Intelligence. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 2004-2010). New York, NY: Springer.

Schuler, M. S., & Rose, S. (2017). Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *American Journal of Epidemiology, 185*(1), 65-73. doi:10.1093/aje/kww165

Schwartz, S., Gatto, N. M., & Campbell, U. B. (2012). Extending the sufficient component cause model to describe the Stable Unit Treatment Value Assumption (SUTVA). *Epidemiologic Perspectives & Innovations, 9*, 3. doi:10.1186/1742-5573-9-3

Schwartz, S., Gatto, N. M., & Campbell, U. B. (2016). Causal identification: a charge of epidemiology in danger of marginalization. *Annals of Epidemiology, 26*(10), 669-673. doi:10.1016/j.annepidem.2016.03.013

Scott, E. M., Hermens, D. F., White, D., Naismith, S. L., GeHue, J., Whitwell, B. G., . . . Hickie, I. B. (2015). Body mass, cardiovascular risk and metabolic characteristics of young persons presenting for mental healthcare in Sydney, Australia. *BMJ Open, 5*(3), e007066. doi:10.1136/bmjopen-2014-007066

Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry, 22*(2), 267-272. doi:10.1038/mp.2016.107

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., . . . Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics, 43*(10), 977-983. doi:10.1038/ng.943

Smith, A. (1982). *Symbol Digit Modalities Test (SDMT)*. Torrance, CA: WPS.

Smith, D. J., Griffiths, E., Kelly, M., Hood, K., Craddock, N., & Simpson, S. A. (2011). Unrecognised bipolar disorder in primary care patients with depression. *British Journal of Psychiatry, 199*(1), 49-56. doi:10.1192/bjp.bp.110.083840

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., . . . Pell, J. P. (2013). Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. *PLoS ONE, 8*(11), e75362. doi:10.1371/journal.pone.0075362

Solé, B., Jimenez, E., Martinez-Aran, A., & Vieta, E. (2015). Cognition as a target in major depression: New developments. *European Neuropsychopharmacology, 25*(2), 231-247. doi:10.1016/j.euroneuro.2014.12.004

Solé, B., Jimenez, E., Torrent, C., Reinares, M., Del Mar Bonnin, C., Torres, I., . . . Vieta, E. (2017). Cognitive Impairment in Bipolar Disorder: Treatment and Prevention Strategies. *The International Journal of Neuropsychopharmacology*. doi:10.1093/ijnp/pyx032

Sparding, T., Silander, K., Palsson, E., Ostlind, J., Ekman, C. J., Sellgren, C. M., . . . Landen, M. (2017). Classification of cognitive performance in bipolar disorder. *Cognitive Neuropsychiatry, 22*(5), 407-421. doi:10.1080/13546805.2017.1361391

Sparding, T., Silander, K., Palsson, E., Ostlind, J., Sellgren, C., Ekman, C. J., . . . Landen, M. (2015). Cognitive Functioning in Clinically Stable Patients with Bipolar Disorder I and II. *PLoS ONE, 10*(1), e0115562. doi:10.1371/journal.pone.0115562

StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.

StataCorp. (2017). Stata Statistical Software: Release 15. College Station, TX: StataCorp LP.

Stewart, R., & Davis, K. (2016). 'Big data' in mental health research: current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology, 51*(8), 1055-1072. doi:10.1007/s00127-016-1266-8

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford: Oxford University Press.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103. doi:10.1207/S15327752JPA8001_18

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., . . . Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology - A proposal for reporting. *JAMA-Journal of the American Medical Association, 283*(15), 2008-2012. doi:10.1001/jama.283.15.2008

Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society Series a-Statistics in Society, 174*, 369-386. doi:10.1111/j.1467-985X.2010.00673.x

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology, 53*, 401-433. doi:10.1146/annurev.psych.53.100901.135220

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine, 12*(3), e1001779. doi:10.1371/journal.pmed.1001779

Svenningsson, P., Westman, E., Ballard, C., & Aarsland, D. (2012). Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. *Lancet Neurology, 11*(8), 697-707. doi:10.1016/S1474-4422(12)70152-7

Swanson, J. M. (2012). The UK Biobank and selection bias. *Lancet, 380*(9837), 110. doi:10.1016/S0140-6736(12)61179-9

Szmulewicz, A. G., Samame, C., Martino, D. J., & Strejilevich, S. A. (2015). An updated review on the neuropsychological profile of subjects with bipolar disorder. *Archives of Clinical Psychiatry, 42*(5), 139-146. doi:10.1590/0101-60830000000064

Szmulewicz, A. G., Valerio, M. P., Smith, J. M., Samame, C., Martino, D. J., & Strejilevich, S. A. (2017). Neuropsychological profiles of major depressive disorder and bipolar disorder during euthymia. A systematic literature review of comparative studies. *Psychiatry Research, 248*, 127-133. doi:10.1016/j.psychres.2016.12.031

Talati, A., Bao, Y., Kaufman, J., Shen, L., Schaefer, C. A., & Brown, A. S. (2013). Maternal smoking during pregnancy and bipolar disorder in offspring. *American Journal of Psychiatry, 170*(10), 1178-1185. doi:10.1176/appi.ajp.2013.12121500

Textor, J., Hardt, J., & Knuppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology, 22*(5), 745. doi:10.1097/EDE.0b013e318225c2be

Textor, J., van der Zander, B., Gilthorpe, M. S., Liskiewicz, M., & Ellison, G. T. H. (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology, 45*(6), 1887-1894. doi:10.1093/ije/dyw341

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*(2), 203-214. doi:10.1016/S0887-6177(03)00039-8

Townsend, P. (1987). Deprivation. *Journal of Social Policy, 16*, 125-146.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test (SNST)*. Lutz, FL: PAR, Inc.

Trotta, A., Murray, R. M., & MacCabe, J. H. (2015). Do premorbid and post-onset cognitive functioning differ between schizophrenia and bipolar disorder? A systematic review and meta-analysis. *Psychological Medicine, 45*(2), 381-394. doi:10.1017/s0033291714001512

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., . . . Ritchie, M. D. (2011). Quality control procedures for genome-wide association studies. *Current Protocols in Human Genetics, Chapter 1*, Unit1 19. doi:10.1002/0471142905.hg0119s68

Vakil, E. (2012). Neuropsychological assessment: Principles, rationale, and challenges. *Journal of Clinical and Experimental Neuropsychology, 34*(2), 135-150. doi:10.1080/13803395.2011.623121

van der Werf-Eldering, M. J., Burger, H., Holthausen, E. A. E., Aleman, A., & Nolen, W. A. (2010). Cognitive Functioning in Patients with Bipolar Disorder: Association with Depressive Symptoms and Alcohol Use. *PLoS ONE, 5*(9), e13032. doi:10.1371/journal.pone.0013032

Vandenbroucke, J. P., Broadbent, A., & Pearce, N. (2016). Causality and causal inference in epidemiology: the need for a pluralistic approach. *International Journal of Epidemiology, 45*(6), 1776-1786. doi:10.1093/ije/dyv341

VanderWeele, T. J. (2016a). Commentary: On Causes, Causal Inference, and Potential Outcomes. *International Journal of Epidemiology, 45*(6), 1809-1816. doi:10.1093/ije/dyw230

VanderWeele, T. J. (2016b). Mediation Analysis: A Practitioner's Guide. *Annual Review of Public Health, 37*, 17-32. doi:10.1146/annurev-publhealth-032315-021402

VanderWeele, T. J., & Hernán, M. A. (2012). Results on Differential and Dependent Measurement Error of the Exposure and the Outcome Using Signed Directed Acyclic Graphs. *American Journal of Epidemiology, 175*(12), 1303-1310. doi:10.1093/aje/kwr458

VanderWeele, T. J., Hernán, M. A., Tchetgen, E. J. T., & Robins, J. M. (2016). Re: Causality and causal inference in epidemiology: the need for a pluralistic approach. *International Journal of Epidemiology, 45*(6), 2199-2200. doi:10.1093/ije/dyw162

VanderWeele, T. J., Valeri, L., & Ogburn, E. L. (2012). The Role of Measurement Error and Misclassification in Mediation Analysis. *Epidemiology, 23*(4), 561-564. doi:10.1097/EDE.0b013e318258f5e4

VanderWeele, T. J., & Vansteelandt, S. (2014). Mediation Analysis with Multiple Mediators. *Epidemiologic Methods, 2*(1), 95-115. doi:10.1515/em-2012-0010

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder. *Epidemiology, 25*(2), 300-306. doi:10.1097/Ede.0000000000000034

Vansteelandt, S., & Daniel, R. M. (2017). Interventional Effects for Mediation Analysis with Multiple Mediators. *Epidemiology, 28*(2), 258-265. doi:10.1097/Ede.0000000000000596

Vienneau, D., de Hoogh, K., Bechle, M. J., Beelen, R., van Donkelaar, A., Martin, R. V., . . . Marshall, J. D. (2013). Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO2 and PM10. *Environmental Science & Technology, 47*(23), 13555-13564. doi:10.1021/es403089q

Vohringer, P. A., Barroilhet, S. A., Amerio, A., Reale, M. L., Alvear, K., Vergne, D., & Ghaemi, S. N. (2013). Cognitive impairment in bipolar disorder and schizophrenia: a systematic review. *Frontiers in Psychiatry, 4*, 87-87. doi:10.3389/fpsyt.2013.00087

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology, 18*(6), 800-804. doi:10.1097/EDE.0b013e3181577654

Walker, E. A., Gelfand, A., Katon, W. J., Koss, M. P., Von Korff, M., Bernstein, D., & Russo, J. (1999). Adult health status of women with histories of childhood abuse and neglect. *American Journal of Medicine, 107*(4), 332-339. doi:10.1016/S0002-9343(99)00235-1

Warrington, E. K., & James, M. (1991). *Visual Object and Space Perception Battery (VOSP)*. Oxford: Pearson Assessment.

Wechsler, D. (2010). *Wechsler Adult Intelligence Scale – Fourth UK edition (WAIS-IV UK)*. London, UK: Pearson Assessment.

Wechsler, D. (2011). *Wechsler Memory Scale – Fourth UK edition (WMS-IV UK).* London: Pearson Assessment.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*, S54-S64. doi:10.1212/WNL.0b013e3182872ded

Westreich, D. (2012). Berkson's Bias, Selection Bias, and Missing Data. *Epidemiology, 23*(1), 159-164. doi:10.1097/EDE.0b013e31823b6296

Westreich, D., & Cole, S. R. (2010). Invited Commentary: Positivity in Practice. *American Journal of Epidemiology, 171*(6), 674-677. doi:10.1093/aje/kwp436

Westreich, D., Edwards, J. K., Lesko, C. R., Stuart, E., & Cole, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology, 186*(8), 1010-1014. doi:10.1093/aje/kwx164

Westreich, D., & Greenland, S. (2013). The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology, 177*(4), 292-298. doi:10.1093/aje/kws412

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399. doi:10.1002/sim.4067

Wilson, B. A., Emslie, H., Evans, J. J., Alderman, N., & Burgess, P. W. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Oxford: Pearson Assessment.

Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach* (6th ed.). Mason, OH: South-Western Cengage Learning.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, 55*(10), 1068-1087. doi:10.1111/jcpp.12295

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Lewis, C. M., . . . Sullivan, P. F. (2017). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRXiv [preprint]*. doi:10.1101/167577

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A Meta-Analysis of Cognitive Remediation for Schizophrenia: Methodology and Effect Sizes. *American Journal of Psychiatry, 168*(5), 472-485. doi:10.1176/appi.ajp.2010.10060855

Yaghootkar, H., Bancks, M. P., Jones, S. E., McDaid, A., Beaumont, R., Donnelly, L., . . . Kutalik, Z. (2017). Quantifying the extent to which index event biases influence large genetic association studies. *Human Molecular Genetics, 26*(5), 1018-1030. doi:10.1093/hmg/ddw433