Understanding prognosis independent of treatment for adults with depression in primary care

Joshua Eusty Jonathan Buckman

A thesis submitted for the degree of Doctor of Philosophy to the Division of Psychology & Language Sciences, University College London (UCL)

Declaration

‘I, Joshua Eusty Jonathan Buckman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.’

Signature:

Date: 17.04.2020

Abstract

This thesis details the rationale and methods for compiling a large individual patient data (IPD) dataset of adults treated for depression in primary care, for the purpose of identifying predictors of prognosis independent of treatment. I present a narrative review of the knowledge of factors associated with prognosis for adults with depression, a narrative review of the merits and difficulties of utilising IPD data, and a protocol for a series of systematic reviews with IPD meta-analyses seeking to meet the above aim. I then present three such IPD meta-analyses. These focus on the associations between depressed patient's pre-treatment characteristics, and i) prognosis independent of a range of treatments for depression in primary care, and ii) attrition from treatment. The first of these IPD meta-analyses is centred on depressive symptom severity and a group of factors that are associated with the degree of severity of a patient's experience of depression, but which are separate from depressive symptoms, (I refer to these as indicators of depressive 'disorder severity'). The second is centred on perceived social support. The third is centred on adverse life events in the six months prior to starting treatment, socio-demographics (age, gender, ethnicity, marital status, employment status, financial wellbeing, housing tenancy, and the highest level of educational attainment), and comorbid long-term physical health conditions. The thesis finishes with a critical review of the implications of these analyses, consideration of future directions and implications for clinical practice.

Impact Statement

This thesis has several potential benefits, most are described within the body of the thesis. Three examples of such benefits are:
i) The development of a large-scale individual patient data dataset which was created to conduct the analyses presented here. This dataset will be available as a resource for further analyses on topics which are not covered in this thesis, and with necessary permissions in place.
ii) The thesis has helped highlight a number of patient characteristics, present prior to treatment, which are associated with clinically important differences in prognosis, independent of treatment. These factors can be assessed with single questions or questionnaire measures which are brief and easy to administer, they might therefore be added to prognostic studies in future without greatly increasing the burden on participants or researchers.
iii) The work presented here also helped identify patient characteristics associated with attrition from treatment. Assessing for these factors pre-treatment may further aid in the clinical management of depression in primary care.

The impact of the work presented in this thesis was enhanced through the publication of a number of papers in peer reviewed academic journals and presentations at a number of research and clinical conferences, workshops and seminars. These are detailed in the Dissemination section below.

There are a number of potential clinical benefits of the work presented here. For example: (i) the finding that it is important to consider ‘disorder severity’ factors alongside depressive symptoms to better understand prognosis might inform the assessment of depressed patients pre-treatment; ii) the finding that social support is associated with outcomes from treatment and single self-report items measuring particular aspects of social support can be informative without the need to use longer scales, might further inform assessment for some depressed patients; and iii) finding that victims of crime or those with debt problems, people not in employment, those not married, and those with poorer socio-economic functioning are all at greater risk

of poor prognoses might highlight factors to assess for and potentially those to consider as treatment targets.

Acknowledgements

I would like to acknowledge the invaluable advice, support and guidance of my supervisors and Fellowship sponsors: Professors Stephen Pilling and Glyn Lewis. The extremely valuable support and opportunities for learning given to me by the collaborators who helped me gain funding for the Fellowship which led to this PhD thesis: Professors Robert DeRubeis, Simon Gilbody, Steven Hollon, Tony Kendrick, and Edward Watkins, as well as the incredibly hard working and supportive project collaborators that have helped me realise this thesis and the related work to that presented here: Drs Gareth Ambler, Sally Brabyn, Katherine Clarke, Eiko Fried, Elizabeth Littlewood; Professors David Kessler, Dave Richards; Christopher Salisbury, Deborah Sharp, Ian White, and Nicola Wiles; Ms Phoebe Barnett, Ms Laura-Louise Arundell, Ms Nomsa Chari, and Ms Iyinoluwa Oshinowo.

I would particularly like to single out three collaborators that have made a huge contribution to the work presented here, providing me with a methodological sounding board, technical guidance, and solutions to the many data problems that arose when compiling and analysing the dataset that is utilised in the studies presented in this thesis: Drs Rob Saunders, Ciaran O'Driscoll and Zachary Cohen. Without these three, the work would have been infinitely harder and probably would not have been completed at all.

I would also like to thank the UCL Research Department of Clinical, Educational and Health Psychology and in particular the staff therein including Professors Peter Fonagy, and Pasco Fearon, Drs John Cape, Josh Stott, and Will Mandy who have acted as mentors and sources of support throughout, and Mr Adam Gudalajtys, Mrs Julia Curl and Ms Verity Armstrong who have helped me navigate the grant management and HR systems to sustain me at UCL and keep me functioning on a day-to-day basis.

My wife Sarah Buckman and our two children, Eloise and Cora, all of whom have enabled me to keep going despite the many setbacks and hurdles along the way, and have been remarkably optimistic and tolerant of this endeavour.

I would also particularly like to thank the Wellcome Trust for the opportunity they have given me to conduct this work, funding me as part of a Fellowship which has provided an excellent training in the research methods and skills required to conduct the work presented here and indeed in managing and organising such research efforts; grant reference number 201292/Z/16/Z. In addition, the studies that make up the Dep-GP IPD database were funded by:

- AHEAD: Health Technology Assessment programme of the National Health Service Research and Development Directorate.
- CADET: UK Medical Research Council (MRC; reference G0701013), managed by the National Institute for Health Research (NIHR) on behalf of the MRC-NIHR partnership
- COBALT: This research was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) programme (project number 06/404/02).
- GENPOD: Medical Research Council and supported by the Mental Health Research Network.
- HEALTHLINES: NIHR under its Programme Grant for Applied Research (Grant Reference Number RP-PG-0108-10011).
- IPCRESS: BUPA Foundation
- ITAS: Primary Secondary Care Interface initiative of the United Kingdom National Health Service Research and Development Programme (PSI 2-58).
- MIR: National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project 11/129/76) and supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol
- PANDA: NIHR Programme Grant for Applied Research (RP-PG-0610-10048).
- REEACT: UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project 06/43/05).
- RESPOND: HTA programme as project number 02/07/04
- TREAD: National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme.

Contribution

I designed, conceptualised, conducted analyses, and wrote up this project based on an initial outline used to secure a Wellcome Trust Fellowship (see acknowledgments). I have overseen all ethical approval and the collection or collation of all data for the project. In order to collect or collate the data I created data sharing agreements, applied to clinical trials units for access to data they hold, created collaborator agreements, and where necessary created or facilitated the creation of research contracts between UCL (my host institution) and institutions acting as custodians of the individual study datasets collated as part of this project. In addition, I provided data and support for an MSc student to conduct an analysis of the association between social support and treatment outcomes in a subset of the studies I used to conduct a larger, similar analysis in this thesis. That MSc project had different research questions, different data, a set of different social support exposure variables, and a different outcome variable. The MSc student also used a different method for the analyses to that presented here. Therefore, these projects were unique and although some of the data overlapped, the MSc student's project makes up no part of this thesis.

Dissemination

Below is a list of papers published or under review based on research conducted as part of this thesis (they are reported or referred to in the body of this thesis):

1) Buckman JEJ, Saunders R, Cohen ZD, *et al.* What factors indicate prognosis for adults with depression in primary care? A protocol for meta-analyses of individual patient data using the Dep-GP database. *Wellcome Open Res* 2019; **4**: 69.
2) Buckman JEJ, Saunders R, Cohen ZD, Barnett P, Clarke K, Ambler G, et al. Indicators of Prognosis Independent of Treatment for Adults with Depression in Primary Care, Going Beyond Baseline Symptom-Severity: A Systematic Review and Individual Patient Data Meta-Analysis. SSRN Electron J [Internet]. 2020; Available from: https://www.ssrn.com/abstract=3520082
3) Buckman JEJ, Saunders R, O'Driscoll C, Cohen ZD, Stott J, Ambler G, et al. Is social support pre-treatment associated with prognosis for adults with depression in primary care? Acta Psychiatr Scand. 2020.
4) Buckman JEJ, Saunders R, Arundell L-L, Oshinowo I, Cohen ZD, O'Driscoll C, et al. Associations between life events and prognosis for adults seeking treatment for depression: a systematic review and individual patient data meta-analysis. *Submitted.* British J Psychiatry.
5) Buckman JEJ, Underwood A, Clarke K, Saunders R, Hollon SD, Fearon P, et al. Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. Clin Psychol Rev. 2018; 64(7):13–38.
6) Buckman JEJ, Cohen ZD, O'Driscoll C, Fried EI, Saunders R, Ambler G, et al. Can the use of symptom heterogeneity give rise to better prognostic predictions for adults with depression ? Open Sci Framew. [Internet]. 2020; 1-32. Available from: 10.31219/osf.io/xkwdc
7) O'Driscoll C, Buckman JEJ, Fried EI, Saunders R, Cohen ZD, Ambler G, et al. The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults : analysis of data from six randomized control trials . Open Sci Framew. [Internet]. 2020;1–21. Available from: 10.31219/osf.io/kwjmn

Conference Presentations:

1) Buckman JEJ, Maintaining Recovery After Depression: Predicting and Preventing Relapses. Conference Symposium at the British Association of Behavioural and Cognitive Psychotherapies 47th Annual Conference 2019; Bath, UK
2) Buckman JEJ, Cohen ZD, Delgadillo J & DeRubeis, JD. Results of the Stratified Medicine Approaches foR Treatment Selection (SMART) Mental Health Prediction Tournament. MQ Mental Health Science Meeting, 2019; London, UK
3) Buckman JEJ, What works for whom under what circumstances? A few examples of 21st Century approaches. Campaign to End Loneliness Conference, 2019; London, UK
4) Buckman JEJ, Predicting prognosis for adults with depression using IPD of UK primary care RCTs. Treatment Selection Ideas Lab Conference, 2018; London, UK
5) Buckman JEJ, Underwood A, Clarke K, *et al.* Risk Factors for Relapse and Recurrence of Depression in Adults and How They Operate: A Four-Phase Systematic Review and Meta-Synthesis. Poster presented at Treatment Selection Ideas Lab Conference, 2018; London, UK
6) O'Driscoll C, & Buckman JEJ. A multisite network model of psychopathology within adults seeking treatment for depression. Poster presented at Treatment Selection Ideas Lab Conference, 2018; London, UK. Awarded 1st Prize in poster competition.

Contents

Figures

Tables

Chapter 1: Introduction

Overview

The purpose of this chapter is to provide an introduction to and the rationale for the aims of the thesis, and in conjunction with Chapter 2, to set out the rationale for the empirical work presented in Chapters 3 to 5. Key concepts will be defined and discussed including: depression; personalized medicine; prognosis, and in particular the distinction between different ways of considering prognosis in relation to treatment; and the use of individual patient datasets. In addition, this chapter will present a narrative meta-review of systematic reviews and meta-analyses discussing the extant knowledge on indicators of prognosis (irrespective of the context in which it was studied) for adults with depression. This review highlights the central role of symptom severity and posits that greater knowledge of prognosis may be gained by broadening the focus on severity to include related factors. Here I refer this broadened construct as depressive 'disorder severity', and will return to it throughout the thesis. This chapter will end by outlining the aims of this thesis as a whole, focussing on the need for greater knowledge of indicators of prognosis over and above symptom severity.

The diagnostic classification of depression

In diagnostic systems depression is classified as a mood or affective disorder. Two diagnostic and classification systems are commonly used: the Diagnostic and Statistics Manual of mental disorders, of which the latest is the 5th edition (1); and the International Classification of Diseases, of which the latest is the 11th edition (2). These systems largely overlap in the way that they guide clinicians to establish a diagnosis of depression: they require that a patient has had two consecutive weeks or more of experiencing 1) low mood or 2) anhedonia (which is a loss of interest or pleasure in activities), leading to withdrawal from everyday life or diminished activity. They then require the patient to have had at least five of the following symptoms: 1) reduced self-esteem or self-confidence; 2) feeling bad about oneself, feeling like a failure, feelings of worthlessness, guilt or of letting oneself or others down; 3) bleak or pessimistic views of the future or hopelessness; 4) thoughts of life not being worth living or of suicide; 5) reduced concentration or attention; 6) disturbed sleep (insomnia or hypersomnia); 7) fatigue, tiredness, or a lack of energy; 8) disturbed appetite; 9) unintended weight loss or gain of 5% or more of one's body weight, and 10) psychomotor disturbance (either more agitated or slowed down/psychomotor retardation). For a diagnosis of major depression in either classification system, the above symptoms must occur nearly every day and must impact functioning in everyday life in social, occupational or educational domains. Other symptoms such as anxiety may also be present, as may some manic or psychotic symptoms, although these latter two would change the diagnosis to one of a Major Depressive Episode with mixed features or with psychotic features respectively (3). For the remainder of this thesis unless otherwise stated, when discussing depression I refer to major depression. That is whether or not a patient would meet criteria for the diagnosis in DSM-5, whether a clinician would be likely to give the diagnosis following ICD-11 guidelines, or in both systems, given the approximately analogous status of the disorder across the classification systems (4).

The need to reduce the burden of depression

Depression is among the most burdensome diseases world-wide (5–7). It is highly prevalent with one in twenty adults (8) or up to 320 million people (9) across the globe suffering from an episode of major depression (8) or a depressive disorder

(including persistent mild depression) (9) every year. Lifetime prevalence estimates give a clearer picture on the pervasive nature of depression, with some estimates suggesting that just over 40% of people will have had at least one episode of depression by the time they are in their mid-thirties (10). Depression has a large impact on functioning both during episodes and between them; it is widely regarded as having a 'relapsing-remitting' course (11), with disabling sub-syndromal symptoms periodically occurring in between discrete episodes (12). Depression is associated with a greater risk of early mortality from physical health conditions and from suicide (6,7). Most episodes of depression last 20 weeks or more (13) and up-to 80% of patients treated for depression will have a recurrence after their episode has remitted (14). The impact of depression on daily life has a knock-on effect on employment and absenteeism from work (15), with one report suggesting that depression will result in approximately £9.2 billion of lost earnings per year in the UK alone by the year 2026, and direct costs in health spending of approximately £3 billion a year (16). Depression is therefore a major public health concern both within the UK (17) and across the world (18–20).

Treatments for depression

A variety of treatments can be helpful for depression including: pharmacotherapies such as selective serotonin reuptake inhibitors (SSRIs), tricyclic antidepressants (TCAs), and monoamine reuptake inhibitors (MAOIs); psychological interventions such as cognitive behaviour therapy (CBT), interpersonal psychotherapy (IPT), behavioural activation (BA), and behavioural couples therapy (BCT); low-intensity psychological interventions based on CBT and psychoeducation such as guided self-help or computerised CBT; structured exercise; and other treatments such as electroconvulsive therapy (ECT), or repetitive transcranial magnetic stimulation (TMS) (21). For the majority of patients with depression, particularly those with mild-to-moderate presentations, many of the treatments are considered to be approximately equally affective: low-intensity psychological interventions, formalised or high-intensity psychological interventions, and the pharmacotherapies are all recommended with the suggestion that low-intensity psychological interventions or structured exercise should be the first-line treatment (21). For those that do not respond to low-intensity psychological interventions or structured exercise then the

recommendation by the National Institute for Health and Care Excellence (NICE) is that an antidepressant (usually an SSRI as these are considered approximately equally affective as other types of antidepressant and are generally better tolerated by patients and have fewer side-effects than other types) or a high-intensity psychological intervention is given, and again, the evidence suggests that specific types of these treatments are approximately equally effective (both within the therapy modality – e.g. SSRIs and TCAs, or CBT and IPT, and across the therapy modalities i.e. antidepressants and high intensity psychological interventions). For those with moderate-to-severe presentations the recommendation is that patients receive a combination of an antidepressant and a high-intensity psychological intervention (CBT or IPT).

The NICE guidelines provide clear recommendations on the format of treatment, how it should be delivered, at what dose, how long for, and for how many sessions (in the case of low or high intensity psychological interventions) according to a summary of the research evidence (21). However, only 16.5% of people with depression world-wide receive even the minimum recommended treatment, with the vast majority of depressed adults receiving no treatment at all (8). This is a pattern mirrored in both high and low income countries; only 29% of those that screen positive for depression in the USA receive any treatment (22).

The majority of those treated for depression in the UK present to general practitioners (GPs) (23), where pharmacotherapies, particularly antidepressant medications are by far the commonest treatment. There is recognition that screening, assessment, and treatment in primary care or by a general physicians (where patients can seek an appointment or consultation directly without having to be referred by another health professional) is essential for the clinical management of depression in many other countries around the world too (8,22). For example, in the USA, recommendations from the US Preventive Services Task Force propose that to support the effectiveness of treatments for depression, there is a need to integrate behavioural health services within primary care (24), as this is where many people initially seek support and large proportions are treated, even if treatment is delivered by specialist teams for patients with more severe presentations (22). Nearly seven million adults or 15.6% of the adult population were written at least one

prescription for an antidepressant by a GP in the UK in the year 2016/17 (25). The rollout of the Improving Access to Psychological Therapies (IAPT) programme across England has contributed to an increase in the uptake of psychological treatments (particularly low-intensity treatments) for depression and other common mental disorders (26), up from 24% in 2007 to 37% in 2014. However, the majority of depressed adults still do not receive any treatment, and the burden of depression in England has not lessened (23), despite a cost of nearly £11billion per year on all treatments for depression (27). The majority of depressed patients will not reach remission with their first treatment; it will take a number of trials of different treatments before they remit, and many will dropout if the first treatment trialled is not successful (28,29). Not reaching full remission is one of the strongest predictors of relapse and recurrence of depression (30).

Given the prevalence of depression, its course, and the limited treatment resources, efforts to lessen the burden of depression are crucial. However, these are hampered by a lack of knowledge of who will and who will not get better, whether given any treatment or none, irrespective of what treatment they are given, independent of treatment, and which of the many treatments available is likely to be most beneficial for any given individual (31). These concepts are discussed further below. It is noteworthy that not only will approximately 30% of patients not recover despite many different treatments trialled (28,29), a further 40% undergo many different treatments over a long period of time before they reach recovery, perhaps unnecessarily extending their episode of depression and in some cases inflicting unnecessary side-effects (32). This can be extremely costly for the individual, those supporting them, and the healthcare services and systems providing treatment (16). In order to reduce the burden of depression it is imperative that we understand more about what confers risk for poor treatment outcomes or for disengaging with treatment before an adequate dose has been received (33,34). Given the importance of primary care in the screening and treatment of adults with depression, determining which factors that might be added to assessments in primary care, to improve knowledge of prognosis after acute-phase treatment, might have important utility.

Prognosis and Personalized Medicine

The term prognosis refers to the “likelihood of future health outcomes in people with a given disease or health condition or with particular characteristics” (34, p.1). We may be interested in knowing the likely prognosis in a number of contexts. These could include “establishing typical prognosis in a broad population, establishing the effect of patients’ characteristics on prognosis, and developing a prognostic model (often referred to as a clinical prediction rule)” (34, p.1). Prognostic studies can identify characteristics or factors that might go into such models or means of determining the most likely prognosis for a given individual. In so doing, such studies can help make predictions about the likely outcomes if no treatment were taken, if one particular type of treatment were taken, irrespective of whatever treatment is taken, or independent of the effects of treatment (i.e. controlling for the effects of treatments to find factors that are generally indicative of prognosis after accounting for the effect of treatment on prognosis) (36).

We might think of these different but similar means of determining prognosis in the following way: when no treatment is taken we can understand prognosis in the “natural state”, or the “natural course” of depression. This is most informative with respect to choices of receiving treatment or not, and may be best studied in longitudinal cohorts or general population surveys with longitudinal measurements or follow-up appointments (30). However, there are two main problems with this. Firstly, presenting to a health professional with a health complaint and providing the necessary data to determine prognosis might be considered an intervention in its own right, particularly if the professional is empathic (37). So, the “natural state” is very difficult to assess in healthcare settings. Knowledge of prognosis for those with a particular health condition is most applicable and has greatest utility to those that seek treatment for such a condition, so studying wholly non-treated populations may not be informative for those that do seek treatment (37). Secondly, there are very few general population surveys or longitudinal cohorts that meet the necessary conditions to be able to study the association between patient characteristics and prognosis independent of the effects of interventions that might influence prognosis (whether intentionally as in the case of treatments, or not, as in the case of a visit to an empathic medical professional irrespective of treatment). Only five such studies could be found that utilise means of continually measuring depressive symptoms in order to not miss episodes in between follow-up visits (38–42), this is essential to

study prognosis accurately and to understand the “natural course” of depression (30). Of those five studies, two involve longitudinal follow-up of RCT patients after randomisation has been broken, those RCTs purposefully selected patients thought to be at high risk of relapse by virtue of having multiple prior episodes and co-occurring personality disorders or substance abuse disorders, so they are not representative of the majority of depressed patients (39,40). Another of the five cohort studies included largely inpatient psychiatric participants with very high levels of co-occurring personality disorder, bi-polar disorder, or depression with psychotic features (38), another two included a mix of patients with depression and primary anxiety disorders (41,42). Crucially, none of those studies enquired about treatment on a regular basis with the exception of one that only recorded patient’s weekly doses of antidepressants and ECT (38), which was in keeping with their mainly inpatient psychiatric sample. This means that there appear to be no cohort studies or general population surveys which allow for a thorough investigation of the relationships between patient characteristics and prognosis for adults with depression independent from the effects of any treatment or other interventions (30).

When the aim is to develop a model or a clinical prediction rule to predict the likely prognosis for a new patient, if a choice of treatments is available we do not initially know the treatment they will receive or that they want (if any). So, models that require treatment to be known when there is more than one treatment to choose from may not be as helpful as those that operate irrespective of treatment (43). Prognosis irrespective of treatment means ignoring treatment in outcome models (43). For such investigations studies where treatment is given are needed, but randomisation to one or other type of treatment is not necessary. A common alternative in prognostic studies of depression is to hold treatment constant, so the question becomes “what is the prognosis for a given individual with a given type of treatment” (44). When there is only one treatment type available or one that is considered superior to all others, this may be particularly useful, but for the majority of depressed patients this is not the case. Understanding prognosis with specific treatments might allow us to consider the differential benefit of one type of treatment compared to another (44). However, this is a question of treatment-by-person interaction; a prescriptive question, and is best answered by prescriptive designs

whereby a controlled treatment is superimposed on top of the “natural course”; thus giving the most robust test of personalising care (30,45).

As will be discussed below and in subsequent chapters, the focus of the studies presented in this thesis was on determining factors that are associated with prognosis such that they might inform clinicians of those things that may be routinely included in their assessments when patients initially present to them with depression. Given that there is considered to be approximate equivalence in the effectiveness of the most common types of treatment for depression, particularly for those with depression that initially present to a primary care clinician (32,46–48), the focus here is on associations between patient characteristics and prognosis independent of treatment. Prognosis independent of treatment means controlling for the effects of different treatments by adjusting for them in a statistical model. It is therefore best assessed where participants are randomised to a number of different types of treatment or there are a number of types of treatment delivered to set standards. The point of this type of study is to control for, or remove the effects of treatment on prognosis. Then, what is left is the association between the potential prognostic factor and the prognostic outcome. Basing such studies on pre-treatment patient characteristics has the advantage of indicating prognosis prior to a decision about which type of treatment to start, and ensuring generalisability at least to the types of treatments controlled for. It therefore means that clinicians can be informed of which variables might be assessed for in order to inform prognoses for their patients regardless of the choice of treatment they or their patients might make. Such studies highlight factors that are associated with prognosis in general, so may guide clinicians to consider those patient characteristics or experiences that might be indicative of poorer outcomes regardless of what treatment they might offer, and considerations of additional forms of support that might be required. Such general factors may also be the most generalisable across health settings where different types of treatment are available (30). Further, this is important because both clinicians and patients often want to know this information (49). However, as will become clear below, this form of prognosis has been largely overlooked in studies of depression.

In other areas of medicine such as cardiovascular disease, where physicians can gather data on known prognostic factors in their consultations (e.g. blood pressure), they are able to use a prognostic algorithm to predict a patient's risk of suffering a cardiovascular event in the next five years (50). This can then help patients and clinicians make a joint decision about whether or not to start treatment to reduce the risk of such an event. We currently lack robust knowledge of the prognostic factors that could be used to build such an algorithm for those with depression (31,33,51). Research that identifies general prognostic factors, independent of treatment, might contribute to the future development of such an algorithm and aid patients and clinicians in considerations of the clinical management of depression. The first step is to identify the prognostic factors, as noted above, this can inform the content of routine assessments and discussions between clinicians and patients about the management of depression. This is separate from the development of a model that can most accurately predict any individual patient's prognosis, for which the best method would be to follow recent conventions to develop and validate prediction models (52–56). The latter would provide valuable information but such models have more modest utility as the interpretation of any individual factor in the models is challenging given the complex interactions often fitted. Unless the predictive model can be used routinely by clinicians, the ability to use the model to inform prognosis in this way is limited (52). This endeavour is also distinct from developing a model which can determine which of two treatments might be most likely to benefit an individual patient. Such information can only come from prescriptive studies in which person-by-treatment interactions are tested (31). Knowledge of the prescriptive effects can be very valuable to clinical practice but as noted above, they require the narrowing of treatment to just two options, and the proper utility of such models can only come from prospective testing and re-calibration of the models to each setting, service, or context in which the patients present (31,45).

Overall aim

In light of the above, the overall aim of this thesis is to investigate the factors associated with prognosis for adults with depression independent of treatment, in primary care. What follows below is a synopsis of the literature reviewed to elaborate detailed aims and objectives for the thesis. Some of this literature is outlined in the

remainder of this chapter, and greater detail is given in the introduction section of each subsequent chapter to provide the rationale for the specific aims of each study presented in this thesis.

Developing knowledge of prognostic indicators

As noted above, studying prognosis independent of treatment might require the use of randomisation to different treatments. However, the restriction of prognostic studies to participants from randomised controlled trials (RCTs) or systematic reviews of RCTs has been criticized as some authors argue that RCTs have samples that are not representative of patients in naturalistic settings (57–59). This then brings in to question the generalisability and utility of findings from such studies to inform clinical practice for individual patients (57–59). Findings from RCTs are criticised when the studies are conducted on small but representative samples from whom findings can be easily generalised, because they include “too few people” (57). The same is true when they are conducted with patients fitting highly selective inclusion criteria as the samples are deemed “unrepresentative” of many patients seen in clinical practice even if the studies have large sample sizes (57). Recruiting a greater range of patients by widening the inclusion criteria is argued to be a strength of pragmatic RCTs so studies of prognosis independent of treatment might be best conducted with such studies (57). However, in order to be more representative pragmatic trials are sometimes criticised for having less well-defined eligibility criteria, resulting in findings that can also be difficult to generalise and make use of in clinical practice, that is unless analyses are conducted on subgroups of patients within such trials (57). Subgroup analyses are problematic though as the vast majority of RCTs are powered to determine overall treatment effects, thus nearly all conceivable subgroup analyses within any trial will be underpowered, giving rise to the risk of false negative results (57,58).

Meta-analyses of RCTs have subsumed the single trial as the most powerful way to determine treatment effects and improved on the ability to consider heterogeneity of effects, though they typically do this at the level of the trials not at the level of individuals, so they too are criticised for not delivering results that are generalisable to many individuals (58). The most common approach to investigating the reasons

for heterogeneity in study-level meta-analyses is by testing interactions between treatment effects and potential moderators in subgroups using meta-regressions (59). These analyses use a summary statistic (usually the mean) so they can be useful in considering effects that apply uniformly across all participants of a given RCT. For that reason they are often constrained to process variables such as whether or not those delivering treatment were blinded to allocation, or to service level variables such as the location of the trial sites. They are however less useful for investigating moderators which vary at the level of the individual (such as age or the number of treatment sessions attended) (59). As study-level meta-analyses do not have individual patient data, meta-regressions therein make the assumption that an aggregate value (such as a mean or median of a particular variable of interest) can be appropriately assigned to all patients within any given trial, and have been criticised for being both inefficient and for introducing additional bias, particularly ecological bias (i.e. making inferences about individual patients from aggregate data which may not be correct) (60).

In order to more appropriately investigate factors that vary at the level of the individual patient (e.g. symptom data), the preferred method is to conduct analyses using the individual patient data (IPD) from each trial that makes up the meta-analysis (46,59). IPD datasets do not suffer from ecological biases and are typically thought to be equipped to reduce other sources of bias such as those related to selective reporting in trials, and to missing data (61–63). IPD datasets also have greater power to conduct subgroup analyses and investigate interactions which study-level meta-analyses and individual trials are often underpowered to conduct (57,64), and can be used to investigate the concurrent contribution of multiple potential prognostic factors that differ at the level of the individual (57,64).

Current knowledge of prognostic factors in depression

One difficulty in better understanding prognostic factors for depressed adults is that there have been apparent inconsistencies in the findings of different RCTs and systematic reviews of RCTs, with subsequent uncertainty about what factors are associated with prognosis (51,65). To better understand the breadth of such findings, literature searches were conducted to identify systematic reviews or meta-analyses

(whether study-level or IPD) that investigated baseline patient-level factors associated with prognosis for adults with depression. Searches were run on the Cochrane Database of Systematic Reviews, the Prospero Register of Systematic Reviews, Embase, and Medline. Details of the search terms and results from the searches can be found in Appendix 1 Table 1.1. Across the databases 632 articles remained after removing duplicates, 71 of these were somewhat relevant to the aim of this thesis and were read in full, from which 29 were directly relevant as they identified patient characteristics associated with prognosis for adults with depression. These 29 studies are summarised in Table 1.1.

As can be seen in Table 1.1 none of the studies investigated factors associated with prognosis independent of a range of treatments, 22 assessed response to particular treatments, six were studies of the "natural course" of depression (although three of these included analyses irrespective of treatment, whether intentionally or not), and one aimed to assess prognosis irrespective of treatment. What follows is a narrative review of these 29 studies.

Table 1.1. Review of systematic reviews, meta-analyses and IPD studies that report on the associations between patient characteristics and prognosis.

Article	Study Type	Studies searched for	Number of studies of depression (K for meta) and sample size (N for meta)*	Sample, Setting, Recruitment	Type of Prognosis Studied	Baseline Patient Prognostic Factors Assessed	Main findings regarding prognostic factors	Limitations for assessing factors associated with prognosis
Noma et al., 2019 (66)	Not Systematic Review but used IPD	Placebo-controlled RCTs conducted in Japan	K=7(7) N=2803(2803)	Unclear	Response to a particular treatment (antidepressants), prescriptive test of interaction	Depressive symptom severity, duration of episode, number of past episodes, age, gender, and age at onset	Depressive symptom severity was associated with outcome both in antidepressant and placebo groups, duration was only associated with outcome in the antidepressant group, age, age at onset, and gender were only associated with outcome in the placebo group. No association between having 3 or more past episodes and prognosis in either treatment condition. Reviewers rated risk of attrition bias as low in 5 studies and unclear in 2 studies.	Different outcome measures were converted via a cross-walk. Two-stage meta-analysis adjusted for treatment but no assessment of heterogeneity making results difficult to interpret. Did not adjust for baseline depressive symptom severity in analyses of other prognostic variables.
Marwood et al., 2018 (67)	Study-level meta-analysis	Neuroimaging studies of patients with depression or anxiety	K=4(2) N=86(33)	Unclear setting; Sample: adults with depression or anxiety scanned before starting treatment	Prognosis irrespective of treatment	Connectivity in neural regions	Too few studies to assess disorder specific effects or make appropriate comparisons between them. Greater activation of the right cuneus cortex at baseline was associated with greater symptomatic improvement across disorders	Very limited by small sample size and small number of studies, particularly of depression. Heterogeneity could not be interpreted. No consideration of attrition. No adjustment for baseline depressive symptom severity.
Wang et al., 2018 (68)	Systematic Review	longitudinal studies	K=23(N/A) N= N/S	Unclear	Natural course	Loneliness, social support	20 studies of depression assessed depressive symptoms at endpoint: lower social support at baseline was associated with higher depressive symptoms at end-point, likewise with higher self-rated loneliness and outcome (in one study)	No quantitative synthesis, uncertainty of sample size, setting and measures used to determine outcomes in the studies qualitatively synthesised. Findings are therefore hard to interpret and associations were most often not adjusted for any treatment. Only 9 studies adjusted for baseline depressive symptoms though findings were still significant in those studies. Attrition rate was unclear in 5 studies and was above 20% in 11 studies.
Haq et al., 2015 (69)	Study-level meta-analysis	Studies of ECT published since 1980	K=51(7 to 32) N= N/S (702 to 1175)	Unclear	Response to a particular treatment (ECT)	Depressive symptom severity, duration of depression, history of treatment with antidepressants, age at onset, age, gender	Shorter durations of depression (7 studies, n=702) were associated with better response to ECT, history of failure to respond to antidepressants (11 studies, n=1175) was associated with worse prognosis. Age was weakly associated with response but there was considerable heterogeneity. Bipolar diagnosis, sex, age at onset, and number of previous episodes were not significant predictors. Association of depressive symptom severity and response to ECT was inconclusive due to high heterogeneity.	Some analyses limited to small number of studies, no weighting for different types of study or sensitivity analyses excluding studies with patients with bi-polar or psychotic depressions making heterogeneity hard to interpret. No assessment of study attrition. Did not adjust for baseline depressive symptom severity in analyses of other prognostic variables.
Johnsen & Friborg,	Study-level meta-analysis	Empirical studies of CBT including RCTs, non-	K=70(70) N=2426(2426)	Unclear	Response to a particular treatment (CBT), interactions with	Depressive symptom severity, age and gender	Age was not related to variation in treatment effects but gender was (studies of women only had higher effect sizes). The number of comorbid diagnoses was not related to	No differential weighting or sensitivity analyses treating the different types of studies differently. Combination of BDI-I and BDI-II scores may have introduced

2015 (70)		randomised studies, uncontrolled studies, and clinical field studies.			time/publication year were also tested		outcomes. The severity of the depressive diagnosis was not associated with outcomes but lower effect sizes were found in studies of mild depression. There were no significant interactions between patient-level characteristics and publication year.	bias particularly to assessment of temporal effects. Recovery was also defined differently across studies making heterogeneity hard to interpret. No assessment of study attrition. Did not adjust for baseline depressive symptom severity in analyses of other prognostic variables.
Sugarman et al., 2014 (71)	Study-level meta-analysis	Industry sponsored placebo-controlled RCTs of Paroxetine registered with FDA for Depression or Anxiety	K=27(27) N=4986(4986)	Unclear	Response to a particular treatment (antidepressants), prescriptive test of interaction	Depressive symptom severity	Higher baseline symptom severity was associated with smaller pre-post effect sizes in both Paroxetine and placebo groups, there was no significant treatment type by severity interaction though.	Included one study of adolescents and two of geriatric populations. Average depression severity ranged from severe to very severe in all studies. Heterogeneity relatively high in both drug and placebo groups. No assessment of study attrition.
Dodd et al., 2014 (72)	IPD	Industry sponsored placebo-controlled RCTs of Duloxetine	K=12(12) N=4987(4987)	Unclear	Response to a particular treatment (antidepressants)	Depressive symptom severity, duration of episode, number of past depressive episodes, age at onset, age, gender, ethnicity and body mass index)	Age was the strongest predictor of response to placebo, followed by duration of depression. Depressive symptom severity and anxiety symptom severity were among the top predictors of remission with Duloxetine and with SSRIs but not as strongly as age was for placebo. Number of past episodes, gender, and ethnicity were less important in all models, age at onset was moderately important in the model of remission with SSRIs but not the placebo or Duloxetine models.	Treated data as a single cohort pooling effects for each treatment across studies without adjusting for between study factors; no adjustment for study or random effects for study fitted, and no assessment of heterogeneity. No fully held-out test data, validation was internal only and did not involve cross-folding training data. Unclear methods of variable selection as separate from process of model building, also unclear on sensitivity analyses and model stability checks. AUC was only metric of model performance and not other important measures such as Brier scores, deviance, log-loss or others. Also no assessment of study attrition.
Steinert et al., 2014 (73)	Systematic Review	Naturalistic cohort studies with at least 3 year follow-up	K=12(N/A) N=4009(N/A)	Community or general practice identified cases	Natural course and prognosis irrespective of treatment	Depressive symptom severity, previous episodes, comorbidities, diagnoses/subtypes of depression, social support, life events, childhood maltreatment, age of onset, educational attainment, and SES.	Owning a home and social support after a negative life-event were associated with recovery. Physical and sexual abuse in childhood as well as adult emotional abuse were associated with a lack of recovery, as was a personal or family history of depression. Personal history of depression (2 studies), onset age (1 study), dysthymia and double depression (1 study), baseline severity of depression (4 studies), and comorbidity (2 studies: anxiety disorder and personality disorders associated with long-term course in one study, personality disorders in the other study) were associated with poorer prognoses. Factors associated with a more favourable course were rare and only pointed out possible worthwhile future approaches. They comprised social support	Mix of effects for recurrence, long-term course, treatment outcomes etc, difficult to isolate effect to prognosis after treatment. Range of sample sizes with no weighting of effects (range from n=33 to n=1996). Authors stated that attrition was not consistently reported and varied considerably across the reviewed studies.

							(1 study), a higher social and educational status (2 studies) as well as a higher onset age (1 study).	
Dodd et al., 2013 (74)	IPD	Industry sponsored Phase III and IV RCTs of Duloxetine	K=15(15) N=5627(5627)	Unclear	Response to a particular treatment (antidepressants),	Number of previous episodes	No main effect was found for the number of previous episodes, comparing three or more to less than three, and none to at least one.	Analyses were conducted with data treated as a cohort, no weighting or adjustment for study and no assessment of heterogeneity. No assessment of study attrition/
Kamp man & Pouta nen, 2011 (75)	Systema tic Review and study-level meta-analysis	studies published 1991-2010 that used the temperament and character inventory	K=10(10) N=938(938)	Non depressed adult community samples, some young adults, some specific populations e.g. school teachers	Natural course; prognosis irrespective of treatment	Types of temperament including harm-avoidance, reward dependence, novelty seeking, persistence, self-directedness, cooperativeness, and self-transcendence.	Harm avoidance temperament was associated with prognosis in clinical samples, others were not.	Mixture of clinical and non-clinical samples, some very small studies included (e.g. n=35). Used fixed not random effects models. Mixing of endpoints (from 6 weeks to 2-years) in same meta-analyses and no harmonisation across different measures. Makes sources of heterogeneity hard to interpret and main findings difficult to interpret too.
Fourni er et al., 2010 (76)	IPD	placebo-controlled RCTs	K=6(6) N=718(718)	Outpatients	Response to a particular treatment (antidepressants), prescriptive test of interaction	Depressive symptom severity,	Baseline depressive symptom severity was associated with prognosis within those that received antidepressants and those that received placebo, there was also a significant interaction so higher severity patients responded better to antidepressants than placebo. Significant interaction between baseline depressive symptom severity and attrition.	Tests of interaction without within study step (falling foul of Fisher et al., 2017 guidelines) only 6 out of 23 eligible studies were included and no use of study-level data from the remaining 17 to consider effects. No control for missing data. Study attrition was high in 2 studies, particularly in the medication arm of one small study (34%).
Bower et al., 2013 (77)	IPD	RCTs reported since 2000 with n>50	K=16(16) N=2470(2470)	Community or primary care settings, included patients with depression and also those with mixed anxiety and depressive disorder	Response to a particular treatments (low intensity CBT), prescriptive test of interaction	Depressive symptom severity	Depressive symptom severity was related to outcome both with LI CBT and controls, there was moderation by severity so LICBT patients experienced better outcomes as severity increased vs control patients.	Only 55% of eligible studies/57% of eligible patients included so a number of potential biases may have affected results. Results may not be generalisable to patients diagnosed with depression due to more lenient inclusion criteria. Differential dropout rates between interventions could have led to systematic bias in moderator analysis. Used a cross-walk for BDI to CORE-OM or vice versa. Attrition assessed but not reported for each study.

Chekroud et al., 2016 (44)	Not Systematic Review but used IPD	N/A convenience sample	K=3(3) N=4326(2234)	Primary care and psychiatric outpatient	Response to a particular treatments (antidepressants), model built with one treatment tested in other treatments without test of interaction	All baseline variables available. Top 25 used in final predictive models included: depressive symptom severity, history of antidepressant medication, history of prior episodes, comorbid anxiety symptoms, comorbid panic attacks, race/ethnicity, employment status and years of education.	Depressive symptom severity, employment status and years of education were among the top predictors of outcomes. Comorbid anxiety, history of antidepressant medication (Sertraline) and race/ethnicity also associated with outcome.	Convenience sample of data, machine learning model used to find significant predictors but direction of effects difficult to determine with complex model. No details on the studies that would have been eligible but that were not included, no assessment of heterogeneity is provided and difficult interpreting reasons for differential model performance across study groups. Study attrition not assessed although last observation carried forward analysis for STAR*D participants is provided.
Weitz et al., 2015 (78)	IPD	RCTs that randomised to CBT vs antidepressants	K=16(16) N=1700(1700)	Not inpatients, otherwise fairly unclear. 3 studies included specific populations (patients with multiple sclerosis, women with low incomes, and women with infertility problems)	Response to a particular treatments (CBT and Antidepressants), with prescriptive test of interaction	Depressive symptom severity	Depressive symptom severity was associated with continuous symptom outcomes independent of treatment group, but not to response (50% reduction in symptoms) outcome, and no evidence was found for differential effect of CBT vs antidepressants.	Imputation conducted across the whole sample not within each study first. Used 1-stage meta-analyses with multi-level effects with individual effects on one level and study level effects at another level of the model. 1/3 of eligible studies did not provide data. Attrition is not assessed, risk of bias due to lack of intention-to-treat (ITT) analysis in studies is reported; four studies did not use an ITT but attrition rates are not reported in the review.
Karyotaki et al., 2017 (79)	IPD	RCTs of internet guided low-intensity CBT	K=13(13) N=3876(3876)	7 of 13 recruited community participants - others unclear	Response to a particular treatment (internet based self-guided CBT), with prescriptive test of interaction	Depressive symptom severity, comorbid anxiety, age, sex, educational level, relationship status, employment status	Depressive symptom severity was associated with outcomes but no moderation by severity was found. None of the other factors were associated with outcomes but adherence moderated outcome between iCBT and controls. Analyses controlled for baseline depressive symptom severity.	A mix of 1-stage and 2-stage meta-analyses and of IPD and study-level analyses, which is perhaps unnecessary. Study focussed on moderation rather than prognostic effects, and had moderate to high heterogeneity which could not be well explained. Most were community samples rather than patient samples so generalisability to clinical settings/treated patients is limited. All studies reported as low risk of bias in all domain but study attrition not reported.
Driessen et al., 2010 (80)	Study-level meta-analysis	RCTs	K=132(16) N=10134(N/S)	Outpatients	Response to particular treatment (psychological therapies), prescriptive test of interaction	Depressive symptom severity	No evidence that pre-treatment severity predicted response to psychological treatment vs control condition; in a subset of studies with within study severity findings reported, psychological therapy was more effective with higher levels of severity	Findings are from meta-regression using mean severity as predictor of response and use prescriptive design to test treatment-type by severity interaction, could be considered to fall foul of appropriate tests of interactions in meta-analyses as per Fisher et al., 2017. Study attrition rates not reported.

Nakabayashi et al., 2018 (81)	IPD	Placebo-controlled RCTs submitted to pharmaceuticals and medical devise agency in Japan.	K=5(5) N=1898(1898)	Adults with MDD and no comorbidities, 2 studies excluded patients with treatment resistant depression. All patients were Japanese in 4 studies, and Japanese or Korean in 1 study.	Response to particular treatment (antidepressants), prescriptive test of interaction	Depressive symptom severity, history of antidepressant medication, age	Baseline depressive symptom severity was associated with prognosis within those that received antidepressants and those that received placebo. History of antidepressant medication and age were associated with response to antidepressants but not placebo, gender was not associated with response to either antidepressants or placebo. Age, and a history of past antidepressant medication moderated the effect of antidepressants compared to placebos.	Regression analyses performed one-step without effects calculated within trial or adjusted for allocation within each trial, and interactions were tested across studies treating data like a cohort (not following recommendations of Fisher et al., 2017). Dropout rates for the included studies were around 10-15 % for all conditions though in one study they were above 20% in each arm. No adjustments were made for baseline severity in assessment of other prognostic variables.
Gariepy et al., 2016 (82)	Systematic Review and study-level meta-analysis	Observational general population studies in "western countries"	K=36(36) N= N/S	General population	Natural course	Social support	Most studies reported associations between social support and protection from depression, most evidence for emotional support followed by instrumental support.	Not treatment seeking sample. 31 studies included were of children or adolescents, 33 were of older adults. Most studies were cross-sectional (28 of 36 in general adult age group) so cannot rule out reverse causality. Only 5 studies rated high quality. Combined estimates from adjusted models in included studies so estimates adjusted for different variables making interpretation of heterogeneity complex. Nearly all studies used different social support measures, only 1/3 used previously validated measures, so comparisons complicated. No assessment of attrition.
Schoemaker et al., 2018 (51)	Meta-review with Qualitative Synthesis	SRs and study-level meta-analyses of placebo controlled RCTs	K=58(N/A) N= N/S	Unclear	Response to particular treatment (placebo)	Depressive symptom severity, duration of illness, duration of episode, comorbid anxiety, comorbid health problems, age, gender, race/ethnicity, body mass index, trial design factors	Very weak evidence for a decrease in response to placebos with: 1) increased baseline severity, 2) increased duration of illness, 3) increased duration of current episode, 4) comorbid anxiety/somatization. Strong evidence for absence of effect on placebo response for age and very strong evidence for same with sex. Weak evidence of absence of effect of race/ethnicity and comorbid physical health conditions, very weak evidence for absence of effect of concomitant medication.	Overlap in RCTs included in the review. Evidence for positive or negative associations all weak or very weak. Results regarding symptom severity come from 3 narrative meta-reviews of meta-analyses with 2 finding a weak effect and one finding no effect. No assessment of attrition within the reviewed reviews or primary studies.
Paykel, 1994 (83)	Review, unclear if systematics	Unclear	K=29 N=N/S	Some outpatients, some general population, some psychiatric patients, generally unclear.	Natural course, and Irrespective of Treatment	Life events	Some evidence that those with life events prior to starting treatment might have poorer prognosis than those without life events.	Methods are somewhat unclear. In narrative review some studies showed effects and are compared to studies of prognosis in a different context, different setting and with very different samples making interpretation difficult. No quantitative synthesis and no comments on heterogeneity of findings, control for other factors, or reverse causality. No assessment of attrition.

Carter et al., 2012 (84)	Systematic Review	Studies published 1998-2008; adults, any study type, n>50.	K=76(N/A) N=Unclear(N/A)	Unclear	Response to particular treatment (antidepressants)	Socio-demographics (age, gender, race/ethnicity, marital status, and SES), Clinical (depressive symptom severity, frequency and duration, comorbid anxiety, comorbid pain, other comorbidities, substance abuse), and Social support	Strong evidence for: baseline severity (14 studies); duration of depression (3 studies); social support (4 studies), SES (4 studies); comorbid pain (4 studies). Good evidence for prognostic associations with: age (10 studies) age at onset (3 studies), gender (9 studies), marital status (4 studies), comorbid anxiety (7 studies), other comorbidities (6 studies). Some evidence for associations with: marital status (4 studies); number of past episodes (1 study) and substance abuse (1 study). Age (3 studies) and gender (1 study) were not associated with attrition. Ethnicity was associated with attrition (1 study) – Caucasians were less likely to dropout than non-Caucasians.	Mixture of reviews and primary studies, some double counting of effects. Effects of several factors claimed to have strong evidence later stated as inconclusive, particularly age and gender. Considerable heterogeneity, unclear of weighting in decisions about what counts as "strong evidence" vs "good evidence" etc. No harmonisation of data so mixed a number of different subtypes of anxiety, or comorbidities, of SES factors, and combined frequency of depressive episodes with duration of depression at baseline, hampering interpretation.
Sockol, 2018 (85)	Systematic review and study-level meta-analysis	Studies of IPT in perinatal women, including RCTs, quasi-randomised trials, and open trials.	K=17(17) N=790(790)	Unclear	Response to a particular treatment (IPT), prescriptive test of interaction	Depressive symptom severity, age, marital status, ethnicity	As the proportion of study participants that were married was increased the effect size of IPT on depressive symptoms increased. The converse was found when the proportion of 'minority' patients increased. Higher depressive symptom severity was associated with larger effect sizes when measuring change in symptoms pre-post treatment. Higher maternal age was associated with smaller effect sizes.	Of the 17 studies most had very small samples (e.g. n=6, 11, 11, 12, 13, 17, 18, 23, 32, 42, 48, 50, 50, 53). Main results regarding prognostic indicators had very high heterogeneity affecting interpretability of the results. Attrition ranged from 0-56% in studies, with a mean of 16%. No adjustments for baseline depressive symptom severity for other prognostic variables.
Cuijpers et al., 2018 (86)	Study-level meta-analysis	RCTs psychological treatments with any comparator. Acute treatment studies not relapse prevention.	K=256(256) N=Unclear	Unclear	Response to a particular treatment (Psychological therapy), prescriptive test of interaction	Age, gender, ethnicity, depressive diagnosis (e.g. chronic, subthreshold)	Studies in students had higher effect sizes than studies of older adults. Studies of adults with general medical disorders had lower effect sizes than other studies. No difference in effect sizes based on gender, ethnicity, or severity of diagnosis.	Prognostic factors not tested specifically in the studies, meta-regression used to test differences where studies were of one particularly population subtype (e.g. students) and compared to another subtype (e.g. older adults). Generally high levels of heterogeneity. Attrition per study was not reported in the review but 50% of studies did not use an intention-to-treat analysis.
Cristea et al., 2019 (87)	Systematic review and study-level meta-analysis	RCTs of psychological therapy with biological markers measures pre-treatment	K=8(1 to 2) N=Unclear	Unclear setting; Mostly adults with MDD but some with other mood disorders or just depressive symptoms, most studies in populations with somatic diseases	Response to a particular treatment (Psychological therapy), prescriptive test of interaction	Biomarkers - functional connectivity, brain metabolism and genetic polymorphisms. Others were examined but not as prognostic factors	One study found that functional connectivity between the subgenual cortex and prefrontal, insula and midbrain regions was associated with outcomes from CBT and antidepressants (Duloxetine, Escitalopram). Another study showed brain metabolism in 6 regions, most notably the right anterior insula was associated with outcome from CBT and Escitalopram. A third study found that those responding to CBT vs Venlafaxine showed increased metabolism in the inferior temporal cortex, and decreased metabolism in the posterior cingulate. No interaction effect was found between treatment and serotonin inhibitor polymorphisms. A study of the brain-derived neurotrophic factor polymorphism reported no prognostic or	Very few studies of prognostic association between biomarkers and outcome. Very difficult to interpret sources of heterogeneity and generalisability of findings questionable. Also overall 10 studies were rated as high risk of bias. 2 studies had high risk of bias due to incomplete data. Specific attrition rates were not reported in the review.

							prescriptive association for the genotype under study although an interaction was found between the genotype and childhood adversity which was associated with prognosis across randomised groups.	
Ebrahim et al., 2012 (88)	IPD	RCTs of CBT vs no treatment, usual care, or minimal treatment, that had disability benefit status as inclusion criterion	K=8(2) N=1502(227)	Unclear	Response to a particular treatment (CBT)	Disability benefit status	Tentative suggestion that effect size may have been higher in patients in receipt of disability benefit compared to those not in receipt of such benefits.	Most studies did not provide IPD. Only 34 people had exposure of interest in IPD, so difficult to interpret results. No adjustment for baseline depressive symptom severity. Attrition ranged from 4%-40% across the studies. No adjustment for baseline depressive symptom severity.
Lin et al., 2019 (89)	Systematic review and study-level meta-analysis	Studies of Venlafaxine with measures of CYP2D6 metaboliser status and pharmacokinetic outcomes	K=14(1 to 3), N=1035(12 to 571)	Included studies of healthy volunteers (K=5) as well as depressed samples (K=9). Settings unclear.	Response to a particular treatment (Venlafaxine)	Phenotypes of CYP2D6 gene	CYP2D6 phenotypes were not associated with response to Venlafaxine.	No control for type of study (health volunteers or patients) and majority of studies had very small sample sizes (e.g. 12, 14, 20); one study with more participants than 12 of the others combined, and most of the meta-analyses had only two or three studies so heterogeneity hard to interpret. 9 studies reported to not have clear reporting of attrition rates or gave no reasons for attrition.
Kloiber et al., 2013 (90)	Not SR but used IPD	N/A convenience sample	K=3(3), N=2256(2256)	338 inpatients and 346 controls from the Munich Antidepressant Response Signature project, a case-control study. Also 672 patients from GENDEP and 980 from STAR*D	Response to a particular treatment (antidepressants)	In meta-analysis only polymorphisms of leptin gene, in single study also leptin mRNA expression and leptin serum levels.	In the meta-analyses no significant associations were found between polymorphisms of the leptin gene and antidepressant treatment outcomes. Such associations were found in the exploration and replication samples from MARS (single study). In the single study (MARS) lower leptin serum levels and reduced leptin mRNA expression were associated with poorer treatment outcomes independent of leptin genotype.	Only three studies included in convenience sample with no statement on the number of other studies that might have been eligible but were not approached for IPD. No details on the methods of meta-analysis or statements about heterogeneity are provided. No assessment of study attrition. No adjustment for baseline depressive symptom severity.
Morris et al., 2009 (91)	Systematic review	Prospective studies measuring the relationship between positive and negative emotionality and course of MDD	K=6 to 22(N/A), N=245 to 3553(N/A)	Adults aged 18-65 diagnosed with MDD. Setting unclear	Natural course	State and trait levels of negative and positive emotionality	Lower levels of positive emotionality were associated with poorer MDD course. Lower levels of state negative emotionality and higher levels of trait negative emotionality were associated with poorer MDD course. The associations in individual studies assessed were often present after controlling for baseline depressive symptom severity.	For state analyses combined a variety of ways or measuring emotionality (e.g. heart rate, skin conductance and self-report) and different measures of depression and different time intervals all in the same analysis. This makes interpretation and understanding sources of heterogeneity very difficult. Similar issues for trait analyses: combining resting EEG asymmetry with extraversion measures, combining treatment response with diagnostic status and relapse, at intervals of 2-36 months. Also combined some

								studies which included treatment with those in non-treated samples (some combined prognosis irrespective of treatment with natural state). No assessment of attrition or control for baseline depressive symptom severity.
Iovieno et al., 2011 (92)	Study-level meta-analysis	Placebo controlled-RCTs of antidepressants in patients with co-morbid long-term health conditions	K=212(190) without Axis-III inclusion criteria & 29(25) with Axis-III inclusion criteria N=46900(46900) without Axis-III & 2338(unclear) with Axis-III	Unclear	Response to a particular treatment (antidepressants)	Comorbid long-term health conditions (“Axis-III” disorders in DSM)	Studies which specifically selected patients with Axis-III disorders comorbid to MDD had higher response rates to antidepressants compared to studies that did not use Axis-III comorbidity as a selection criteria. There was a non-significant trend towards the same effect in those randomised to pill placebos. Analyses adjusted for baseline symptom severity.	All included studies were efficacy trials so results may not be generalisable to wider MDD population. Combining results across studies with very different inclusion/exclusion criteria may make inferences invalid, and although studies for comparison did not have Axis-III conditions as an inclusion criterion it is not known what proportion of their participants had Axis-III conditions. No assessment of study attrition.

* Note numbers do not correspond to total numbers of studies and participants assessed overall in each study, instead they represent numbers of studies (k) and participants (n) of relevance to this review.

Depressive symptom severity

The factor most commonly assessed for an association with prognosis in the 29 reviewed studies was baseline depressive symptom severity (16 studies). Findings across these studies were mixed although the majority reported some level of prognostic association between symptom severity and outcome (13 studies), one study found it to be inconclusive, and two studies found no effect of symptom severity. Further, most studies that assessed a prescriptive effect, i.e. moderation of response to one compared to another type of treatment by baseline symptom severity (10 studies) found no evidence for such an interaction effect (8 studies).

Considering these studies in greater detail, all but two of the study-level systematic reviews identified in the searches found depressive symptom severity to be associated with prognosis, however the direction and size of the effect was not consistent across the reviews (80).

1) A study-level meta-analysis of 132 trials of psychotherapy including 10,134 patients found that baseline depressive symptom severity was strongly associated with treatment outcomes. However, this was only conducted on a subgroup of 16 studies, some with small samples and those conducted with different patient populations (e.g. working age adults, older adults only, those recruited in clinical settings and those not receiving treatment in community settings) bringing in to question both the ability to interpret heterogeneity in the analyses and the generalisability of the findings.

2) A second study-level meta-analysis of 17 studies of Interpersonal Psychotherapy (IPT) with perinatal women reported that higher depressive symptom severity at baseline was associated with larger overall effect sizes (85). However, there was very high heterogeneity and again most studies had very small samples ($n<50$) making it difficult to interpret the reasons for and meaning of the heterogeneity.

3) A third study-level meta-analysis, this one of 32 ECT studies, found no conclusive evidence regarding the association between symptom severity and outcome due to high levels of heterogeneity (69). However, there was

no differential weighting or sensitivity analyses presented to allow for the fact that some studies were randomised and others were not, or that some were studies of adults with bi-polar disorder or depression with psychotic features, while others were studies of MDD. It is noteworthy too that the variation in levels of severity in the studies of ECT was lower than would be expected to be the case for most patients with depression, certainly patients treated in primary care, and was lower than in other studies that have found evidence for such an association.

4) A fourth review of 70 CBT studies (of randomised and non-randomised designs) reported that the severity of depressive symptoms was not associated with outcomes, although lower effect sizes were reported in studies of mild depression (70). This study used different definitions of recovery in the same analyses and combined studies with different versions of the same symptom measure, again making it difficult to interpret sources of heterogeneity.

5) A fifth review investigated effects in 27 RCTs of Paroxetine conducted by pharmaceutical companies, and found that higher baseline severity was associated with smaller effect sizes in both the drug and placebo groups (71). However, this included one study of adolescents and two of geriatric samples, and the average levels of depressive symptom severity were severe to very severe in all studies, limiting the variance of severity.

6) One study-level review assessed the “natural course” of depression in cohort studies, three primary studies reported positive associations between depressive symptom severity and poorer prognoses (73), there was a mix of outcome variables (including recurrence, “natural course”, and treatment outcomes) making the interpretation of the results difficult.

7) A meta-review of 58 systematic reviews found very weak evidence for an association between depressive symptom severity and response to pill placebos (51). The results regarding symptom severity came from three narrative meta-reviews of meta-analyses with two studies finding a weak

effect and one finding no effect, and no adjustment for the reviews including the same primary studies. For further details see Table 1.1.

The greatest limitation to these reviews in regards to investigating associations between patient characteristics and prognosis is perhaps that they were limited to group level analyses of factors shared across primary studies. So, they were assessing whether studies that had different mean levels of baseline symptoms had different effect sizes for the treatment of interest (or the association with depressive symptoms at a particular time point in groups of patients that apparently did not receive treatment), potentially introducing ecological bias (60). As noted above, in order to assess the association of several potential prognostic factors in conjunction and to not propagate an ecological fallacy, IPD data may be required (57,93). Eight studies were found that used IPD datasets to study prognostic factors for adults with depression. There was a limited range of treatments included in each study as most included studies of particular antidepressants or particular forms of psychological intervention, further, the setting for the patient samples was for the most part unclear in these studies, potentially limiting generalisability, see Table 1.1.

1) Two of these IPD studies assessed effects of low-intensity CBT, one included 16 RCTs published since the year 2000 (77) and another more recent review included 13 RCTs (79). Both studies reported an association between baseline depressive symptom severity and treatment outcomes in both the active and control group conditions. However, in the first of these studies IPD were gained from only 55% of the eligible studies and some studies included participants without diagnosed depression, potentially limiting generalisability to patients seeking treatment for depression (77). The second of these studies included a majority of trials with community samples also limiting generalisability and limiting the ability to interpret the quite high levels of heterogeneity found (94).

2) A third IPD assessed prognostic factors in 16 RCTs of CBT and antidepressant medication (78). This study found that depressive symptom severity was associated with the level of depressive symptoms post-treatment (continuous score) but not to a binary outcome using the same symptom

measures (response). As with one of the studies above, this IPD dataset missed a number of RCTs that may have impacted the results, and in addition, the authors conducted imputations of outcomes across all patients from all studies, potentially biasing results by not including some variables unique to certain studies which could have improved the imputation models, and not allowing for differences between the studies in their imputation models (64,95).

3) Five of the eight IPD studies to report on the association between depressive symptom severity and prognosis focussed on antidepressant treatments. These studies either used niche datasets (66,81) to find eligible studies, did not run systematic searches for such studies (44), the nature of the literature searching was not clear (72), or the majority of eligible studies were not able to be included in the analyses (76), see Table 1.1 for details. They all included a small number of trials with large sample sizes and reported associations between depressive symptom severity and prognosis. However, the strength of these associations, the direction of them, and reasons for heterogeneity were not always able to be interpreted (or heterogeneity was not assessed). The settings for the primary studies were most often unclear, and as the specificity of the treatments (e.g. Duloxetine (72)) or of the country in which the studies had to have been situated (Japan (66,81)) limited the samples in most of them, the generalisability of the results to most adults with depression is questionable (30,64).

Taken together, although not all studies reported the same direction of effects, it would appear that there is a degree of consistency in the finding that baseline depressive symptom severity is associated with prognosis with specific treatment modalities, in community samples, or irrespective of treatment. Whether this also holds independent of treatment is yet to be determined, as is the strength of any such association. That severity is related to outcome holds with the ‘common-sense’ view of most illnesses, depression included, but a number of other factors which may be related to, but separate from depressive symptom severity, which I will call indicators of depressive ‘disorder severity’ (these factors are listed as sub-headings below in the next section), were also found to be associated with prognosis in the

reviewed studies in Table 1.1. I will briefly review the related findings here, more details about and a critique of each review are provided in Table 1.1.

Depressive 'disorder severity'

Duration of depression or chronicity

Five studies reported on the association between the duration of depression at baseline and treatment outcomes with some mixed results.

1) An IPD meta-analysis of RCTs found evidence of an association between the duration of depression and prognosis in patients treated with antidepressants but did not find such evidence among those randomised to pill placebos (66).
2) Another review of RCTs including treatment with Duloxetine found the opposite result such that there was evidence of an association between the duration of depression and response to pill placebos but not Duloxetine, although this study only reported a single binary outcome. As shown above, some studies have reported different directions of some effects when using continuous compared to binary outcomes (78), so there may be uncertainty as to whether the effects would have also been found using a continuous outcome. The results of this study were based on a machine learning model using 14 baseline characteristics so independent association of each variable with the outcome was difficult to determine. The study also treated data from all primary studies as a single cohort without adjustments for differences between the studies, potentially invalidating the results (72).
3) A meta-review study found very weak evidence of an association between duration of depression and response to placebos, although this involved no quantitative synthesis (51).
4) A study-level systematic review reported strong evidence of an association between depressive duration and antidepressant treatment response, although this was based on just three primary studies and again this did not involve any quantitative synthesis (84).
5) A further study-level review found shorter durations of depression to be associated with better response to ECT (62), but as noted above, this review included primary studies of patients with a variety of types of depression, including bi-polar disorder, and a higher degree of baseline symptom severity

than other studies, and so the findings may not be applicable to those with unipolar MDD.

Overall, it would appear that the duration of depression maybe associated with response to some treatments, but perhaps not others, it is therefore debateable whether it might be associated with prognosis in general, and to clarify this the duration of depression needs to be investigated independent of treatment.

Comorbid mental health symptoms or disorders

Five studies also reported associations between comorbid mental disorders or comorbid symptoms of anxiety and prognosis, although the generalisability of these findings and the implications of them for patients seeking treatment for depression in primary care are somewhat uncertain. These studies included:

1) A systematic review of 76 primary studies of any design that involved participants treated with antidepressant medications (84), reported to have found good evidence for an association between comorbid anxiety and treatment response. However, only seven primary studies reported on this and there was no quantitative synthesis or consideration of heterogeneity.
2) An IPD study with no systematic review, comprising three RCTs of antidepressant medication that found higher levels of somatic anxiety to be associated with worse treatment response to Citalopram, and Escitalopram (whether taken alone or in combination with Buproprion) (44).
3) A systematic review of cohort studies that found some evidence that both anxiety disorders and comorbid personality disorders were associated with relapse or recurrence in the two primary studies to investigate these associations (73).
4) A meta-review that reported very weak evidence for an association between comorbid anxiety and poorer response to placebos (51).
5) Lastly, an IPD meta-analysis that found no evidence of an association between comorbid anxiety symptoms and response to internet based low-intensity CBT (79).

Taken together these studies might point to an association between comorbid anxiety and poorer prognosis that is specific to antidepressant medications;

assessment of the association between comorbid anxiety and prognosis independent of a range of treatments is warranted.

History of depression or of treatment for depression

Eight studies investigated the association between some measure of a history of depression and prognosis: including the number of past episodes (4 studies), a history of antidepressant medication (3 studies), and a history of depression irrespective of the number of prior episodes and irrespective of any past treatment (1 study).

1) All three studies that assessed the association between a history of antidepressant medication and response to antidepressants in the primary studies reviewed (albeit one study assessed response to ECT combined with antidepressants) reported that such a history was associated with worse treatment outcomes (44,69,81).

There were mixed findings regarding the number of past episodes:

2) An IPD of antidepressant trials found no association with treatment outcomes (66), another IPD found a weak association with outcomes from treatment with Duloxetine (72), and a third IPD using similar studies of Duloxetine found no association with treatment outcome (74). A fourth study found a very weak association between the number of past episodes of depression and the outcome from antidepressant treatment, based on one primary study (84).

3) A study of naturalistic cohorts found two primary studies that reported a weak association between a history of depression and poorer prognosis (73).

From these studies we might conclude that a history of past failure to respond to antidepressants is associated with a lower probability of response to antidepressants in the present. Evidence for associations with the number of past episodes of depression and treatment response seems inconclusive, and there was a lack of

studies assessing a history of depression and prognosis irrespective of past treatment.

Age at initial onset of depression

Three of the reviewed studies assessed the association between the age at initial onset and prognosis, overall the findings were inconclusive:

8) One study found this not to be associated with response to ECT (69), a second found it to be moderately important in the response to SSRIs but not associated with response to Duloxetine or placebos (72), and a third study found no association between age at onset and response to antidepressants but did find an association with placebo response (66).

Comorbid substance use disorders

In addition, one study found an association between comorbid substance use disorders and prognosis with antidepressant treatment (84), but that was based on one primary study that investigated this association.

Social support

Four of the reviewed studies found associations between social support and prognosis.

1) Three of these studies investigated this in community samples (68,73,82). The generalisability of the findings are further limited as one review combined effects for various types of prognostic outcomes (73), and there was uncertainty of the sample size, setting and measures used to determine outcomes in the third of these reviews (68).
2) A fourth study found an association between social support and response to antidepressants in four primary studies (84).

The evidence for an association between social support and prognosis is limited but all studies that have assessed this reported that higher social support was associated with better prognosis, see Table 1.1 for details.

Life events

Only two reviews reported on the association between life events and prognosis, both assessed the “natural course” of depression and also assessed prognosis

irrespective of treatment, and both found some limited evidence that life events recent to the onset of depression were associated with poorer prognosis (73,83). The degree of evidence for an association with prognosis is hard to assess as one of these reviews included just one primary study that investigated this association, and that was in the context of adults who also had low social support (73), and the study also had a high degree of dropout so findings may have been subject to selection biases (96). The other review included studies of people with other mental health disorders (e.g. Schizophrenia) and a number of the included studies were cross-sectional studies so reverse causality could not be ruled out (83).

Socio-demographics and Long-term health conditions

Ten reviews assessed prognostic associations with socio-demographics, specifically: age (10 studies), gender (7 studies), race or ethnicity (5 studies), marital status (3 studies), employment status (3 studies), socio-economic status (3 studies), and educational attainment (2 studies).

Age

1) Overall the evidence for an association between age and prognosis with particular treatments was mixed: one study-level meta-analysis found higher effect sizes for psychological therapies in a subgroup of trials of students compared to the effect size in studies of older adults (86). Similarly, a study-level meta-analysis of studies of IPT for perinatal women found studies with a higher mean maternal age showed lower effect sizes (85). However, two other reviews (one IPD and one study-level meta-analysis) of psychological therapies found no association between age and outcomes (70,79). Two reviews of antidepressant treatment response (72,84) also found associations between prognosis with those treatments and age, one of which reported age to be the strongest predictor of outcomes with placebo (72), but the direction of the association was unclear due to the use of a complex machine learning model to determine the presence of the association. An IPD of antidepressants found that prognosis with antidepressants and placebo improved as age at baseline increased (66) although another IPD of antidepressant treatments found the opposite effect (81).

Gender

1) There were similarly mixed findings regarding gender with four studies finding that it was associated with prognosis (66,70,72,84), at least in some groups, and a further four studies finding no association (51,79,81,86).

Ethnicity

1) One study found that ethnicity was a significant predictor of outcomes with Duloxetine but it was not considered an important predictor relative to others such as age (72), and the two other studies to assess ethnicity found that it was not associated with treatment outcomes (51,86).

Marital status

1) Perinatal women that were married were found to have better response to IPT than unmarried women in one review (85), marital status was not associated with outcomes from low-intensity CBT in an IPD study (79), although there was an association between marital status and outcomes from antidepressant treatments in one study-level systematic review (84).

Employment status

1) Employment status was among the top 25 predictors of treatment response in one IPD study of antidepressant treatment response (44) but not in another IPD study of response to low-intensity CBT (79). There was reported to be strong evidence for an association of employment status and outcome from antidepressant treatments in one study-level review (84), but that finding was based on just a single study of 542 adults treated with antidepressants.

Socio-economic status

1) There was reported to be strong evidence for an association of socio-economic status and outcome from antidepressant treatments in one study (84) and with the course of depression without treatment in another review (73). An IPD study found that the effect size of CBT may have been higher in patients with disability benefits compared to those without disability benefits, but this was based on just 34 participants with the exposure of interest (88).

Educational attainment

1) There was some evidence that educational attainment was associated with prognosis in the two studies to assess this. One was a study-level systematic review (73) that found adults with more years of education had a better course of depression or shorter duration of depressive episodes irrespective of

treatment. The other was a non-systematic IPD study that found educational attainment to be among the top 25 predictors of response to antidepressant medications, though as noted above, the nature and direction of the effects was difficult to determine as the associations came from a complex machine learning variable selection and separate machine learning model building process (44).

With the exceptions of age and gender, the findings regarding the association between socio-demographic factors and prognosis were based on few primary studies and so the level of evidence for their associations is limited. A more thorough investigation of such associations is warranted.

Long-term health conditions

Three reviews reported on the association between long-term health conditions and prognosis.

1) One systematic review reported that there were isolated findings regarding particular comorbidities such as heart disease and diabetes, which did not replicate across the six primary studies reviewed therein (84). All six of those studies were said to be of moderate or low quality.
2) A meta-review reported weak evidence for the absence of an effect of comorbid long-term health conditions overall on the prognosis with pill placebos (51) although this was based on one primary review which itself reported a non-significant trend towards an effect (92).
3) That latter review was a study-level meta-analysis that found higher effect sizes for a variety of antidepressants in RCTs that had long-term health conditions as an inclusion criterion compared to studies that did not (92). However, the interpretation of this association is not straightforward because an unknown number of the participants in the trials that did not have long-term physical health conditions ("Axis-III conditions") as an inclusions/exclusion criterion would have also had such conditions.

Overall, it would appear that IPD might be required to assess the association between long-term health conditions and prognosis in greater detail, and that such an association should be assessed independent of treatment.

Other factors

A number of other factors were each assessed in one review, they will not be considered further in this thesis as the implications of these studies suggest that there is either insufficient evidence of an effect of those factors on prognosis, or there is a lack of immediate impact of these factors for adults with depression presenting in primary care as the factors are not currently measurable at scale in primary care settings. I will however detail the findings regarding these factors here, and further details are given in Table 1.1:

1) A systematic review of pre-dominantly neuro-cognitive predictors of treatment response found eight RCTs reporting on prognostic associations of such factors with response to psychological interventions of various sorts (mostly CBT or IPT) (87). This review found very limited evidence (coming from a single study with a sample of 122 participants) for associations between connectivity in the subgenual cortex, prefrontal cortex and midbrain regions and response to CBT, Duloxetine, or Escitalopram (87); even more limited evidence (coming from a single study of 38 participants) for an association between metabolism in six regions, most notably the right anterior insula, and response to CBT or Escitalopram; and very limited evidence (coming from one primary study of just 24 participants) that increased metabolism in the inferior temporal cortex and decreased metabolism in the posterior cingulate was associated with response to CBT compared to Venlafaxine. Finally, that review reported no association between polymorphisms of the brain-derived neurotrophic factor gene and prognosis.

There were several methodological issues with this review which affect the implications of the findings: the authors reported on 51 RCTs but only eight of the studies reported on biomarkers as predictors of treatment response, so this greatly reduced the power to detect effects. Nearly all of the studies included in the review themselves had small samples of fewer than 50 patients in each arm, making it difficult to interpret the inconsistencies found in the review. In addition, the primary studies had a small range of treatments, the vast majority included CBT or IPT as the psychological interventions and waitlist or a non-active control as the comparison

condition, so it is possible that any findings maybe specific to those treatments. There were also significant risks of bias in many of the studies, and 35 of the studies were not primarily studies of depression but instead were studies of physical health conditions, whether chronic (e.g. diabetes) or acute (e.g. stroke), so again the generalisability of the findings to most depressed patients is questionable.

2) Another systematic review and meta-analysis combined results from a case control study of inpatients with single episode MDD, recurrent depression or bipolar disorder, with randomly selected healthy controls, and with participants of two RCTs: 672 patients treated with Escitalopram or Nortriptyline in GENDEP (97) and 980 patients treated initially with Citalopram in STAR*D (98). In the meta-analysis no evidence was found for an association between polymorphisms of the leptin gene and response to the above antidepressant medications (90). Such associations were found in the exploration and replication samples from the case control study as were associations between lower leptin serum levels and reduced leptin mRNA expression with poorer treatment outcomes.

3) A systematic review and study-level meta-analysis was conducted on 14 RCTs which included 1035 patients treated with Venlafaxine to consider the associations of three phenotypes of the CYP2D6 metaboliser gene on treatment response. The study found no evidence for associations between phenotypes and response to Venlafaxine in terms of depressive symptom improvement or in terms of adverse events (89). However, the review included primary studies of healthy controls and patients with MDD and did not separate these participants in the meta-analyses, and one included study had a larger sample size than 12 of the others combined, making it difficult to interpret between-study heterogeneity.

4) Another systematic review aimed to assess the associations between neural connectivity in any brain regions and prognosis irrespective of treatment. However, just four studies were included, they all had very small sample sizes, all assessed different regions of interest, and they included patients

with either depression or anxiety disorders with too few participants to determine disorder specific effects (67).

5) A further systematic review of studies that used the temperament and character inventory included ten primary studies of treated and non-treated samples and assessed the “natural course” of depression or prognosis irrespective of treatment in a study-level meta-analysis (75). The review authors reported that harm avoidance temperament was associated with prognosis in treated samples, and other forms of temperament were not. However, six of the ten studies with treated samples were case control studies and the results were presented for the whole studies in meta-analyses, not just the MDD cases, and some of the primary studies had small samples (e.g. n=35). In addition, the authors included results from studies with endpoints varying between six weeks and two years post-baseline, making it difficult to interpret sources of heterogeneity and to consider the implications of their findings.

6) A final review of prospective studies that reported associations between positive or negative emotionality and the course of MDD, found between six and 22 primary studies reporting on facets of emotionality (91). The review authors reported associations between low positive emotionality and poorer MDD course, as well as lower levels of state negative emotionality and higher trait negative emotionality with poorer MDD course. However, the results are difficult to interpret because the authors combined a variety of ways or measuring emotionality (e.g. heart rate, skin conductance and self-report), different measures of depressive symptoms, different MDD course outcomes (e.g. remission, relapse, recurrence), and different time intervals (e.g. ranging from 2-36 months in analyses of state emotionality), all in the same analysis.

Summary and Aims for this Thesis

The above review suggests that depressive symptom severity is probably associated with prognosis but as no systematic reviews or meta-analyses assessed patient characteristics associated with prognosis independent of treatment, whether this is generalisable to all treatments is yet to be determined, as is the magnitude of any

such association. On balance, most reviewed studies found that longer durations of depression at baseline were associated with worse outcomes, as were comorbid anxiety disorders, or particular types of anxiety symptoms. There was more limited evidence regarding other indicators of 'disorder severity' so further studies of these factors will be reviewed in Chapter 3. Those reviewed here suggested that a history of depression maybe associated with poorer outcomes, particularly if that history involves failure to respond to antidepressant treatments and the current treatment being received involves antidepressants. There was only one review and one primary study within it that assessed associations with comorbid substance abuse and prognostic outcomes, which is somewhat surprising given the high rates of comorbidity between the two (99,100).

It seems possible that social support is associated with prognosis although there was limited evidence with which to assess such an association, and so further studies of social support will be reviewed in Chapter 4. There were mixed findings regarding life events and few studies to have assessed the association (so further studies of life events will be considered in Chapter 5).

Of the socio-demographic factors assessed only age and gender have been well studied and there were mixed findings regarding both, so further studies of these two factors and the other socio-demographics will be considered in Chapter 5. Long-term health conditions were only reported on in three reviews, one of which was a meta-review and included another of the three reviewed here, and the evidence of an effect on prognosis was inconclusive due to a number of methodological limitations to the reviewed studies. A number of other factors were considered in one review each though these appear to have limited utility and generalisability to adults with depression in primary care, so will not be focussed on further in this thesis. It is noteworthy too that none of the included reviews considered the association between personality disorders or traits and prognosis. Such factors have been found to be associated with poorer treatment outcomes in IAPT among patients with either depression or anxiety conditions (101,102) and associated with a greater risk of relapse (40).

In light of the above, further to the overall aim of this thesis stated above (and re-stated here: overall aim: to investigate factors associated with prognosis for adults with depression independent of treatment, in primary care), the following sub-aims are outlined below.

Sub Aims:

1) To determine whether and the degree to which depressive symptom severity is associated with prognosis for adults with depression independent of treatment, in primary care.
2) To determine which depressive 'disorder severity' factors (of those listed in the 'disorder severity' subsection above) pre-treatment are associated with prognosis independent of depressive symptom severity and independent of treatment.
3) To determine whether other factors including social support, life events, socio-demographics (age, gender, ethnicity, marital status, employment status, financial wellbeing, housing status, and the highest level of educational attainment) and long-term health condition status are associated with prognosis independent of symptom severity, 'disorder severity', and treatment.
4) To determine whether all of the above factors are associated with attrition from studies, independent of treatment and independent of depressive symptom severity.

Chapter Summary and Next Steps

In summary, this chapter has shown that depression is both prevalent and disabling, that a number of treatments might be helpful for some patients; at a population level many treatments might be considered to be approximately equally effective, but even when received as recommended, they are not effective for large proportions of adults with depression. We know relatively little about what confers risk for poor prognoses aside from depressive symptom severity before starting treatment, and so are currently ill equipped to accurately determine prognosis for an adult seeking treatment for depression. We also have little to inform clinicians of what factors they might assess for routinely to inform prognosis for their patients. In particular, which factors might be most valuable in that regard after accounting for depressive

symptom severity, which clinicians routinely measure in some form or other. I have discussed the importance and uses for understanding prognosis and have highlighted the importance of investigating prognosis independent of treatment to learn more about general prognostic factors and to understand indicators of prognosis beyond symptom severity. This chapter has also briefly discussed why individual patient datasets which include studies with a variety of treatments might provide the best opportunity to study prognosis in that context, and therefore might be required in order to meet the aims outlined above. How such an IPD dataset might be developed, the methodological considerations when compiling and analysing such data, and a systematic processes for forming such an IPD will be discussed in greater depth in Chapter 2. The findings from analyses of the IPD dataset formed as part of this thesis are presented in Chapters 3-5.

Chapter 2: Methodological considerations for assembling and analysing IPD datasets

Overview

In Chapter 1 I discussed some of the methodological weaknesses of study-level meta-analyses in investigating indicators of prognosis, and argued that one promising avenue for such investigations lies in forming a dataset from individual patient data (IPD). Further, I discussed the importance of investigations of factors associated with prognosis being conducted independent of treatment. Such studies would benefit from an IPD dataset including a breadth of trialled treatments and a consistent means of measuring both depressive symptom severity and depressive 'disorder severity' factors. In Chapter 2 I will discuss some of the methodological issues that arise when forming an IPD dataset and in analysing IPD, will go on to describe how the Depression in General Practice (Dep-GP) IPD dataset was formed, and will discuss the attempts made to account for such methodological issues. I will finish the chapter with a protocol for a series of analyses of the Dep-GP IPD which will form the basis for the three succeeding chapters of this thesis.

Benefits of IPD datasets and findings from IPD datasets on prognosis of depression

As outlined in Chapter 1, identifying who is most likely to experience a good or bad outcome from treatment is fundamental to reducing the burden of depression (30,64,103). Although, randomised controlled trials (RCTs) and meta-analyses of RCTs had become the gold-standard for determining treatment effects in medicine (104), single trials usually lack the power to explore patient characteristics that can help predict prognosis (64), and study-level meta-analyses are vulnerable to a number of biases. In particular, ecological biases or the 'ecological fallacy' (62,77,78), reporting and publication biases, and biases introduced by the loss of important data due to differing approaches to missing data across trials, or the loss of unreported intermediate endpoint data (62). Further, both RCTs and study-level meta-analyses can be legitimately criticised for lacking robust methods to consider subgroup analyses that might allow findings to be generalizable to a greater range of patients (57–59). In contrast, meta-analyses that use the data from all of the individual participants in each trial (IPD) and particularly those IPD that represent all the trials in a given population, are more adequately powered for sub-group analyses, and do not suffer from most of the biases discussed above (77,80). Such IPDs can therefore be powerful tools to investigate areas of heterogeneity and to consider indicators of prognosis independent of treatment (62,77,80,105,106), and may therefore be considered the new gold-standard for such investigations (107,108).

IPD datasets have become increasingly popular as interest has grown in personalized medicine. Understanding a range of factors, particularly when in combination, often requires considerably larger sample sizes than might be feasible with individual studies, and a richness of data not available when using study-level summary statistics (61,108). Health research funders are increasingly demanding greater collaboration or publishing data 'open-source' to maximize the impact of the research they fund, and computing hardware and software have improved to the point of significantly reducing the barriers to conducting complex analyses on large datasets. This has given rise to a great number of IPD datasets being established and to an increased range of research questions and analytical techniques applied in IPD studies (61). However, while IPDs have great potential to be used to investigate

prognosis, there are a number of issues that can still introduce significant degrees of bias, can affect the generalisability of findings, and issues with the way that they are sometimes analysed. This has led one group of authors to call the analytical approaches in some IPD studies either “deft, daft or deluded” (64). IPD studies are therefore not without their pitfalls and constraints. Below I will discuss a number of these issues and outline how I have attempted to avoid or mitigate them when setting up an IPD dataset of studies of depression that recruited participants in their GP practices: “the Dep-GP” IPD dataset.

Setting up the Dep-GP IPD dataset

Objective

The objective was to set up an IPD dataset that could be analysed to address the aims of this thesis. As the setting was unclear for many of the clinical populations studied in the reviews (and the primary studies reviewed therein) outlined in Chapter 1, potentially limiting the generalisability of their findings, and as the majority of depressed patients present for treatment initially to primary care physicians or general practitioners (23), it was decided that this IPD dataset would focus on studies that recruited in primary care alone. Guidance on how to set up an IPD dataset avoiding some of the methodological difficulties (detailed above) is limited, but prior research has recommended that for an IPD dataset the identification and selection of studies should be rigorous as would be the case with a systematic review (57,106,107), and that every effort should be made to include all eligible trials (59,108).

Over the last decade it has been commonly stated that the heterogeneity of depression as a diagnosis is stifling research that could greatly improve understandings of prognosis (109). Some authors propose that given the variety in symptoms experienced by those meeting diagnostic criteria for depression, there are as many as 1030 ‘variations’ of major depressive disorder (110), of which 166 variations can be considered likely to present in clinical practice (110,111). Given this, further to the above recommendations, if an IPD dataset is to be useful in determining indicators of prognosis that are generalizable to large proportions of adults with depression, the studies therein should have used a uniform method for measuring a broad range of symptoms that might reflect the proposed ‘variations’ of

major depression. It will also be important that such an IPD has used uniform means of diagnosing the severity of depression. As such, the IPD would need to focus on studies that used the same method to derive diagnoses. The use of diagnostic schedules such as the Clinical Interview Schedule Revised (CIS-R) (112) could achieve this. The CIS-R can be used in a self-administered computerised format that measures depression and anxiety disorder symptoms (excluding posttraumatic stress disorder), calculates primary and secondary diagnoses in accordance with the ICD-10 (113), and has been used in a number of primary care and community based RCTs, and in large epidemiological studies (23,114).[1]

Identification and Selection of Studies

The search strategy to identify studies was refined over several iterations after running scoping searches. From running the scoping searches it became clear that the most commonly used comprehensive measure of the variety of prognostic indicators of interest used in RCTs in primary care was the Revised Clinical Interview Schedule (CIS-R) (112). Several studies used other measures such as the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (115) or the Structured Clinical Interview for DSM (SCID) (116), but these were used considerably less often. It was more common that only the depression module from the SCID was used. As such, it was decided that searches would be restricted to studies that used the CIS-R. Searches were not restricted to any particular type of RCT, so both pragmatic and explanatory RCTs (117) could be included.

Studies were identified using a combination of keyword and subject heading searches on the bibliographic databases below, hand-searching through references of studies identified in the searches, and by contacting experts for unpublished or missed studies. Searches were run on the Cochrane CENTRAL Trial Register (searched on 20th March 2019), Embase 1947 to 2019 Week 12, International Pharmaceutical Abstracts 1970 to March 2019, Ovid MEDLINE 1946 to March Week 3 2019, and PsycINFO 1806 to March Week 3 2019. Search terms included variations of phrases such as “depression” or “major depression”, “RCT” or “Randomised Controlled Trial” or “Clinical Trial”, and as detailed in Chapter 1, in

[1] Details in the following section are set out in a protocol paper: Buckman et al., 2020, pp.3-19.

order to achieve the needed uniformity in measurement of depressive “disorder severity” factors, a further condition for the searches was that studies returned should include terms such as “CIS-R” or “Clinical Interview Schedule”. Full details of the searches are provided in Appendix 1.

A single reviewer (JB) screened titles and abstracts of potentially eligible studies, these were then read in full and judged against inclusion/exclusion criteria by two reviewers (JB and GL) with consultation with a third (SP) to resolve any uncertainties by consensus.

Inclusion & Exclusion Criteria

Studies were included if they:

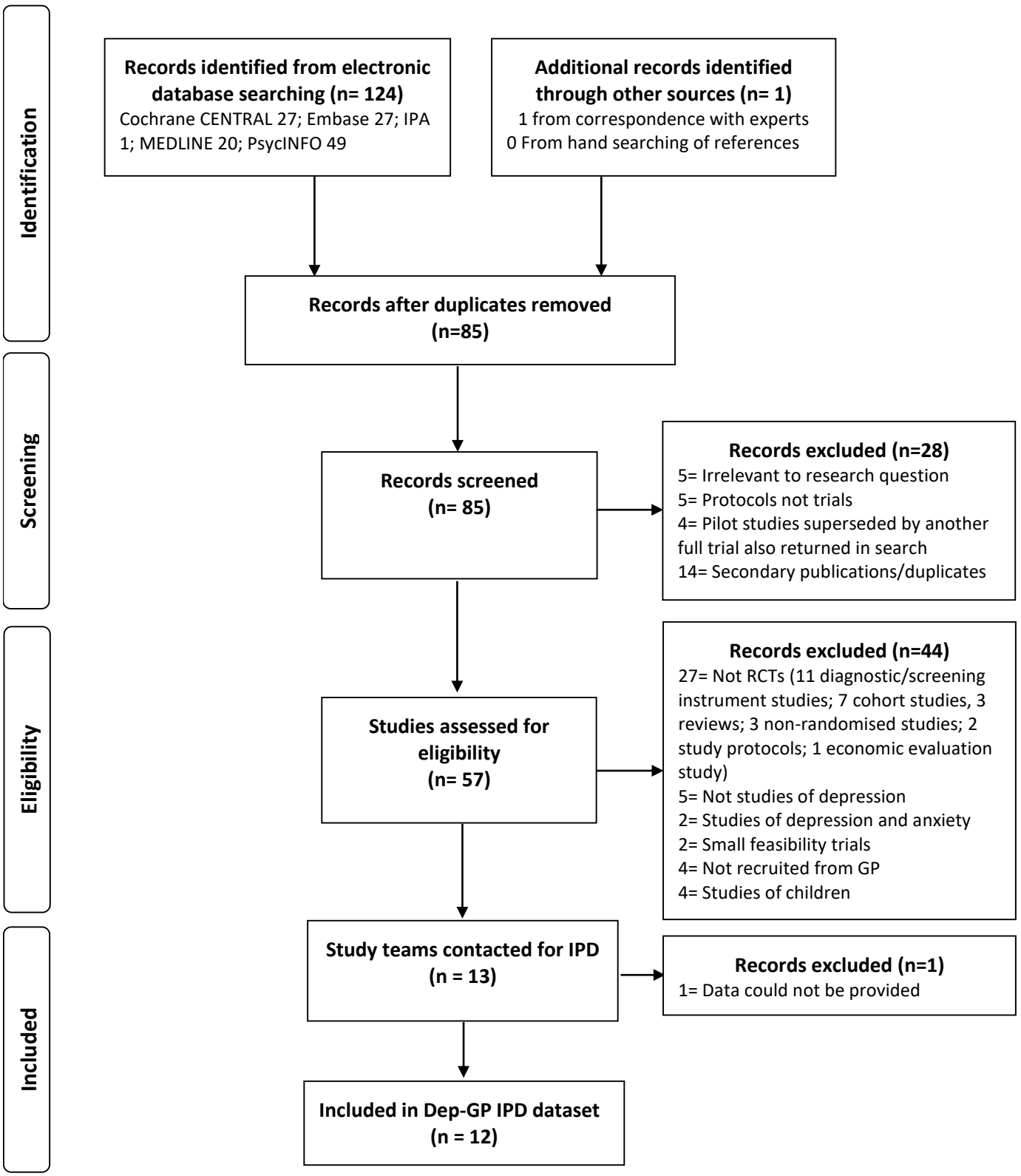
- Were randomised clinical trials (RCTs) of adults (aged 16 or over) with unipolar depression, or with depressive symptoms significant enough for them to seek treatment, or a CIS-R score of ≥12.
- Recruited from primary care centres.
- Had at least one active treatment arm.
- Used the CIS-R at baseline to measure symptoms and to determine diagnoses.

Studies were excluded if they: did not meet the above criteria and if they:

- Included patients with depression as a secondary diagnosis in studies of adults with personality disorders, psychotic conditions, or neurological conditions.
- Were studies of adults with bi-polar disorder or psychotic depression.
- Were studies of children or adolescents.
- Were feasibility studies only.
- Did not recruit participants from primary care.

For the analyses presented in this thesis studies were also excluded if they were trials of adults with either depression or an anxiety disorder, rather than a primary depression with or without comorbid anxiety.

Figure 2.1. Flow diagram of study selection.



Characteristics of the included studies

Thirteen RCTs (n=6175) were identified as meeting inclusion criteria for the IPD, of which 12 provided individual patient-data, the remaining study (n=151) was completed over 20 years ago and the data were no longer available (118), see Figure 2.1. A description of each study can be found in Table 2.1 and descriptive statistics and degrees of missing data for key predictor and outcome variables discussed below are presented in Appendix 2.

Table 2.1 Description of studies included in the Dep-GP IPD dataset.

Study	N	Pragmatic RCT (Y/N)	Inclusion criteria Mean (SD)	Age Mean (SD)	Gender % Female	T0 Depressive Symptom Severity Mean(SD)	Remission* % at 3-4 months	Interventions	Outcome Measure Primary (additional)
AHEAD (119)	327	Y	Adults with new depressive episodes diagnosed by GP	43.1(15.4)	67%	HADS depression=10.5(3.9)	62%	TCA vs SSRI vs Lofepramine	HADS (CIS-R)
CADET (120)	527	Y	Adults ≥18, ICD-10 Depressive Episode	44.4(13.2)	72%	PHQ-9=17.7(5.1)	41%	Collaborative Care vs TAU	PHQ-9
COBALT(121)	469	Y	Adults 18-75 with treatment resistant depression, scoring ≥14 BDI-II	49.6(11.7)	72%	BDI-II=31.8(10.7)	34%	CBT+TAU vs TAU	BDI-II (PHQ-9)
GENPOD(122)	601	N	Adults 18-74 with depressive episode	38.8(12.4)	68%	BDI-II=33.7(9.7)	41%	Citalopram vs Reboxetine	BDI-II (HADS)
HEALTHLINES (123)	609	Y	Adults ≥18, PHQ-9 score ≥10, confirmed diagnosis of depression with CIS-R, internet access	49.5(12.9)	69%	PHQ-9=16.9(4.6)	30%	Healthlines telecare + TAU vs TAU	PHQ-9
IPCRESS(124)	295	Y	Adults scoring ≥14 BDI-II and GP confirmed diagnosis of depression	34.9(11.6)	68%	BDI-II=33.2(8.8)	34%	iCBT+TAU vs TAU + waiting list for iCBT	BDI-II
ITAS(125)	798	Y	Adults ≥16, scored ≥12 on CIS-R	43.2(14.8)	68%	GHQ=7.7(3.2)	N/A; at 6-8 months 46%	Recommendation + TAU vs TAU	GHQ-12
MIR(126)	480	Y	Adults ≥18 taking SSRIs or SNRIs at adequate dose for≥ 6 weeks, and scored ≥14 on BDI-II	50.7(13.2)	69%	BDI-II=31.1(9.9)	30%	Mirtazapine vs Placebo	BDI-II (PHQ-9)
PANDA(127)	652	Y	Adults presenting with low mood or depression to GP in last 2 years, free of ADM for 8 weeks up to baseline	39.7(15.0)	59%	BDI-II=23.9(10.3)	69%	Sertraline vs Placebo	PHQ-9 (BDI-II)
REEACT(128)	685	Y	Adults with PHQ-9≥10 presenting to GP with depression	39.9(12.7)	67%	PHQ-9=16.7(4.3)	53%	Moodgym vs Beating the Blues vs TAU	PHQ-9
RESPOND(129)	220	Y	Women meeting criteria for MDD within 6-months post-partum	28.7(6.4)	100%	EPDS=17.6(3.4)	56%	ADM vs Listening intervention	EPDS
TREAD(130)	361	Y	Adults 18-69 who met diagnostic criteria for MDD and scored ≥14 on BDI-II	39.8(12.6)	66%	BDI-II=32.1(9.2)	35%	Physical Activity + TAU vs TAU	BDI-II

Abbreviations: ADM – antidepressant medication; BDI-II – Beck Depression Inventory; EPDS – Edinburgh Postnatal Depression Scale; GHQ-12 – General Health Questionnaire 12 item version; HADS-D – Hospital Anxiety and Depression Scale – depression subscale; iCBT (internet based therapist delivered cognitive behavioural therapy); MDD – Major Depressive Disorder; T0 - Baseline; TAU – treatment as usual; TCA – tricyclic antidepressant

* definitions of remission in each study are given in Table 2.2 below

Measures used in included studies

The relevant measures of symptoms, key potential indicators of prognosis, covariates, or outcomes included in the identified studies are outlined in Table 2.1.

Table 2.2. Measures used across the studies of the Dep-GP IPD database.

Measure	Details	Scores and Cut-offs for Remission
The CIS-R (112)	Consists of 14 symptom subsections scored 0-4 covering core features of depression, depressive thoughts (scored 0-5), fatigue, concentration/forgetfulness, and sleep, generalized anxiety, worry, irritability, obsessions, compulsions, health anxiety, somatic concerns, phobic anxiety (split into agoraphobia, social phobia, and specific phobia), and panic. A final section measures general health, impairment and weight change.	The total score ranges from 0-57 with a cut-off of ≥12 used to indicate likely common mental disorder, primary and secondary diagnoses using ICD-10 criteria are given as are binary indictors of diagnosis for all the disorders assessed. The duration of each type of problem is also assessed for the present episode (or subsyndromal episode) upto the point of completing the CIS-R. Duration items are measured in five categories: 1) less than two weeks; 2) between two weeks and six months; 3) between six months and one year; 4) between one and two years; and 5) more than two years.
Beck Depression Inventory 2nd Edition (BDI-II) (131)	Consists of 21 items to assess depressive symptoms, each item is scored 0-3.	There is a maximum score obtainable of 63, and a cut-off of ≥10 is used indicate significant symptoms of depression, scores of <10 are therefore used to indicate remission in those that were previously depressed/scored ≥10.
Patient Health Questionnaire 9-item version (PHQ-9) (132)	This is a depression screening measure, with respondents asked to rate how often they have been bothered by each of the nine symptom items over the preceding two weeks. Each item is scored 0-3	There is a maximum score of 27 with a cut-off of ≥10 is used to indicate "caseness" for depression, a score of 9 or below for those that were previously depressed is therefore considered to indicate remission
Hospital Anxiety and Depression Scale (HADS) (133)	Measures symptoms on two subscales, depression and anxiety, each with 7 items scored 0-3. Respondents are asked to endorse a statement relating to frequency or severity of problems in each symptom domain over the preceding 7 days	A total score of 21 is obtainable on each subscale, with a cut-off for caseness on the depression subscale of ≥8. Scores <8 are therefore used to indicate remission
General Health Questionnaire (12-item version) (GHQ-12) (134)	Consists of 12 items related to present and recent health over the "few weeks" prior to completion. Each item is related to depression or generalised anxiety, they are scored 0-0-1-1 for the four response options.	A cut-off of ≥2 is used to indicate the likely presence of common mental disorder, and so scores of <2 for those formally scoring above this would be considered to indicate remission
Edinburgh Postnatal Depression Scale (EPDS) (135)	This measures symptoms of depression among women in the post-natal period. It consists of 10 items relating to symptoms of depression, each one is rated in relation to feelings over the week prior to completion. Each item is scores 0-3.	The maximum obtainable score is 30, with scores of ≥13 are indicative of a depressive episode, and scores of <13 indicative of remission among the formally depressed.
Generalised Anxiety Disorder Scale, 7-item version (GAD-7) (136)	This measures symptoms of generalised anxiety with the same scaling and structure of questions as used in the PHQ-9.	A maximum score of 21 is obtainable across the 7 items. A cut-off of ≥8 is used to determine 'caseness' for GAD.
Social Support Scale - adapted for use in Adult Psychiatric Morbidity Surveys from the Health and Lifestyles Survey (137)	An 8-item instrument, the first seven items of which come from the Health and Lifestyles Survey assessing the degree to which participants rated the social support of their friends and family in each of the following domains: 1) being accepted for who one is; 2) feeling cared about; 3) feeling loved; 4) feeling important to them; 5) being able to rely on them; 6) feeling well supported and encouraged by them; 7) being made to feel happy by them; and 8) feeling able to talk to them whenever one might like. Items are scored 1-3, with total scores ranging from 8-24; higher scores indicate higher levels of perceived social support. The authors of the Health and Lifestyles Survey suggested the maximum score for social support (which was 21 on that scale) indicated 'no lack of social support' scores between 18-20 indicated a 'moderate lack of social support' and scores of 17 or below indicated a 'severe lack of social support'	N/A
Life events: the **Social Readjustment Rating Scale** (138)	Participants are asked to say yes/no to whether they have suffered any of eight events within the last six months e.g. a death/bereavement; being physically attacked/injured; or going through a divorce/separation. Each item is scored yes (1) or no (0) and the total score is the sum of all the items.	N/A

Alcohol use: the **alcohol use disorder identification test primary care version (AUDIT-PC)** (139).	Used to assess alcohol misuse, this includes five items scored 0-4. A cut-off of ≥5 indicates hazardous alcohol use that may be harmful to one's health	N/A
Health related quality of life: EQ-5D-3L & EQ-5D-5L (140).	The EQ-5D is a generic measure of health status in five domains – mobility; self-care; usual activities; pain/discomfort; and anxiety/depression. Each domain in the 3L version has three response categories ranging from no problem present (1) to extreme problems in the given domain (3), the 5L version has five response options ranging from "I have no problems…" (1) to "I am unable to…" or "I have/am extreme/extremely…" (5). A total score is derived from summing the score on the five items with higher scores indicating more severe health problems than lower scores. A cross-walk of scores from the 3L and 5L versions will be used to derive a continuous index score representing the EQ-5D total score in the present study (141).	N/A

CIS-R was used in all 12 studies, for depression subscale scores and durations n=5686, for anxiety scores n=5415, for anxiety durations and individual diagnoses n=5088. BDI-II was used in 6 studies (COBALT, GENPOD, IPCRESS, MIR, PANDA, & TREAD), n=2858 ; PHQ-9 was used in 6 studies (CADET, COBALT, HEALTHLINES, MIR, PANDA, & REEACT) n=3416 ; HADS was used in two studies (AHEAD & GENPOD) n=925; GHQ was used in ITAS only n =796; EPDS was used in RESPOND n=220 ; GAD-7 was used in 5 studies (CADET, COBALT, HEALTHLINES, MIR & PANDA) n=2110; the Social Support Scale was used in 6 studies (COBALT, GENPOD, IPCRESS, MIR, PANDA, & TREAD) n =2858; the Life Events, Social Readjustment Rating Scale was used in 7 studies (COBALT, GENPOD, IPCRESS, ITAS, MIR, PANDA, & TREAD) n=3656; the AUDIT-PC was used in 6 studies (COBALT, GENPOD, IPCRESS, MIR, PANDA, & TREAD) n=3028 ; EQ-5D was used in 8 studies (AHEAD, CADET, HEALTHLIENS, IPCRESS, MIR, PANDA, REEACT, & TREAD) n=3931.

Methodological issues that arise with IPD datasets and how these were dealt with in Dep-GP

Once studies have been identified as eligible for inclusion in an IPD dataset there are a number of important considerations that will affect the utility and validity of findings from that dataset (61). The formation of IPD datasets is incredibly resource intensive requiring money, time and commitment from a group of researchers (46,59), and can be hampered by requiring data from studies that finished many years ago, making retrieval of data challenging and sometimes impossible (59). As will be discussed below there are a variety of other methodological issues that need to be considered when compiling an IPD dataset, particularly regarding agreements on data-sharing, access to cleaned IPD data, publication policies, and how data are managed and stored, cleaned, harmonized, and checked for integrity prior to analysis. All of these could have a significant impact on the validity of analyses of the Dep-GP IPD dataset to achieve the aims of this thesis laid out in Chapter 1.

Data sharing and access agreements and publication policies

Guidelines on the compilation and reporting of IPD datasets e.g. the preferred reporting items for systematic reviews and meta-analysis of individual patient data (PRISMA-IPD) (108) highlight the importance of data sharing and access agreements, and of publication policies. The suggestion is that such agreements should be set up prior to access to any data in order to maintain successful collaborations and interest from the many involved parties (46,59,61). Following guidance on the compilation of such agreements (142), all chief investigators of the studies identified as meeting inclusion criteria for the Dep-GP IPD dataset (Figure 2.1) were written to and asked to confirm any specific stipulations they would require to go into a data sharing agreement or on the publication policy, and all parties were sent a proposed data management and analysis plan setting out the framework for the use and management of the data (see Appendix 3). Policies were redrafted to account for stipulations from any of the relevant parties, after which all parties (chief investigators, nominated principal investigators, sponsors and collaborators) signed and agreed to be bound be the terms of a revised data-sharing agreement (Appendix 3). In addition, some studies required additional contracts to be signed which would stand in conjunction with the data-sharing agreement (also see Appendix 3).

The basic terms of the data-sharing and publication policy agreements were that:

1) All data would be treated in accordance with the Data Protection Act 1998, data would be pseudonymised, stored securely and treated as confidential.

2) In conducting this study the Sponsor, UCL would comply with all laws and statutes, as amended from time to time, applicable to the performance of research studies with human participants including, but not limited to: The Human Rights Act 1998, The Data Protection Act 1998, The Freedom of Information Act 2000, The Research Governance Framework for Health and Social Care, issued by the UK Department of Health (Second Edition 2005) or the Scottish Health Department Research Governance Framework for Health and Community Care (Second Edition 2006)

3) The studies using data from the IPD dataset would be conducted in accordance with the World Medical Association Declaration of Helsinki entitled 'Ethical Principles for Medical Research Involving Human Subjects' (1996 version) and in accordance with the terms and conditions of the ethical approval given to the trial.

4) All publications based on data included in the IPD would include each Chief Investigator from the included RCTs as co-authors and potentially nominated Principal Investigators where appropriate.

5) All articles to be submitted for publication would be sent to all co-authors for approval prior to submission.

6) Data from the IPD would not be used for any purposes outside of the approved investigations on understanding and predicting prognosis for adults with depression and closely related topics. Any additional analyses would be subject to agreement from the chief investigators of the individual studies.

7) Data from the IPD would not be shared with external parties, any requests for data sharing would be signposted to the chief investigators or custodians of the data for each individual RCT included in the IPD dataset and it is expected

that these requests would then be reviewed as usual by those individuals or teams.

It should be noted that these terms were laid out in advance of the introduction of the General Data Protection Legislation (GDPR) implemented into European Union law in May 2018.

Data Collection Processes

Once the above terms had been agreed to and any additional contractual elements had been drawn up and agreed to, the means and methods of access to the data from each study had to be agreed. One chief investigator had already shared IPD data from one study meeting eligibility criteria for the Dep-GP (AHEAD) with one of the study sponsors, so that chief investigator suggested using those same data and there were therefore no additional requirements on accessing those data. In addition, one of the study sponsors was the chief investigator and custodian of the data for three of the eligible RCTs (GENPOD, ITAS, & PANDA) so agreed to give access to those data without any further requirements too. For all of the other studies bespoke systems for securely accessing the data had to be agreed, this involved the use of data encryption software and the use of online portals for data transfers on a cloud server hosted by the data custodians' universities or clinical trial units, guaranteeing data security under the terms of their own institutions.

Compiling the datasets, harmonizing data, and data integrity checks

PRISMA-IPD requires that details of the compilation and storage of IPD datasets be reported (108), however it has been noted by several authors that in compiling IPD datasets data are often transferred in multiple formats, across various software and so bringing these datasets together into one single IPD dataset is both time consuming and challenging (46,61,63,143). For the Dep-GP IPD dataset study teams sent their data in Stata 15.0 (144), Microsoft Excel, SPSS (145) and text (notepad) formats and most study teams sent multiple datasets containing various sections of their study data. All datasets/files for each study were merged in Stata and then entered into Microsoft Excel to keep a uniform format for original datasets, initial cleaning of each study was performed in Excel using formulae bringing original data into a new "clean" worksheet and then all studies were appended and further cleaned in Stata with a series of 'do files' used to record all of the cleaning steps.

A key stage of compiling the overall dataset involved harmonizing data from each included study (46,63). As an inclusion criterion for Dep-GP was the use of the CIS-R at baseline and as most studies used the same computer programme version of CIS-R with the same demographics and history of depression variables included, little harmonization was required of such data. However, there were some variations: two studies conducted more recently than some of the others (MIR and PANDA) included additional categories for all duration items on the CIS-R, as such these were recoded so the highest categories ("five years or more" and "between two and five years") fitted with the other 10 studies (this involved recoding of the top two categories to be equal to the top category in the other studies "two years or more"). One study only included data on the depression subscales of the CIS-R (HEALTHLINES) so the total score on that measure could not be harmonised with total scores on the CIS-R from the other studies. Harmonizing of the two versions of the EQ5-D questionnaire used in many of the studies was conducted using a validated cross-walk of the total scores on each measure to a single index score (as detailed in the Measures section above).

Of great importance is consideration of whether or not to harmonize (and if so how to do so) across different outcome measures used in the different studies (63). Some authors propose the use of standardised means or z-scores across all different measures (46), others propose the use of cross-walks (143) when appropriate ones exist (e.g. PROMIS for the BDI-II and PHQ-9) (146), and other authors propose the use of multiple imputation techniques to impute missing outcome data across studies, negating the need to harmonize across different outcome measures (63). The choice of which approach to take depends on the outcome of interest and the research question under investigation, as such, the approach taken in Dep-GP varied depending on such factors and is detailed below in discussion of the specific data analysis plans of relevance to this thesis.

Once data were uniformly formatted, all individual trial datasets underwent integrity checks (108), checking all baseline and each reported endpoint variables in each trial against those reported in the publications about each study. This often required splitting the data into the randomised groups and re-categorising some variables to

match those used in the publications. This led to a number of discrepancies which were checked with the chief investigators and where applicable with data managers for each trial. The following issues arose:

1) In AHEAD despite collecting data on the CIS-R as the study team did not use the computerised version only total score and subscale scores were available, not duration items or individual items of the CIS-R. The study team searched archive folders for the original raw data but could not access the missing CIS-R data and so these items were lost.
2) In CADET the publication (120) reported on 581 participants but data for only 527 were received, this was because 54 people were withdrawn/asked to withdraw from the study and so their data were removed at the end of the trial. Numbers of participants in each category of categorical variables and summary statistics were therefore different to those reported by the study team (120) but were largely similar in form and distribution.
3) In ITAS the publication reported that inclusion criteria were being aged 16 or over, being a case on the GHQ-12 and scoring ≥12 on the CIS-R, the study team reported that 762 participants met those criteria (147). The study team sent raw data on 805 participants, and applying their inclusion criteria resulted in a sample of 798. In discussion with that study team it was suggested that for the purposes of Dep-GP we should use data on all 798 as they contributed data at each endpoint. Given the passage of time since this study was conducted the precise reason for the discrepancy in the number of participants could not be determined.
4) In GENPOD there was a slight error in how the study team had calculated the HADS depression score when data were sent, after discussing the formulae used by the study team to calculate this variable we discovered a small error which was able to be corrected and data were resent.
5) In IPCRESS the publication on the study reported 297 participants (124) but two participants were missing data on most variables so were removed, leaving a sample of 295 participants. An additional issue occurred in that the individual items of the CIS-R were not available as they had not been extracted from the original raw files created by the CIS-R computer software programme. This meant that the IPCRESS study team had to send all the raw files for the CIS-R and it was necessary to create a bespoke data-extraction

tool based on those used in other studies but changing the variables to fit IPCRESS. Once this was done it was possible to extract all the necessary data and create the CIS-R individual items. As data on these were not reported in the publications about IPCRESS, integrity checks involved ensuring that total scores in the published analyses matched the total scores derived by adding all items together in the format specified by the CIS-R scoring ("proqsy") programme.

6) In TREAD the BDI-II scores sent were slightly different from summary statistics reported in the publication about this study (130). Inspection of the IPD revealed that 30 participants had scores of zero at baseline and that the total score was not a simple summation of the 21 individual items. In discussion with the study team we identified an error that had occurred when the variables requested for Dep-GP were split off from the larger dataset held by the study team, they were therefore able to correct this error and resend the data, after which the summary statistics matched those in the publication (130).
7) PANDA and MIR were not published when the data were initially sent so the integrity checks involved checking data against the protocols for those studies and then later once such publications were available, checking against those. It was noted that in PANDA there was a change in the stratification by CIS-R score from the data analysis plan for the study to the published protocol (148). The protocol paper reported that the top stratification category would be a score ≥20 on the CIS-R, this was previously listed as >20 in the data analysis plan which was all that was available when the data were initially provided by the PANDA study team.
8) HEALTHLINES were not able to send the CIS-R data in the original data transfer, as with IPCRESS a programme was written to extract data from the text files automatically generated by the CIS-R computer programme during the study. When doing this it was found that only the depression subscales of the CIS-R were used. As the only data from CIS-R reported in the publication about HEALTHLINES (123) was the ICD-10 depression diagnosis determined by CIS-R, the other CIS-R data were not able to be checked against the publication. As with IPCRESS I therefore checked that the sum scores for each domain of the CIS-R with observed data added up to the

expected amount when checked against the CIS-R “proqsy” (scoring) programme.

9) It was agreed with the study sponsors and collaborators that we would remove cases if they had missing data on over 75% of the variables at baseline or were missing all CIS-R variables, this resulted in two patients being removed from IPCRESS and one from PANDA.

After this phase all studies were appended together in Stata and underwent further cleaning and harmonizing to ensure all variables were in the same formats, were usable in analyses, and the same integrity checks were performed for each study in the collated IPD dataset. The data cleaning steps and integrity checks were independently cross-checked by the study sponsors and a collaborator with permission to access the data. Discrepancies in the data-cleaning and integrity checks were highlighted and discussed with the chief investigator for the relevant studies such that errors could be resolved. This resulted in very few changes to the cleaned data but did help highlight two of the errors noted above, including the error on the BDI-II total score in TREAD and the discrepancy in the application of the inclusion criteria noted in ITAS.

Interim Summary

I have discussed the importance of IPD datasets in general, why some authors have argued that they are the “new gold-standard” for investigations of indicators of prognosis, and hence the rationale for forming the Dep-GP IPD dataset to achieve the aims of this thesis. I have outlined some of the methodological considerations of relevance to creating and forming an IPD dataset, and have outlined how these were accounted for in the formation of the Dep-GP IPD dataset. I will now go on to describe methodological considerations relating to the analysis of IPD datasets and how these have been accounted for in the data analysis plans for Dep-GP.

Analysing the Dep-GP IPD to investigate indicators of prognosis independent of treatment

As discussed above, the Dep-GP IPD dataset should lend itself particularly well to research questions focussed on the identification of indicators of prognosis (106),

however there are a number of key considerations regarding the analysis of the Dep-GP data that will greatly influence the ability for these data to be used in this way. To start with, a number of authors have noted the complications of determining primary outcomes and finding a unique time-point for such outcomes in IPD datasets (61,63).

Outcomes

End-point data

Eleven of the studies included in the Dep-GP collected endpoint data between three and four months post-baseline (see Table 2.1). Guidance on the length of treatments for depression suggests that low-intensity psychological treatments should take place over 9-12 weeks, high-intensity psychological therapy over 12-16 sessions spread over 3-4 months, and structured exercise over 12-14 weeks (21). Further, depending on the level of baseline symptom severity antidepressant medications usually take 2-4 weeks to have some effect with most suggesting up to 6-12 weeks to have a clinically meaningful impact (21,32). Given the intention to have a breadth of treatments in the Dep-GP IPD dataset that represent the options in primary care, it was decided that the primary endpoint of interest for the analyses should be a combination of the endpoints at 3-4 months. This also would ensure that the prognostic outcome being assessed was capturing symptoms after acute-phase treatment; later endpoints might include outcomes from maintenance or continuation phase treatments or capture some relapses or even recurrences of depression after earlier remission (149). An additional advantage of having a chronological endpoint (such as 3-4 months post-baseline) as opposed to a time varying one (such as at the end of treatment) is that findings here might have more pragmatic utility. They might inform conversations with patients about their potential prognoses within 3-4 months of presenting for treatment rather than at an either undefined endpoint or one that might vary depending on treatment choices. Additional end-points between six and eight months, and nine and twelve months post-baseline were used for sensitivity analyses (see Table 2.3). Endpoints prior to three months or after 12 months were excluded from the analyses though as detailed below, they were used for the imputation of missing data.

Table 2.3. Endpoints in months and time from baseline in weeks in each study in the Dep-GP database.

	Endpoint (m) and time from baseline in weeks (w)		
	3-4m	**6-8m**	**9-12m**
Study	**12-18w**	**24-32w**	**36-52w**
AHEAD	12w	26w	52w
CADET	16w		52w
COBALT	12w	26w	36w
GENPOD	12w		
HEALTHLINES	16w	32w	52w
IPCRESS	16w	32w	
ITAS		26w	
MIR	12w	24w	52w
PANDA	12w		
REEACT	16w		52w
RESPOND	18w		44w
TREAD	16w	32w	52w

Note: only 3-4 and 6-8 month endpoints were used for analyses presented in this Thesis; 9-12 month endpoint data were only used for imputation of missing data.

Primary outcomes

As the studies using Dep-GP aim to investigate associations with prognosis independent of treatment, the primary outcome for the analyses presented in this thesis is the endpoint symptoms of depression. This was captured in two ways:

1) The standardised mean or z-score of the primary depressive symptom measure score used at 3-4 months post-baseline in each study. The means and standard deviations were calculated separately for each measure at 3-4 months post-baseline.

2) The natural logarithm of 3-4 months post-baseline depression scale scores combined across all studies irrespective of the measure used (the type of measure used across studies were controlled for in all models by including the random allocation in each study in all models of prognosis, as detailed further below). When the regression coefficient is exponentiated it provides an estimate of the proportional difference in symptoms per unit-change in the independent variable relative to the mean; it gives the difference between groups expressed as a proportion. This avoids the need to standardise scores across different measures.

It was expected that the two methods of capturing endpoint symptoms of depression would give broadly similar results but that the natural logarithm or “log outcome”

might have greater clinical utility as the proportions can be expressed as percentage differences. These might be more easily understood by patients and clinicians than differences in symptoms worked out as fractions of one standard deviation as is the case with the standardised mean "z-score outcome". Further, the use of the log outcome allowed for the consideration of whether degrees of difference in prognostic factors at baseline were greater or less than the levels considered to be clinically important; i.e. they were judged against proportional considerations of the minimal clinically important difference (MCID) in depressive symptoms.

Secondary outcomes

The same type of standardised mean (z-score) outcome as described above was also calculated for the endpoints between 6-8 months post-baseline. This was calculated using the same mean and standard deviation as were used to create z-scores at the 3-4 month endpoint, removing the potential for bias that could be introduced by ignoring the within-study variability in standard deviations across different time points (63). An additional secondary outcome was remission on each of the primary outcome measures used in each study (scores below the cut-off for 'caseness' (i.e. the level at which symptoms are considered to be of sufficient severity that is a likely the person would meet criteria for a clinical diagnosis) on each measure detailed in Table 2.2). In order to obtain potentially more easily interpretable findings, remission was modelled both as a binary outcome (with logistic models), and as a count outcome (with Poisson models). As such, when the regression coefficient was exponentiated it would give the difference in the proportion of patients reaching remission between groups (with different levels of the baseline prognostic indicator variable).

Attrition was also considered as a secondary outcome. This was defined as a participant dropping out from their randomised treatment or from the study by: withdrawing from the study or being withdrawn by a clinician or the study team, or the participant being lost to follow-up, between the baseline assessment and the 3-4 month endpoint. As the Dep-GP does not have any data on the 54 participants withdrawn from CADET, estimates of attrition at the 3-4 month end point from that study might have been overestimated as the denominator in such estimates could be lower than would otherwise have been the case, or underestimated as the numerator

could be lower than it would otherwise have been. If estimates from the available data in Dep-GP for CADET were considerably different from those of other studies, it would have been removed from analyses in sensitivity analyses to determine the impact this had on overall estimates (see Sensitivity Analyses section below for more details).

For some of the planned analyses which involved subsets of the Dep-GP studies, all studies used the BDI-II at baseline, and all but one study did so at 3-4 post-baseline. The one study that did not use the BDI-II at 3-4 months post-baseline (COBALT) used the PHQ-9 at 3-4 months instead, and then used the BDI-II again at 6-8 months post-baseline. For such analyses a sensitivity analysis was conducted using a conversion of the scores on those measures to a continuous variable: the PROMIS T-score (146). Individual item level data on the BDI-II and PHQ-9 in the studies of the Dep-GP IPD were used to directly estimate the latent trait depressive symptom severity score (PROMIS T-Score) using the expected *a posteriori* parameter from a multidimensional item-response theory (IRT) based conversion tool (150). The use of this well validated cross-walk removed the need for imputing systematically missing outcome data across studies (63). In addition, further sensitivity analyses were conducted using the BDI-II scores at the 3-4 month endpoint in the five studies not systematically missing such data.

Prognostic indicators under consideration

The PRISMA-IPD statement suggests that protocols for analyses of IPD data are preferred (108), and the PRISMA-P (for protocols) statement recommends that all data items that will be requested from authors of primary studies and those to be used in the analyses should be pre-specified prior to conducting the analyses (151). To that end, below are the baseline factors that were investigated as potential indicators of prognosis.

1. Depressive symptom severity taken as scores on the depressive symptom measures detailed above.
2. Depressive ‘disorder severity’ factors at baseline, from self-reported:
    a. the sum of the scores on the depressive sub-scales of the CIS-R (12 studies)

b. the sum of the scores on the anxiety sub-scales of the CIS-R combined, and individually by subscale (11 studies)
c. the number and type of comorbid anxiety disorders (10 studies)
d. the duration of depression (see table 2.2 for the way this was measured) (11 studies)
e. the duration of all anxiety sub-types measured in the CIS-R (see table 2.2) (10 studies)
f. whether or not participants have a history of depression (12 studies)
g. whether or not participants have a history of previous treatment for depression (11 studies), and whether or not participants have a history of antidepressant treatment (12 studies)
h. whether or not participants were experiencing significant functional impairment at baseline (10 studies)
i. alcohol misuse as measured with the AUDIT-PC questionnaire (6 studies)

3. Social support and specific items of the Social Support Scale (23,137) (6 studies) (see table 3 for details)
4. The occurrence of recent stressful life events and specific types of life events included in the Social Readjustment Rating Scale (138) (6 studies) (see table 2.2 for details)
5. Demographic factors
    a. Age (12 studies)
    b. Gender (12 studies)
    c. Ethnicity (10 studies)
    d. Employment status (11 studies)
    e. Marital status (10 studies)
    f. Highest level of educational attainment (9 studies)
    g. Financial wellbeing (7 studies)
    h. Housing tenancy status (9 studies)
6. Presence or absence of a long-term physical health condition (9 studies).

Of the above factors 1 and 2 were assessed in Chapter 3, to meet sub-aims 1 and 2 of the thesis (see Chapter 1 section on Summary and Aims for this Thesis). Data for these factors were available for between 10 (for durations of anxiety problems and

anxiety disorder comorbidities) and 12 (depressive symptom severity, history of depression and history of antidepressant medications, and depressive subscales of CIS-R) studies in Dep-GP IPD dataset, with the exception of alcohol misuse which was only measured in six studies (see Table 2.2 for details). Item 3 (social support) was analysed in Chapter 4 and the remaining items were all analysed in Chapter 5. Six of the Dep-GP studies were included in the analyses in Chapters 4 and 5, see the method section in each chapter for details.

Controlling for Covariates

Although all studies included in the Dep-GP IPD dataset were RCTs and within each study the randomisation should have removed the potential for known or unknown confounding factors to influence the effect of randomisation to the trialled treatments, the benefits of randomisation do not apply when combining studies in an IPD (59). Therefore standard considerations for the control of confounders are important to any investigation of causal effects. However, in the analyses outlined here there was no attempt to determine causal relationships, therefore the need for consideration of different confounding factors in relation to each prognostic factor under investigation was not necessary (152). This notwithstanding, the independent association of each prognostic factor with each outcome was of primary importance here and so factors which might be covariates in the association of each independent variable with the outcome, should be adjusted for (153). In addition, adjusting for variables that are associated with the outcome reduces the residual variance in linear regression models so estimates are more accurate.

Determinations of which factors to include in the meta-analytic models as covariates were made based on *a priori* considerations of the relationships under investigation and the relationships between the covariate and both the prognostic indicator and outcome. On this basis, age and gender were adjusted for in all models in which they were not the prognostic factor under investigation. In addition to age and gender, only factors that were independently associated with both the prognostic factor and the outcome, were not potentially caused by the prognostic factor, and affected the association between the prognostic factor and outcome were considered as potential covariates, as recommended by several authors (152). Treatment allocation, i.e. the

randomisation in each study was controlled for in all multivariable models, in order to investigate associations with prognosis independent of treatment.

Data Handling and Data Management

Pre-processing

Data from the 12 trials were received and cleaned on an individual study basis before combining all studies into a single aggregated dataset as detailed above. PRISMA-IPD suggests that data pre-processing should be reported in detail, including how variables were/will be re-categorised prior to analysis (108). To this end, a number of baseline variables were re-categorised into higher-order categories due to small numbers, see Table 2.4 for details of the categories of the variables that were sent by the study teams who provided data for Dep-GP, and how these variables were re-categorised. Of note, there was poorer data-coverage across the IPD dataset on information about the number of past depressive episodes than there was on a separate question about whether or not the participant had any previous episodes, see Appendix 2. Further pre-processing for the analyses specified below was also considered. The distributions of all variables were inspected prior to imputation (discussed further below). Continuous variables that were non-normally distributed were either transformed to normality prior to imputation or where log-transformation did not result in approximate normality of the distribution of these variables, predictive mean matching (154) was used for imputation of missing data as part of the multiple imputation with chained equations approach discussed further below.

Table 2.4. Categorisation of variables during data pre-processing.

Variable	Original categories when data were sent	New categories used in analyses of Dep-GP IPD
Ethnicity	White	White
	Mixed	Other
	Black	
	Asian	
	Chinese	
	Other	
Employment Status	Full time employed	Employed
	Part time employed	
	Student	Not seeking employment
	Retired	
	House-person	
	Other	
	Unemployed jobseeker	Unemployed
	Unemployed due to ill-health	
Marital Status	Married/cohabiting	Married/cohabiting
	Single	Single
	Separated	No longer married
	Divorced	
	Widowed	
Highest level of education	Degree or higher	Degree or higher
	Foundation Degree/Diploma	A-level or Diplomas
	A-level	
	GCSE	GCSE
	Other qualifications	None or Other
	No formal qualifications	
Financial Wellbeing	Living Comfortably	OK financially
	Doing alright	
	Just about getting by	Just about getting by
	Hard to make ends meet	Struggling financially
	Very hard to make ends meet	
Long-term Health Condition Status	None	No long-term physical health conditions
	Mental Health Only	
	Diabetes	At least one long-term physical health condition
	Asthma or COPD	
	Arthritis	
	Heart Disease	
	Stroke	
	Cancer	
	Kidney Disease	

Missing Data

In IPD datasets data may be systematically missing (also called missing by-design), because a study did not collect data on the given variable, or non-systematically missing (also called inadvertently missing), where responses were unknown, a participant accidentally did not answer a given question, or a participant did not wish to answer the question. The reason for data being missing influences the way data can be used and the possible approaches to dealing with biases introduced by missing data (61). Although IPD datasets have larger sample sizes than individual studies, and as such, list-wise deletion (or 'complete case analyses') might not greatly reduce the overall sample size for analyses, it will still introduce a degree of bias. If there were high degrees of attrition between baseline and endpoints in the studies included in Dep-GP, then the degree of bias introduced by list-wise deletion might become unacceptably high (154–156).

For the analyses presented in this thesis missing data were imputed using multiple imputation with chained equations (MICE) in Stata 15.0 (144). This approach uses regression models to impute missing values. A number of imputed datasets are produced to reflect the uncertainty/variability in the imputation process. Where predictive mean matching (PMM) was required (as discussed above regarding the distributions of continuous variables) it was conducted via a k-nearest neighbours approach as this is considered to be more appropriate for non-normal continuous variables (157), following convention, here I used the ten nearest neighbours for each missing data point (k=10) (157). Linear regression was used for approximately normally distributed continuous variables, logistic regression models for binary variables, and ordinal and multinomial regression models for ordered and unordered categorical variables respectively. All imputation models were built using data on baseline and outcome variables following conventions (156). Only variables with less than 50% missing data were imputed (see Appendix 2 Table 1 for degrees of missing by variable). All imputation models were run to produce 50 imputed datasets. If the primary analysis (detailed below) were to show that results differed considerably when studies with systematically missing baseline data were included/excluded from the meta-analytic models, then a separate imputation approach would have been taken to impute these systematically missing data: multiple imputation with multilevel random effects for study (158). The analyses

demonstrating that the latter approach was not necessary are presented in Chapter 3 below. Data were imputed by study, maximising the accuracy of the imputation models by including variables that were not of relevance to the analyses here or that were systematically missing between the studies, and removing the potential for between-study variability on the variables included in the imputation models to bias the imputations for any given study (61).

Software & Packages

Stata SE 15 (144): admetan and ipdmetan (93), MICE (159), and mi impute pmm (154) packages. R Studio (160): metafor (161).

Primary Analyses

To investigate sub-aim 1 linear regression models of the score on the depressive symptom scales at 3-4 months post-baseline were built in each study (using both the z-score and log outcomes), adjusting for the random allocation in each study and for any important covariates. The same was done for sub-aim 2 with the addition of adjusting for baseline depression scale scores in each study, and then separately for other 'disorder severity' related factors listed above. Estimates from each study were then pooled in random effects meta-analyses. This can be considered a 'two-stage' meta-analytic approach as data are first aggregated at the level of the study before being combined in a meta-analysis, and is considered the most appropriate method for conducting analyses such as those proposed here (60,64). 'One-stage' approaches which derive across-study summary estimates without calculating effects within each study first, allow for greater complexity in meta-analytic models and as such have been favoured in other IPD meta-analyses (46,162). However, one-stage models have been criticised for increasing the opportunity to introduce biases by failing to separate within-study from between-study effects, a lack of ability to cope with random-effects in time-to-event data, and for a difficulty in using such approaches to display meta-analytic findings appropriately in a forest plot (64,93). As such, the one-stage approach is less favourable than a two-stage approach when complex model structures are not considered necessary (64,93). Further, the difference in the ability of the approaches to deal with complex models is considered negligible by some authors (93). As no multilevel models were required to meet the aims of this thesis a two-stage approach is favoured and was used here.

A multivariable model of outcome was built considering all of the 'disorder severity' factors that were significantly associated with outcome after adjusting for baseline depressive symptom scale scores alone. This was initially only conducted with variables that were not systematically missing between the studies, such models were built firstly on all studies and then on all studies that did not have systematically missing covariates that could otherwise have been included in the multivariable model. These models were compared and had there been considerable differences in the effects, systematically missing variables would then have been imputed, as described above. Decisions on which factors to include/exclude in the multivariable models were be led by consideration of the unique contribution to the models by each variable, the amount of variance explained (adjusted R^2) when modelled with and without the given factor, and to tests of the assumptions of linear regression models. Where there were high degrees of multicollinearity the variable(s) explaining most variance in outcome were retained in the model while the other(s) were removed. Link tests were performed to consider the appropriateness of the linear link function. Multivariate normality, homoscedasticity, and overly influential data points were considered by plotting residuals, and assessing Cook's distance in the residuals plotted against leverage. Forest plots of the variables included in the final model were built in R Studio (160) using the Metafor package (161) in order to visualise all variables in the same figure.

For sub-aim 3 - separate meta-analyses were conducted with each of the prognostic indicators under consideration, unadjusted (with the exception of adjustments for the random allocation in each study, age, and gender) and separately adjusted for i) depressive symptom severity, and ii) the depressive 'disorder severity' factors found to be independently associated with outcomes in the analyses for sub-aim 2. This was the same process as for the analyses for sub-aim 4, albeit the latter analyses were conducted with attrition as the sole outcome.

There were therefore three models of each outcome built for each prognostic factor assessed and an additional model just for the association of baseline depressive symptom severity with each outcome independent of treatment, age and gender (Model 1):

1. Baseline depressive symptom scale score adjusted for treatment, age and gender.
2. As in 1 but with the addition of each prognostic factor under consideration (one by one).
3. As in 2 but with the addition of covariates specific to each prognostic factor.
4. As in 3 with the addition of all 'disorder severity' factors that were significant or otherwise important in 2, and then removing factors that are no longer significant when modelled together in multivariable models.

Meta-analyses were conducted using the "ipdmetan" package in Stata (93), a wrapper in that package: "admetan" was used for primary analyses over imputed datasets, and ipdmetan was used for sensitivity analyses of observed 'un-imputed' data. For the z-score and log outcomes meta-analytic models were fitted using linear regressions and logistic regression models were fitted for remission and attrition as outcomes. All meta-analyses were conducted using DerSimonian and Laird random effects models which take into account heterogeneity of coefficients between studies. The degree of heterogeneity was assessed using prediction intervals and its impact assessed using the I^2 statistic (163). To determine the proportional difference in the depressive symptom scale scores at 3-4 months post-baseline per unit change in the prognostic factor under consideration, the coefficients from the linear models using the logarithm primary outcome were exponentiated (to base e). Poisson models of remission at 3-4 months post-baseline were also fitted and as above, the coefficient for each prognostic indicator was then exponentiated in order to calculate the proportional difference in remission for each unit difference in the prognostic indicator.

Sensitivity Analyses

As discussed in the PRISMA-IPD statement and by several authors, consideration of heterogeneity of effects between studies in an IPD are as important as they are in any meta-analysis (61,108), so sub-group analyses should be considered to better understand effects (57,61). In the Dep-GP IPD dataset if heterogeneity between the studies were considerable based on guidance from the Cochrane Collaboration e.g.

with I^2 above 75% or where the effect in one study was considerably different from that of all other studies after inspecting the forest plot (164), sensitivity analyses were performed removing studies contributing most to the heterogeneity from the meta-analyses to consider their impact on the summary statistics. If the same variables were found to have considerable amounts of heterogeneity when analysed in each of the four models above, sensitivity analyses would be conducted for the model controlling for the most other variables, e.g. symptom severity and covariates (model 3) only. In addition, for variables in the final model(s), sensitivity analyses were similarly planned where the threshold for substantial heterogeneity was met (I^2 above 50%) (164). Further sensitivity analyses were planned to remove any studies with moderate or high risks of bias or to that offered a low quality of evidence for the effects investigated (see Risk of Bias section below). For sub-aims 1 and 2 in the first instance, further sensitivity analyses were conducted using the endpoint at 6-to-8 months in bivariate meta-analyses in order to include the one study that did not have an endpoint in the 3-to-4 month post-baseline time period. This was initially conducted only to assess the prognostic indication of baseline depressive scale scores adjusted for the covariates specified. If it were found that this led to considerable variation in the results then this method would have similarly been used in the analyses of the other potential prognostic factors. Analyses showing this was not necessary are presented in the sensitivity analyses in Chapter 3.

Risk of Bias

It is important to consider biases introduced by the methods of each reviewed study and the quality of the studies (or the level of evidence provided by such studies) included in an IPD dataset. There are particular checklists that are commonly used to assess the quality of each study to provide evidence for the effects being investigated here I used the most commonly used measure: the GRADE framework (35,165). There are risk of bias rating systems specific to IPD datasets but they require the included studies to be predictive modelling studies (166) which was not the case for the present IPD. Therefore, for Dep-GP risk of bias assessments were conducted using the Quality in Prognosis Studies (QUIPS) tool (36). Two reviewers (myself and a collaborator) independently rated study quality (using GRADE) and the risk of bias (using QUIPS) in each study related to: i) study participation; ii) study attrition; iii) prognostic factor measurement; iv) outcome measurement; v) study

confounding; and vi) statistical analysis and reporting. Studies were then given ratings of “high”, “moderate” or “low” in relation to quality, and also in relation to risk of bias. The quality ratings were conducted in relation to each set of prognostic factors investigated.

Summary and Next Steps

Above I have discussed the importance of treating the identification of studies for an IPD like one would a systematic review, and have outlined the systematic review-like methods utilised to form the Dep-GP IPD dataset. I have outlined the importance of setting up IPD datasets well from the outset with collaborators and contributors all involved in formalising and then committing to data sharing, data access, and publication policy documents. The importance of harmonizing data and compiling individual trial datasets in such a way that the integrity can be checked both before and after the overall IPD is compiled has also been discussed. I have reported how these things were done in Dep-GP, the errors that were found as a result, and how these errors were dealt with.

I identified 13 studies that met inclusion criteria, 12 gave IPD data to help form the Dep-GP dataset. The one study unable to provide IPD data had a considerably smaller sample (n=151) (167) than the others included in the Dep-GP and so the data obtained to form this IPD dataset represents 98% of the participants in all studies meeting inclusion criteria. I have explained the rationale and methods for using data from different subsets of the 12 studies to meet the aims of this thesis for identifying factors associated with prognosis independent of treatment.

Now that these analyses have been described and a protocol for these analyses has been published (168), the subsequent three chapters of this thesis will discuss the results of each set of analyses, considering them in turn: starting with sub-aims 1 and 2 in which the associations between symptom severity and ‘disorder severity’ factors as indicators of prognosis independent of treatment are addressed; then looking at those factors beyond ‘disorder severity’ that may or may not be indicative of prognosis independent of treatment and independent of ‘disorder severity’ to address sub-aim 3. This will include investigating social support in Chapter 4, and

then looking at life events, socio-demographics and long-term health condition status in Chapter 5. Each chapter will also present analyses of associations with attrition, to address sub-aim 4. I will finish the thesis by discussing implications of these analyses altogether, considering their potential utility and further work that might lead on from these analyses in Chapter 6.

Chapter 3: The associations between factors related to severity and prognosis.

Overview

In Chapter 2 I discussed the methods for forming an individual patient dataset using the data from 12 randomised controlled trials of common treatments for depression for adults recruited in primary care. I also outlined a protocol for the analysis of that dataset in order to determine whether factors related to a patient's experience of depression are indicative of their prognosis independent of treatment, and independent of the severity of their depressive symptoms. In this chapter I report the findings from that analysis and consider some clinical and research implications of them, these findings are also reported in a recent paper (169).

Abstract

Background

There is evidence that depressive symptom severity is associated with treatment outcomes. However, we lack evidence for this effect independent of treatment, and for the magnitude of impact of this factor on prognosis. There is limited evidence that some other factors associated with the severity of depressive illnesses but separate from depressive symptom severity may also be associated with treatment outcomes. Again, there is a lack of evidence for such effects independent of treatment.

Methods

Individual patient data were gathered from 12 RCTs. Two-stage random-effects meta-analyses were undertaken to ascertain the independent association between each potential prognostic factor and depressive symptoms at 3-4 months post-baseline, remission, depressive symptoms at 6-8 months post-baseline, and attrition at 3-4 months post-baseline. Risk of bias was calculated using QUIPS and quality was assessed using GRADE. PROSPERO registration: CRD42019129512.

Results

Baseline depressive symptom severity was strongly associated with prognosis independent of treatment; there was a 30.74%(95%CI: 24.94 to 36.82) difference in depressive symptoms at 3-4 months per z-score increase at baseline. The duration of anxiety, duration of depression, comorbid panic disorder, and a history of antidepressant treatment were also associated with prognosis independent of depressive symptom severity, albeit the latter with weaker evidence. When combined there was a difference of 36.25%(95%CI: 12.35 to 65.23) independent of treatment and of depressive symptom severity. Adding these variables to a model of prognosis independent of treatment improved the amount of variance explained from 16% using depressive symptom severity alone to 27%. After adjusting for depressive symptoms only the severity of health anxiety symptoms was significantly associated with attrition at 3-4 months post-baseline. Risk of bias was low in all studies, quality was high and heterogeneity was within acceptable limits for most associations. No substantive differences were found in sensitivity analyses.

Conclusions

Depressive symptom severity was the most impactful indicator of prognosis independent of treatment considered here. Accounting for the ‘disorder severity’ factors above when assessing adults with depression pre-treatment could lead to more accurate prognoses for large numbers of patients.

Introduction

The rationale for this analysis was described in Chapter 1 and the background to the dataset used and the methods of analysis are described in Chapter 2. They are also outlined in a protocol paper (168) and a registration document (PROSPERO registration: CRD42019129512 (01/04/2019)). In brief, knowledge of prognosis independent of treatment might be particularly informative for patients and clinicians alike, and may influence the clinical management of depression, but we currently lack robust knowledge of the factors that could be used to predict prognosis for people with depression (31,33,51). In particular, we lack knowledge of the factors associated with prognosis independent of treatment for adults with depression, because as noted in Chapter 1, no previous systematic reviews or meta-analyses appear to have addressed prognosis in this context. The extant literature has demonstrated with a high degree of consistency that the severity of depressive symptoms pre-treatment is associated with the outcomes of a number of different treatments for people with depression (44,66,76–79,81). However, there is the possibility that these associations are limited to people receiving those particular types of treatment only, and there were a number of important methodological limitations to previous studies as outlined in Chapter 1, so there is uncertainty about the magnitude of the association between depressive symptom severity and prognosis. In addition, those studies either rarely included data from primary care settings, or they did not give enough information about how the patients were recruited in the primary studies they reviewed to know if the results may be generalizable to other health service settings. In the UK the vast majority of people with depression initially seek help from primary care settings (23), this is true of a number of other countries, and there is an effort to increase the screening, assessment, and treatment of depression in primary care or by general physicians in many other countries too (8,22). Identifying prognostic factors in a primary care setting would therefore have important utility as findings could be generalizable to the largest proportion of adults seeking treatment for depression (23).

On the basis of past research there is further uncertainty as to whether features of depressive illness, such as duration and comorbidity are related to prognosis (170).

As outlined in Chapter 1, I refer to such factors as indicators of depressive 'disorder severity', in contrast to depressive symptom severity. Some of these 'disorder severity' factors have been reported to be associated with response to particular treatment(s) but with considerably less consistency across studies than for depressive symptom severity, and the associations have been investigated in far fewer studies. In addition, the existing literature has not established whether these factors are associated with prognosis independent of treatment and after controlling for depressive symptom severity. Knowledge of the potential benefit of assessing each 'disorder severity' factor in addition to depressive symptom severity could help inform both clinical practice and future research.

Aims

The analyses presented in this chapter aimed to meet sub-aims 1 and 2 of this thesis, and to partly meet sub-aim 4, re-stated here for reference:

1) To determine whether and the degree to which depressive symptom severity is associated with prognosis for adults with depression independent of treatment, in primary care.
2) To determine which depressive 'disorder severity' factors pre-treatment are associated with prognosis independent of depressive symptom severity and independent of treatment.
4) To determine whether all of the above factors are associated with attrition from treatment, independent of treatment and independent of depressive symptom severity.

Methods

The methods for this systematic review and meta-analysis of individual patient data have been described in Chapter 2 as well as in protocol (168) and registration documents (Open Science Framework: DOI 10.17605/OSF.IO/UX95Q; PROSPERO: CRD42019129512 (01/04/2019)). In brief: this study involved compiling an IPD dataset from RCTs of adults with depression that sought treatment in primary care and were randomised to any type of treatment. This was in order to be able to assess effects independent of the variety of treatments available to many clinicians in primary care settings.

Measures

Details of the measures used in each study can be found in Table 2.2. One measure was used across all studies: the CIS-R (112). This analysis used the scores and durations measured on each of the symptom subscales which are also used to determine ICD-10 (113) diagnoses, and the degree of functional impairment.

Ethical considerations and trial registrations

All included studies were granted ethical approvals by NHS Research Ethics Committee, see Appendix 4. No additional NHS ethical approval was required for this study: HRA reference 712/86/32/81 confirmed 8th August 2019.

Data Extraction

Raw data were extracted for each study participant on all variables outlined in Chapter 2. These data were cleaned one study at a time, independently by two reviewers and cross-checked with publications and via liaison with chief investigators for each study. Issues were resolved by consensus between four reviewers (myself, my two sponsors/supervisors and a collaborator).

Primary outcomes

The primary outcome was endpoint symptoms of depression captured in two ways: 1) the standardised mean ("z-score") of the primary depressive symptom measure score used at 3-4 months post-baseline in each study. For the primary analyses this was based on four different depressive symptom measures (see Tables 2 and 3): the means and standard deviations were calculated separately for each measure at 3-4 months post-baseline. 2) The natural logarithm ("log") of 3-4 months post-baseline depression scale scores combined across all studies irrespective of the measure used (the type of measure used across studies was controlled for in all models by including the random allocation in each study in all models of prognosis, as detailed further below).

Secondary outcomes

1) Remission on the primary depression measure in each study at 3-4 months post-baseline (see Table 2.2 for how this was defined), both the odds ratio for remission and the percentage difference in remission per unit change in the prognostic indicator variable were calculated, see Chapter 2 for details.
2) Endpoint depressive symptoms at 6-8 months post-baseline also captured with the z-score of depressive symptom measures at 6-8 months post-

baseline, using the mean and standard deviation for the scores at 3-4 months post-baseline, and the natural logarithm of scores at that time point as described above.

Data Analysis

Details of the analysis plan are provided below, for further details see (168) and see Chapter 2.

Primary analyses

Two-stage random effects meta-analyses were conducted for each prognostic factor adjusted for the treatment allocation, analysing within each study before aggregating across studies, using Stata 15 (144). This approach removes variance due to the different depressive symptom measures used across the studies. There were three models of each primary outcome built for each prognostic factor assessed and an additional model just for the covariates and the baseline depressive symptom scale scores (Model 1):

Model 1: Baseline depressive symptom scale score adjusted for age, gender and treatment.

Model 2: Each 'disorder severity' factor (one by one) adjusted for age, gender and treatment.

Model 3: As in 2 with the addition of baseline depressive symptom scale score.

Model 4: As in 3 with the addition of covariates specific to each prognostic indicator.

Then final models were built with the primary outcome (using both the z-score and log outcomes), adding each prognostic indicator to the model in order of magnitude of effect from model 4 (one by one). Those no longer significantly associated with prognosis (at the 5% significance level) after adding subsequent factors were removed. If similar items were able to be included, in order to avoid multi-collinearity those contributing least to the model were also removed. The association with prognosis by each prognostic indicator was then tested after removing any highly collinear items, adjusted for age, gender, treatment, baseline depressive symptom scale score, specific covariates, and the other 'disorder severity' factors. Factors that

were used to make up other variables already included in the full models (e.g. individual subscale scores and the total score from the same questionnaire measure) were not tested as potential prognostic indicators in the final models. In the final models, any ordinal 'disorder severity' variables were re-categorised to assess the associations with prognosis in clinically meaningful groups (e.g. duration items were re-categorised into durations at baseline of less than one year and durations greater than one year). The percentage differences in the mean depressive symptom scores attributable to a one category (or unit) difference (at baseline) in each variable included in the final model, after adjusting for all other variables included in the model, were then assessed using the log outcome. This allowed for the calculation of magnitudes of difference in endpoint symptoms for each category (or unit) and thus consideration of clinically important differences. This was done by using the proportional MCID which has previously been calculated to be approximately 17.5% on the BDI-II (171). The final models were evaluated for their explanatory utility by considering the amount of variance in depressive symptom scale scores at 3-4 months post-baseline explained by the models when adding each variable in the final model one at a time (using the adjusted R^2 statistic). This was calculated both for the z-score and the log outcomes.

Means and standard deviations are presented to one decimal place, with percentages and I^2 rounded to the nearest whole number. Effect estimates (including percentages for the log outcome), confidence intervals and model fit statistics are presented to two decimal places, p-values are presented to two decimal places or the first significant figure.

Secondary and sensitivity analyses

As noted in Chapter 2, sensitivity analyses were planned if heterogeneity was considerable based on guidance from the Cochrane Handbook for Systematic Reviews of Interventions (164) (I^2 above 75%) for models 1-3, if heterogeneity was substantial (I^2 above 50%) for variables in the final model(s) (model 4) (164). Further sensitivity analyses were conducted using the endpoint at 6-8 months in bivariate meta-analyses including the study that did not have an endpoint in the 3-4 month post-baseline time-period using the mvmeta package in Stata (172). The impact of variables that could not be imputed as they were not collected in any one of the Dep-

GP studies was assessed by comparing results of meta-analyses with and without studies systematically missing any potential 'disorder severity' factor. The two above sensitivity analyses were initially run to assess the prognostic indication of baseline depressive symptom severity adjusted for the covariate factors specified. If it were found that either of the above led to considerable variation in the results then the bivariate meta-analytic method was planned to be similarly used in the analyses of the other potential prognostic factors, or the systematically missing variables imputed and all analyses run over those data, accordingly. In addition, all analyses were also run on observed ("un-imputed") data to consider the impact of imputation.

Additional sensitivity analyses were conducted using all 11 studies that had an endpoint at 3-4 months post-baseline, to build a full 'disorder severity' model of outcome using only those variables available in all 11 studies. In order to include a measure of anxiety symptom severity, as the total of the anxiety subscales from CIS-R was not available in two studies (AHEAD & HEALTHLINES) (119,123), the z-score of anxiety symptoms on all measures of anxiety used in the studies including the HADS, GAD-7 and CIS-R anxiety subscales was calculated in the same way as depressive symptom severity, using the mean and standard deviation of each symptom measure across all studies at baseline.

Risk of Bias

Risk of bias assessments were conducted using the Quality in Prognosis Studies (QUIPS) (36), and the quality of evidence for each prognostic indicator was assessed using the GRADE framework (165). These were conducted independently by two reviewers (a collaborator and myself).

Results

Characteristics of the included studies

In total, 13 RCTs were identified as meeting inclusion criteria (Figure 2.1), 12 provided IPD. A description of each included study can be found in Table 2.1. Two reviewers independently judged the risk of bias in each study to be low in most domain, although half of the studies were judged as moderate risk of bias due to attrition (Table 3.1). Based on the GRADE framework the quality of evidence in

regards each prognostic indicator was considered to be high, see Table 3.2; interrater reliability: (Cohen's Kappa) k=0.98 for QUIPS and k=1.00 for GRADE.

A key question in this study was whether or not adjusting for depressive symptom severity ameliorates the associations between depressive 'disorder severity' factors and prognosis independent of treatment, therefore descriptive statistics are presented stratified by a median split of depressive symptom severity at baseline, using observed data only, see Table 3.3. Across the dataset, those with higher depressive symptom severity were more likely to have: identified as female; a greater number of comorbid mental health problems; mental health problems of longer durations; lower social support; lower health-related quality of life; more adverse life events; and greater social disadvantages, than those with lower baseline scores. Only ethnicity, whether or not participants had a self-reported long-term health condition, and scores on a measure of alcohol misuse, did not significantly differ between those with higher compared to lower baseline depressive symptoms (Table 3.3). Details on the degrees of missing data for prognostic indicators and outcomes are available in Appendix 2.

Table 3.1. QUIPS risk of bias ratings.

Study	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
AHEAD	Low	Moderate	Low	Low	Low	Low
CADET	Low	Moderate	Low	Low	Low	Low
COBALT	Low	Low	Low	Moderate	Low	Low
GENPOD	Low	Low	Low	Low	Low	Low
HEALTHLINES	Low	Moderate	Low	Low	Low	Low
IPCRESS	Low	High	Low	Low	Low	Low
ITAS	Low	Low	Low	Low	Low	Low
MIR	Low	Moderate	Low	Low	Low	Low
PANDA	Low	Low	Low	Low	Low	Low
REEACT	Low	Moderate	Low	Low	Low	Low
RESPOND	Low	Moderate	Low	Low	Low	Low
TREAD	Low	Low	Low	Low	Low	Low

Table 3.2. GRADE quality rating for each evidence for each type prognostic factor assessed.

	Prognostic Indicator									
Study	Depressive symptom severity	CIS-R score items	Depression Duration	CIS-R duration items	Comorbid Diagnoses	Number of comorbid diagnoses	Functional Impairment	History of Depression	Past ADM use	Past treatment for depression
AHEAD	High	High	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CADET	High	High	High	High	High	High	High	High	High	High
COBALT	High	High	High	High	High	High	High	High	N/A	N/A
GENPOD	High	High	High	High	High	High	High	High	High	High
HEALTHLINES	High	High	High	High	High	High	High	High	High	High
IPCRESS	High	High	High	High	High	High	High	High	High	High
ITAS	High	High	High	High	High	High	High	High	High	High
MIR	High	High	High	High	High	High	High	High	High	High
PANDA	High	High	High	High	High	High	High	High	High	High
REEACT	High	High	High	High	High	High	High	High	High	High
RESPOND	High	High	High	High	High	High	High	High	High	High
READ	High	High	High	High	High	High	High	High	High	High
Overall	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**

Table 3.3. Baseline characteristics of Dep-GP sample stratified by median split of baseline z-score of depressive symptom scale scores using complete data.

Self-reported Baseline Characteristics	Factor	Low Symptom Severity N(%) or Mean(SD)	High Symptom Severity N(%) or Mean(SD)	χ^2 or t-test p-value
Total		**2978(50)**	**3033(50)**	
Age	Mean(sd)	44.0(14.6)	41.4(13.6)	<.0001
Gender	Female	2005(67)	2127(70)	.02
	Male	973(33)	906(30)	
	Other	0	0	
Ethnicity	White	2262(94)	2319(93)	.32
	Non-White	143(6)	165(7)	
Employment status	Employed	1574(50)	1413(49)	<.0001
	Not seeking employment	817(29)	749(26)	
	Unemployed	385(14)	739(26)	
Marital Status	Married/cohabiting	1333(55)	1231(47)	<.0001
	Single	663(27)	852(32)	
	No longer married	437(18)	557(21)	
Educational Attainment	Degree or higher	691(32)	469(23)	<.0001
	A-level or Diplomas	529(25)	518(25)	
	GCSE	634(30)	690(33)	
	None or Other	289(14)	409(20)	
Financial status	Doing OK	957(52)	581(32)	<.0001
	Just about getting by	573(31)	597(33)	
	Struggling	322(17)	616(34)	
Housing status	Home owner	1359(56)	1130(44)	<.0001
	Tenant	789(33)	1126(44)	
	Other	263(11)	326(13)	
Long-term conditions	No	1653(75)	1773(73)	.10
	Yes	539(25)	646(27)	
Social Support	Mean(sd)	21.3(3.3)	19.2(4.1)	<.0001
Number of recent life events	Mean(sd)	1.4(1.3)	1.8(1.5)	<.0001
AUDIT-PC score	Mean(sd)	2.8(3.0)	2.7(3.2)	.71
Hazardous Alcohol misuse	No	1224(79)	1146(78)	
	Yes	327(21)	329(23)	.41
EQ5D Index Score	Mean(sd)	0.7(0.2)	0.7(0.2)	<.0001
History of Depression	No	815(29)	736(23)	.0004
	Yes	1995(71)	2415(77)	
History of ADM treatment	No	1073(38)	1035(33)	.0001
	Yes	1739(62)	2120(67)	
Any past treatment	No	964(39)	940(32)	<.0001
	Yes	1519(61)	1963(68)	
CIS-R Total Score	Mean(sd)	21.9(8.1)	31.3(8)	<.0001
Functional Impairment	No impairment	344(14)	124(5)	<.0001
	Things more difficult but get everything done	1184(49)	902(34)	
	Impaired in one activity	376(15)	394(15)	

	Impaired in more than one activity	533(22)	1223(46)	
		Mean(sd)	Mean(sd)	
CIS-R scores	Compulsions	0.6(1.0)	1.0(1.3)	<.0001
	Concentration	1.7(1.4)	2.4(1.5)	<.0001
	Depression	2.3(1.2)	3.2(1.0)	<.0001
	Depressive thoughts	2.5(1.4)	3.5(1.1)	<.0001
	Fatigue	3.0(1.1)	3.3(0.9)	<.0001
	Generalized Anxiety	1.8(1.4)	2.5(1.5)	<.0001
	Health Anxiety	0.8(1.1)	1.3(1.3)	<.0001
	Irritability	2.0(1.3)	2.6(1.3)	<.0001
	Obsessions	1.0(1.5)	1.5(1.7)	<.0001
	Panic	0.4(1.0)	1.0(1.4)	<.0001
	Phobias	0.9(1.1)	1.5(1.4)	<.0001
	Sleep	1.9(1.2)	2.6(1.2)	<.0001
	Somatic concerns	1.3(1.4)	1.8(1.4)	<.0001
	Worry	2.1(1.4)	2.8(1.2)	<.0001
	Compulsions	1.2(1.8)	1.5(1.9)	<.0001
CIS-R durations	Concentration	2.6(1.6)	3.1(1.4)	<.0001
	Depression	3.3(1.4)	3. 5(1.3)	<.0001
	Fatigue	3.1(1.4)	3.3(1.3)	<.0001
	Generalized Anxiety	2.3(1.9)	2.5(1.9)	.0001
	Health Anxiety	1.8(1.9)	2.3(1.9)	<.0001
	Irritability	2.7(1.6)	3.0(1.5)	<.0001
	Obsessions	1.2(1.8)	1.7(1.9)	<.0001
	Panic	0.6(1.4)	1.4(1.9)	<.0001
	Phobias	1.2(1.9)	2.0(2.1)	<.0001
	Sleep	2.7(1.8)	3.1(1.6)	<.0001
	Somatic concerns	2.3(1.8)	2.7(1.7)	<.0001
	Worry	3.0(1.6)	3.2(1.4)	<.0001
Average Anxiety Duration		1.8(0.9)	2.2(1.0)	<.0001
Number of comorbid CMDs		1.5(1.0)	2.3(1.1)	<.0001

Note: numbers do not add up to total N due to missing data

Association between prognostic indicators and depressive symptom scores at 3-4 months post-baseline

All of the prognostic indicators assessed were associated with prognosis at 3-4 months post-baseline independent of treatment, apart from a comorbid diagnosis of specific phobias (mean difference in depressive symptom scale scores at 3-4 months post-baseline for those with a comorbid specific phobia compared to those without this comorbidity = 0.08(95%CI: -0.01 to 0.17)) and hazardous alcohol misuse (mean difference in depressive symptom scale scores at 3-4 months post-baseline for those with vs without hazardous alcohol misuse =0.06(95%CI: -0.06 to 0.19)), see Table 3.4. Most of the CIS-R anxiety scores and durations, and history of

depression variables were associated with prognosis at 3-4 months post-baseline after also adjusting for depressive symptom severity and covariates. Those with longer durations of anxiety problems and those with a history of depression or history or treatment for depression had poorer prognoses than those with shorter durations, or those without such histories, see Table 3.4. However, there was no evidence that functional impairment (mean difference in depressive symptom scale scores at 3-4 months = 0.04(95%CI: -0.05 to 0.14)) or most comorbid diagnoses were associated with prognosis after adjusting for depressive symptom severity and covariates (e.g. Generalised Anxiety Disorder: mean difference in depressive symptom scale scores at 3-4 months = 0.04(95%CI: -0.02 to 0.10)), with the exception of Chronic Fatigue Syndrome, and Panic Disorder.

Overall, depressive symptom severity was strongly associated with prognosis at 3-4 moths post-baseline. For every one standard deviation increase in depressive symptoms at baseline, after adjusting for treatment and covariates, the mean difference in depressive symptom scale scores at 3-4 months post-baseline was 0.44(95%CI: 0.41 to 0.47). The standard deviation of BDI-II scores at baseline was 10.53 in the five studies using the BDI-II at 3-4 months post-baseline, the difference in BDI-II scores at 3-4 months per standard deviation increase at baseline was approximately 7.07 points. Likewise, the standard deviation of PHQ-9 scores at baseline in the six studies that used that measure was 5.49 and the difference in PHQ-9 scores at 3-4 months post-baseline per standard deviation increase at baseline was 4.78 points. Using the log outcome the scores at 3-4 months post-baseline were 30.74%(95%CI: 24.94 to 36.82) higher on average, see Table 3.4.

Other variables were also strong associated with prognosis after adjusting for depressive symptom severity and covariates. Each category increase in the duration of depression was associated with higher depressive symptom scores at 3-4 months post-baseline (mean difference in depressive symptom scale scores at 3-4 months = 0.08(95%CI: 0.05 to 0.11)). Likewise with each category increase in the average duration of anxiety problems (0.11(95%CI: 0.07 to 0.16)). Having any past treatment for depression was associated with marginally worse prognoses (0.11(95%CI: 0.05 to 0.18)). Having comorbid panic disorder was strongly associated with an increase

in depressive symptoms at 3-4 months post-baseline relative to those without comorbid panic disorder (0.21(95%CI: 0.07 to 0.34)).

The findings when using the standardised mean and natural logarithm were consistent across prognostic factors and across the models used to investigate associations with prognosis. There was however a notable exception where the two ways of considering the primary outcome gave different results. In the models adjusted for treatment, age, gender, depressive symptom severity and covariates, using the z-score there was some evidence that each of the three variables capturing history of depression (history irrespective of past treatment, history of antidepressant treatment, or history of any treatment for depression) were significantly associated with outcome. However, there was no evidence for such associations when using the log outcome, see Table 3.4.

Table 3.4. Outcomes at 3-4 months (difference in z-score of depressive symptoms, and percentage difference in depressive symptoms) per unit increase in baseline prognostic indicators.

	Difference in z-score of depressive symptoms or % difference in depressive symptoms at 3-4 months post-baseline per unit increase in baseline prognostic indicator											
	Adjusted for Treatment, Age and Gender[ρ]				Depressive symptom severity adjusted*				Depressive symptom severity and covariate adjusted* or [‡]			
Prognostic Indicator	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**
Depressive symptom severity	0.44(0.41 to 0.47)	16	30.74(24.94 to 36.82)	78	0.44(0.41 to 0.47)	16	30.74(24.94 to 36.82)	78	0.44(0.41 to 0.47)	16	30.74(24.94 to 36.82)	78
CIS-R Total Score	0.04(0.03 to 0.04)	23	13.10(10.87 to 15.37)	62	0.02(0.01 to 0.02)	48	6.16(3.88 to 8.49)	50	0.02(0.01 to 0.02)	48	6.16(3.88 to 8.49)	50
Depressive Subscales Total[‡1]	0.09(0.08 to 0.11)	71	27.68(21.03 to 34.7)	69	0.04(0.02 to 0.05)	40	9.95(5.04 to 15.09)	42	0.03(0.02 to 0.05)	46	8.22(2.98 to 13.73)	45
Anxiety Subscales Total[‡2]	0.04(0.03 to 0.05)	76	13.47(9.98 to 17.07)	71	0.01(0.01 to 0.02)	69	4.84(1.83 to 7.95)	57	0.01(0.00 to 0.02)	68	4.50(1.37 to 7.72)	55
Compulsions Score	0.13(0.10 to 0.17)	44	8.93(5.78 to 12.18)	61	0.05(0.01 to 0.09)	59	3.62(0.53 to 6.80)	63	0.05(0.01 to 0.09)	59	3.62(0.53 to 6.80)	63
Compulsions Duration[‡1]	0.06(0.04 to 0.09)	44	4.30(2.45 to 6.19)	49	0.03(0.00 to 0.05)	43	1.77(0.08 to 3.50)	45	0.02(0.00 to 0.04)	31	1.69(0.16 to 3.25)	35
Concentration Score	0.16(0.11 to 0.20)	72	9.81(6.47 to 13.26)	69	0.05(0.02 to 0.07)	0	2.73(0.98 to 4.51)	0	0.05(0.02 to 0.07)	0	2.73(0.98 to 4.51)	0
Concentration Duration	0.12(0.09 to 0.16)	62	8.16(5.49 to 10.90)	61	0.06(0.03 to 0.09)	49	3.96(1.95 to 6.02)	37	0.06(0.03 to 0.09)	49	3.96(1.95 to 6.02)	37
Depression Score[‡3]	0.2(0.15 to 0.25)	67	13.97(10.59 to 17.46)	55	0.04(0.01 to 0.08)	42	3.88(1.42 to 6.39)	19	0.03(-0.01 to 0.07)	42	3.13(0.23 to 6.11)	22
Depressive Thoughts Score[‡1]	0.22(0.18 to 0.25)	52	15.5(12.21 to 18.88)	52	0.08(0.06 to 0.11)	6	6.29(3.85 to 8.79)	19	0.07(0.04 to 0.10)	8	4.90(2.50 to 7.36)	10
Depression Duration[‡1]	0.15(0.12 to 0.18)	27	10.23(7.36 to 13.18)	51	0.09(0.06 to 0.12)	49	6.61(3.96 to 9.33)	49	0.08(0.05 to 0.11)	46	5.87(3.24 to 8.56)	49
Fatigue Score[‡1]	0.09(0.03 to 0.14)	70	6.59(1.72 to 11.70)	74	0.03(0.00 to 0.07)	30	3.20(0.22 to 6.27)	37	0.03(0.00 to 0.07)	21	3.16(0.12 to 6.28)	31
Fatigue Duration	0.12(0.09 to 0.15)	35	8.11(5.31 to 10.98)	52	0.08(0.05 to 0.10)	28	5.73(3.63 to 7.86)	25	0.08(0.05 to 0.10)	28	5.73(3.63 to 7.86)	25
Generalised Anxiety Score[‡1]	0.10(0.07 to 0.13)	33	6.51(4.36 to 8.71)	42	0.02(0.00 to 0.04)	1	1.33(-0.27 to 2.96)	7	0.02(0.00 to 0.04)	13	1.47(-0.23 to 3.19)	10
Generalised Anxiety Duration	0.06(0.04 to 0.08)	0	4.43(2.63 to 6.26)	32	0.03(0.01 to 0.05)	0	2.49(1.01 to 3.98)	14	0.03(0.01 to 0.05)	0	2.49(1.01 to 3.98)	14
Health Anxiety Score[‡1]	0.15(0.12 to 0.18)	41	10.04(8.04 to 12.08)	0	0.07(0.04 to 0.10)	30	4.32(2.18 to 6.51)	18	0.05(0.02 to 0.08)	26	3.26(0.92 to 5.65)	28
Health Anxiety Duration[‡3]	0.08(0.06 to 0.09)	1	5.49(4.20 to 6.79)	0	0.04(0.02 to 0.05)	23	2.8(1.58 to 4.04)	1	0.03(0.01 to 0.04)	0	2.20(0.99 to 3.43)	0
Irritability Score[‡3]	0.09(0.06 to 0.11)	34	6.51(4.57 to 8.49)	0	0.00(-0.02 to 0.02)	0	0.69(-1.27 to 2.68)	10	0.01(-0.02 to 0.03)	9	1.18(-0.78 to 3.18)	0
Irritability Duration[‡2]	0.07(0.04 to 0.10)	49	5.64(3.28 to 8.07)	45	0.03(0.00 to 0.06)	53	3.05(0.75 to 5.41)	48	0.04(0.01 to 0.07)	56	3.28(0.91 to 5.71)	50
Obsessions Score	0.06(0.04 to 0.09)	41	3.45(1.33 to 5.60)	52	0.01(-0.02 to 0.04)	50	0.20(-1.69 to 2.14)	47	0.01(-0.02 to 0.04)	50	0.20(-1.69 to 2.14)	47
Obsessions Duration	0.05(0.03 to 0.07)	0	3.02(1.39 to 4.68)	35	0.01(0.00 to 0.03)	17	0.69(-0.82 to 2.22)	34	0.01(0.00 to 0.03)	17	0.69(-0.82 to 2.22)	34
Panic Score[‡2]	0.13(0.07 to 0.19)	82	8.59(5.04 to 12.25)	72	0.05(0.00 to 0.09)	72	3.05(0.30 to 5.87)	55	0.05(0.00 to 0.1)	74	3.10(0.11 to 6.17)	59
Panic Duration[‡1]	0.10(0.08 to 0.13)	45	6.81(4.74 to 8.92)	53	0.05(0.02 to 0.07)	36	3.10(1.40 to 4.83)	34	0.04(0.02 to 0.06)	33	2.63(0.95 to 4.34)	35
Phobias Score[‡1]	0.15(0.10 to 0.20)	75	10.00(6.70 to 13.40)	62	0.06(0.02 to 0.11)	64	4.07(1.38 to 6.83)	46	0.04(0.01 to 0.08)	53	2.92(0.50 to 5.39)	32
Phobias Duration[‡3]	0.08(0.05 to 0.11)	70	5.59(3.53 to 7.68)	63	0.03(0.01 to 0.05)	56	2.28(0.70 to 3.88)	40	0.02(0.00 to 0.05)	52	1.89(0.39 to 3.41)	35

Sleep Score[‡3]	0.13(0.11 to 0.16)	25	8.25(6.46 to 10.07)	45	0.04(0.01 to 0.06)	1	2.61(0.93 to 4.32)	0	0.02(0.00 to 0.05)	10	1.79(-0.17 to 3.78)	2
Sleep Duration[‡1]	0.10(0.08 to 0.13)	35	7.15(5.10 to 9.23)	44	0.06(0.03 to 0.09)	54	4.55(2.49 to 6.65)	51	0.05(0.03 to 0.08)	45	4.04(2.13 to 5.99)	45
Somatic Score[‡3]	0.10(0.07 to 0.13)	50	6.61(4.13 to 9.15)	47	0.05(0.02 to 0.07)	5	3.16(1.51 to 4.83)	0	0.04(0.02 to 0.06)	0	2.70(1.02 to 4.41)	0
Somatic Duration[‡3]	0.09(0.07 to 0.11)	0	6.62(5.17 to 8.09)	0	0.05(0.04 to 0.07)	0	4.25(2.88 to 5.65)	0	0.05(0.03 to 0.06)	0	3.82(2.47 to 5.19)	0
Worry Score[‡1]	0.11(0.08 to 0.14)	36	7.00(4.86 to 9.19)	24	0.02(-0.01 to 0.04)	18	0.93(-0.84 to 2.74)	0	0.02(-0.01 to 0.04)	6	0.82(-1.03 to 2.71)	0
Worry Duration	0.09(0.07 to 0.11)	0	6.57(4.56 to 8.63)	21	0.05(0.03 to 0.07)	0	3.71(2.04 to 5.4)	0	0.05(0.03 to 0.07)	0	3.71(2.04 to 5.4)	0
Average Duration of Anxiety[‡3]	0.26(0.21 to 0.31)	58	19.62(14.64 to 24.83)	67	0.13(0.08 to 0.18)	53	10.29(6.23 to 14.5)	51	0.11(0.07 to 0.16)	45	9.23(5.35 to 13.24)	47
Number of Comorbid CMDs[‡3]	0.21(0.12 to 0.29)	87	14.47(8.49 to 20.78)	82	0.06(0.00 to 0.12)	70	4.59(0.62 to 8.71)	56	0.05(-0.01 to 0.11)	72	4.06(-0.01 to 8.31)	60
Agoraphobia[‡1]	0.34(0.19 to 0.49)	51	25.72(16.51 to 35.66)	11	0.14(0.02 to 0.26)	30	10.4(3.09 to 18.24)	0	0.09(-0.03 to 0.20)	25	5.96(-1.05 to 13.47)	0
CFS[‡2]	0.31(0.20 to 0.43)	64	26.10(15.62 to 37.54)	60	0.08(0.00 to 0.15)	20	9.80(4.02 to 15.91)	0	0.09(0.01 to 0.17)	29	10.79(4.77 to 17.15)	5
GAD	0.24(0.18 to 0.31)	4	17.54(11.99 to 23.36)	0	0.04(-0.02 to 0.10)	0	4.03(-0.84 to 9.13)	4	0.04(-0.02 to 0.10)	0	4.03(-0.84 to 9.13)	4
MADD	-0.24(-0.30 to -0.18)	54	-12.53(-16.84 to -8.00)	59	-0.05(-0.12 to 0.01)	0	-2.25(-7.08 to 2.83)	28	-0.05(-0.12 to 0.01)	0	-2.25(-7.08 to 2.83)	28
OCD	0.34(0.22 to 0.46)	30	21.51(13.10 to 30.56)	9	0.01(-0.10 to 0.12)	19	-1.29(-9.18 to 7.29)	29	0.01(-0.10 to 0.12)	19	-1.29(-9.18 to 7.29)	29
Panic Disorder	0.41(0.19 to 0.64)	72	33.07(19.08 to 48.71)	48	0.21(0.07 to 0.34)	34	15.01(6.92 to 23.72)	0	0.21(0.07 to 0.34)	34	15.01(6.92 to 23.72)	0
Social Phobia	0.24(0.08 to 0.39)	55	18.25(6.91 to 30.79)	47	0.10(-0.02 to 0.22)	36	8.06(-0.16 to 16.95)	20	0.10(-0.02 to 0.22)	36	8.06(-0.16 to 16.95)	20
Specific Phobias[‡1]	0.08(-0.01 to 0.17)	0	6.28(-0.02 to 12.98)	0	-0.04(-0.12 to 0.04)	0	-1.15(-6.82 to 4.87)	0	-0.02(-0.10 to 0.06)	0	0.21(-5.50 to 6.27)	0
History of Depression[‡3]	0.19(0.12 to 0.26)	42	7.98(2.71 to 13.52)	59	0.11(0.05 to 0.17)	30	3.26(-1.54 to 8.30)	55	0.09(0.02 to 0.16)	13	2.64(-2.62 to 8.18)	53
History of ADM Treatment	0.17(0.10 to 0.25)	20	8.46(1.89 to 15.45)	43	0.10(0.03 to 0.17)	24	3.73(-2.12 to 9.92)	40	0.10(0.03 to 0.17)	24	4.53(-0.65 to 9.98)	40
Any past Treatment	0.19(0.13 to 0.26)	0	10.24(3.9 to 16.96)	24	0.11(0.05 to 0.18)	13	5.21(-0.65 to 11.43)	26	0.11(0.05 to 0.18)	13	5.21(-0.65 to 11.43)	26
Functional Impairment[‡3]	0.27(0.15 to 0.39)	72	20.7(11.32 to 30.87)	63	0.04(-0.05 to 0.14)	58	4.58(-2.35 to 12.00)	48	0.02(-0.07 to 0.11)	54	2.8(-3.88 to 9.94)	46
Hazardous Alcohol misuse[‡2]	0.06(-0.06 to 0.19)	31	7.21(-4.1 to 19.86)	45	0.01(-0.09 to 0.10)	0	3.25(-5.59 to 12.92)	28	0.00(-0.09 to 0.09)	0	2.68(-5.86 to 11.98)	24

[ρ] adjusted for treatment allocation, age, and gender; *additionally adjusted for baseline depression scale z-score; additionally adjusted for: [‡1]employment status; [‡2] marital status; [‡3] employment status and marital status; [Z] per 1 z-score increase.

Association between prognostic indicators and remission at 3-4 months post-baseline

There were very few differences in the factors associated with remission and those associated with the primary outcomes, however heterogeneity was lower in the associations with remission than in regards to the primary outcomes.

The odds of remission were lower with each standard deviation increase in depressive symptom severity scores at baseline (OR=0.49(95%CI: 0.44 to 0.54)), and 31.01%(95%CI: 26.61 to 35.15) fewer patients reached remission. Those scoring higher on the depression subscales of the CIS-R were less likely to reach remission even after adjusting for baseline depressive symptom severity measured on the primary symptom measures (as outlined in Table 2.1) (OR=0.95(95%CI:0.92 to 0.98)), likewise with those scoring higher on the anxiety subscales (OR=0.98(95%CI: 0.96 to 0.99)).

After adjusting for treatment, age, gender, depressive symptom severity and covariates, longer durations of either depression or of depressive symptoms of various sorts (e.g. sleep problems or fatigue) were associated with lower odds of remission: e.g. for depression duration OR=0.85(95%CI: 0.79 to 0.92) and for fatigue duration OR=0.86(95%CI: 0.81 to 0.91)). Again, this was similar for patients with longer durations of anxiety related problems: for each unit increase in the average across all anxiety durations measured in CIS-R, the odds of remission were lower at 3-4 months post-baseline (OR=0.79(95%CI: 0.71 to 0.88)), and approximately 11% fewer patients remitted: 11.42%(95%CI: 7.68 to 15.01). There was limited evidence that the number of anxiety disorder comorbidities was associated with lower odds and proportions of patients in remission at 3-4 months, after adjusting for treatment, age, gender, depressive symptom severity, and covariates: OR=0.89(95%CI: 0.80 to 1.00), percentage difference in remission per additional comorbid disorder =5.68%(0.63 to 10.49).
Of the individual anxiety disorder comorbidities, for patients with comorbid panic disorder the odds of remission were approximately 36% lower (OR=0.64(95%CI: 0.49 to 0.83)) and approximately 23% fewer patients reached remission relative to those without comorbid panic disorder (23.26%(95%CI: 9.02 to 35.27)).

Having a history of antidepressant medication treatment was associated with lower odds of remission independent of treatment, age, gender, depressive symptom severity and covariates (OR=0.83(95%CI: 0.70 to 0.98)). However, as with the

primary outcomes, there was no evidence for an association between functional impairment and remission after having adjusted for baseline depressive symptom severity (OR=0.90(95%CI: 0.76 to 1.07)), or for an association between hazardous alcohol misuse and prognosis independent of treatment, age and gender (OR=0.78(95%CI: 0.59 to 1.03)).

Table 3.5. Difference in odds of remission and percentage difference in odds of remission at three-to-four months post-baseline per unit increase in baseline prognostic indicators.

	Difference in odds of remission and % difference in odds of remission at 3-4 months post-baseline per unit increase in baseline prognostic indicator											
	Adjusted for Treatment, Age and Gender[ρ]				Depressive symptom severity adjusted*				Depressive symptom severity and covariate adjusted* or [‡]			
Prognostic Indicator	**OR(95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**OR(95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**OR(95%CI)**	**I^2**	**%(95%CI)**	**I^2**
Depressive symptom severity	0.49(0.44 to 0.54)	39	-31.01(-35.15 to -26.61)	63	0.49(0.44 to 0.54)	39	-31.01(-35.15 to -26.61)	63	0.49(0.44 to 0.54)	39	-31.01(-35.15 to -26.61)	63
CIS-R Total Score	0.94(0.93 to 0.95)	31	-13.83(-15.31 to -12.33)	0	0.97(0.96 to 0.98)	18	-7.18(-9.21 to -5.10)	0	0.97(0.96 to 0.98)	18	-7.18(-9.21 to -5.10)	0
Depressive Subscales Total[‡1]	0.86(0.83 to 0.89)	50	-26.93(-31.32 to -22.27)	51	0.94(0.91 to 0.96)	24	-10.34(-14.57 to -5.91)	0	0.95(0.92 to 0.98)	29	-9.29(-13.98 to -4.35)	0
Anxiety Subscales Total[‡2]	0.94(0.92 to 0.96)	67	-15.39(-18.31 to -12.35)	38	0.98(0.96 to 0.99)	42	-6.43(-9.96 to -2.77)	35	0.98(0.96 to 0.99)	38	-6.01(-9.86 to -1.99)	36
Compulsions Score	0.81(0.75 to 0.88)	49	-10.69(-14.56 to -6.63)	41	0.91(0.84 to 0.99)	48	-4.82(-8.99 to -0.45)	43	0.91(0.84 to 0.99)	48	-4.82(-8.99 to -0.45)	43
Compulsions Duration[‡1]	0.91(0.87 to 0.95)	25	-4.96(-7.47 to -2.39)	33	0.96(0.92 to 1.00)	16	-1.90(-3.99 to 0.25)	15	0.96(0.92 to 1.00)	3	-1.73(-3.51 to 0.08)	0
Concentration Score	0.79(0.72 to 0.86)	63	-11.71(-14.73 to -8.57)	38	0.94(0.89 to 1.00)	0	-3.21(-5.89 to -0.44)	0	0.94(0.89 to 1.00)	0	-3.21(-5.89 to -0.44)	0
Concentration Duration	0.82(0.78 to 0.87)	33	-9.77(-12.63 to -6.82)	40	0.90(0.86 to 0.95)	8	-4.55(-7.40 to -1.61)	34	0.90(0.86 to 0.95)	8	-4.55(-7.40 to -1.61)	34
Depression Score[‡3]	0.73(0.66 to 0.79)	52	-14.24(-17.06 to -11.32)	28	0.93(0.85 to 1.01)	38	-3.31(-6.95 to 0.47)	29	0.95(0.86 to 1.06)	44	-2.24(-6.42 to 2.13)	30
Depressive Thoughts Score[‡1]	0.71(0.67 to 0.76)	15	-14.41(-17.18 to -11.53)	41	0.87(0.82 to 0.93)	0	-4.72(-7.08 to -2.30)	0	0.89(0.83 to 0.95)	0	-4.09(-6.67 to -1.44)	0
Depression Duration[‡1]	0.78(0.73 to 0.84)	40	-12.01(-15.59 to -8.28)	50	0.83(0.77 to 0.90)	41	-7.84(-11.79 to -3.72)	57	0.85(0.79 to 0.92)	39	-6.91(-10.53 to -3.15)	48
Fatigue Score[‡1]	0.85(0.76 to 0.96)	64	-8.22(-12.03 to -4.24)	40	0.93(0.84 to 1.02)	36	-2.82(-6.08 to 0.55)	19	0.94(0.86 to 1.03)	22	-2.47(-5.50 to 0.66)	3
Fatigue Duration	0.82(0.77 to 0.86)	0	-9.31(-11.6 to -6.96)	4	0.86(0.81 to 0.91)	0	-6.58(-9.64 to -3.42)	29	0.86(0.81 to 0.91)	0	-6.58(-9.64 to -3.42)	29
Generalised Anxiety Score[‡1]	0.86(0.81 to 0.90)	29	-7.97(-9.95 to -5.96)	0	0.96(0.92 to 1.01)	0	-2.10(-4.29 to 0.13)	0	0.96(0.91 to 1.01)	0	-2.51(-4.84 to -0.13)	0
Generalised Anxiety Duration	0.89(0.86 to 0.93)	0	-5.56(-7.32 to -3.76)	0	0.93(0.89 to 0.97)	0	-3.16(-4.93 to -1.37)	0	0.93(0.89 to 0.97)	0	-3.16(-4.93 to -1.37)	0
Health Anxiety Score[‡1]	0.80(0.75 to 0.85)	23	-11.29(-14.51 to -7.96)	27	0.90(0.85 to 0.96)	2	-4.87(-7.97 to -1.67)	12	0.93(0.87 to 0.99)	3	-3.79(-7.02 to -0.44)	6
Health Anxiety Duration[‡3]	0.89(0.85 to 0.92)	16	-6.14(-8.30 to -3.94)	23	0.94(0.90 to 0.97)	0	-2.92(-4.99 to -0.79)	18	0.95(0.91 to 0.99)	0	-2.02(-3.81 to -0.19)	0
Irritability Score[‡3]	0.87(0.83 to 0.92)	0	-6.48(-8.72 to -4.19)	0	1.00(0.94 to 1.05)	0	-0.05(-2.49 to 2.44)	0	0.98(0.93 to 1.04)	0	-0.86(-3.50 to 1.86)	0
Irritability Duration[‡2]	0.88(0.83 to 0.92)	29	-6.61(-9.44 to -3.70)	40	0.92(0.87 to 0.97)	20	-3.72(-6.26 to -1.10)	30	0.92(0.87 to 0.97)	20	-4.00(-6.50 to -1.44)	27
Obsessions Score	0.91(0.86 to 0.97)	54	-4.68(-7.44 to -1.84)	41	0.98(0.92 to 1.04)	46	-0.94(-3.88 to 2.10)	43	0.98(0.92 to 1.04)	46	-0.94(-3.88 to 2.10)	43
Obsessions Duration	0.93(0.88 to 0.97)	38	-4.17(-6.49 to -1.80)	27	0.98(0.93 to 1.02)	36	-1.30(-3.69 to 1.15)	30	0.98(0.93 to 1.02)	36	-1.30(-3.69 to 1.15)	30
Panic Score[‡2]	0.82(0.74 to 0.9)	65	-10.67(-15.36 to -5.73)	60	0.92(0.85 to 1.00)	44	-4.51(-8.88 to 0.07)	46	0.93(0.85 to 1.02)	47	-3.99(-8.95 to 1.24)	50

Panic Duration‡1	0.85(0.81 to 0.90)	26	-9.04(-11.48 to -6.54)	6	0.92(0.88 to 0.96)	0	-4.55(-6.98 to -2.05)	0	0.93(0.89 to 0.97)	0	-4.05(-6.49 to -1.55)	0
Phobias Score‡1	0.78(0.71 to 0.86)	63	-12.69(-16.20 to -9.03)	42	0.89(0.82 to 0.95)	33	-6.56(-9.66 to -3.35)	14	0.91(0.86 to 0.97)	2	-5.14(-8.15 to -2.03)	0
Phobias Duration‡3	0.88(0.83 to 0.93)	63	-7.28(-9.68 to -4.81)	37	0.94(0.90 to 0.98)	29	-3.54(-5.53 to -1.50)	9	0.95(0.91 to 0.99)	25	-3.06(-5.00 to -1.08)	6
Sleep Score‡3	0.81(0.77 to 0.85)	48	-10.57(-12.70 to -8.40)	12	0.92(0.87 to 0.97)	38	-3.76(-6.12 to -1.35)	36	0.94(0.88 to 1.00)	49	-2.97(-5.65 to -0.22)	46
Sleep Duration‡1	0.84(0.81 to 0.88)	5	-8.14(-10.01 to -6.24)	6	0.89(0.85 to 0.93)	4	-5.32(-7.57 to -3.02)	26	0.89(0.86 to 0.93)	0	-4.50(-6.36 to -2.61)	5
Somatic Score‡3	0.88(0.82 to 0.94)	53	-7.06(-9.68 to -4.36)	19	0.95(0.89 to 1.01)	28	-2.8(-5.32 to -0.20)	11	0.96(0.91 to 1.02)	13	-2.23(-4.73 to 0.33)	0
Somatic Duration‡3	0.88(0.85 to 0.91)	0	-6.29(-8.09 to -4.45)	0	0.92(0.89 to 0.96)	0	-3.37(-5.13 to -1.57)	0	0.93(0.89 to 0.97)	0	-3.02(-4.8 to -1.21)	0
Worry Score‡1	0.83(0.79 to 0.87)	0	-8.85(-10.94 to -6.70)	0	0.96(0.91 to 1.01)	0	-1.93(-4.29 to 0.48)	0	0.96(0.91 to 1.02)	0	-1.71(-4.24 to 0.9)	0
Worry Duration	0.85(0.81 to 0.89)	0	-7.75(-9.71 to -5.75)	0	0.89(0.85 to 0.94)	0	-4.29(-6.27 to -2.26)	0	0.89(0.85 to 0.94)	0	-4.29(-6.27 to -2.26)	0
Average Duration of Anxiety‡3	0.65(0.58 to 0.74)	58	-21.29(-24.98 to -17.41)	24	0.78(0.70 to 0.86)	36	-12.36(-16.37 to -8.16)	15	0.79(0.71 to 0.88)	30	-11.42(-15.01 to -7.68)	0
Number of Comorbid CMDs‡3	0.72(0.61 to 0.83)	79	-16.5(-21.65 to -11.01)	65	0.89(0.80 to 0.99)	47	-6.19(-10.96 to -1.17)	38	0.89(0.8 to 1.00)	49	-5.68(-10.49 to -0.63)	38
Agoraphobia‡1	0.56(0.45 to 0.70)	0	-29.44(-39.03 to -18.33)	0	0.76(0.60 to 0.96)	0	-15.57(-26.87 to -2.52)	0	0.82(0.65 to 1.04)	0	-11.65(-23.55 to 2.09)	0
CFS‡2	0.61(0.49 to 0.76)	52	-23.57(-29.68 to -16.91)	18	0.86(0.73 to 1.00)	0	-7.05(-13.74 to 0.16)	0	0.83(0.71 to 0.98)	2	-8.43(-15.03 to -1.33)	0
GAD	0.67(0.58 to 0.77)	26	-19.53(-24.94 to -13.74)	0	0.90(0.78 to 1.04)	0	-5.61(-11.91 to 1.14)	0	0.9(0.78 to 1.04)	0	-5.61(-11.91 to 1.14)	0
MADD	1.45(1.26 to 1.67)	11	17.93(10.45 to 25.93)	39	1.12(0.96 to 1.30)	0	7.29(0.57 to 14.46)	0	1.12(0.96 to 1.30)	0	7.29(0.57 to 14.46)	0
OCD	0.57(0.46 to 0.70)	0	-27.25(-36.31 to -16.89)	0	0.95(0.75 to 1.19)	7	-3.67(-15.73 to 10.13)	0	0.95(0.75 to 1.19)	7	-3.67(-15.73 to 10.13)	0
Panic Disorder	0.47(0.33 to 0.67)	44	-35.19(-48.87 to -17.85)	42	0.64(0.49 to 0.83)	0	-23.26(-35.27 to -9.02)	0	0.64(0.49 to 0.83)	0	-23.26(-35.27 to -9.02)	0
Social Phobia	0.66(0.51 to 0.86)	25	-21.48(-31.73 to -9.68)	0	0.81(0.65 to 1.03)	0	-10.6(-22.06 to 2.54)	0	0.81(0.65 to 1.03)	0	-10.6(-22.06 to 2.54)	0
Specific Phobias‡1	0.86(0.72 to 1.03)	0	-7.43(-16.07 to 2.10)	0	1.04(0.86 to 1.26)	0	2.26(-7.05 to 12.50)	0	1.03(0.85 to 1.25)	0	2.20(-7.06 to 12.39)	0
History of Depression‡3	0.73(0.63 to 0.85)	43	-15.36(-21.16 to -9.13)	33	0.80(0.68 to 0.94)	35	-10.79(-16.69 to -4.46)	24	0.86(0.72 to 1.02)	31	-7.87(-14.44 to -0.81)	25
History of ADM Treatment	0.76(0.65 to 0.89)	23	-14.02(-19.64 to -8.01)	0	0.83(0.7 to 0.98)	26	-9.79(-15.76 to -3.40)	6	0.83(0.70 to 0.98)	26	-9.79(-15.76 to -3.40)	6
Any past Treatment	0.73(0.62 to 0.87)	26	-14.69(-20.37 to -8.60)	0	0.8(0.67 to 0.96)	30	-10.67(-16.85 to -4.03)	9	0.8(0.67 to 0.96)	30	-10.67(-16.85 to -4.03)	9
Functional Impairment‡3	0.65(0.54 to 0.78)	42	-22.44(-27.86 to -16.62)	0	0.9(0.76 to 1.07)	21	-5.98(-12.99 to 1.59)	7	0.94(0.79 to 1.12)	22	-3.82(-11.17 to 4.14)	11
Hazardous Alcohol misuse‡2	0.78(0.59 to 1.03)	23	-11.25(-24.52 to 4.36)	23	0.86(0.67 to 1.10)	0	-4.79(-16.12 to 8.07)	8	0.86(0.67 to 1.11)	0	-3.41(-13.26 to 7.55)	0

ρ adjusted for treatment allocation, age, and gender; *additionally adjusted for baseline depression scale z-score; additionally adjusted for: ‡1employment status; ‡2 marital status; ‡3 employment status and marital status; Z per 1 z-score increase.

Association between prognostic indicators and depressive symptom scores at 6-8 months post-baseline

Seven studies had an end-point at 6-8 months post-baseline. All of the severity factors associated with outcome at 3-4 months were also associated with the outcome at 6-8 months post-baseline, although there was greater heterogeneity at this later time-point, see Table 3.6. There was one other factor that was significantly associated with prognosis at this time point, which was not significantly associated with outcomes at 3-4 months post-baseline: a comorbid diagnosis of mixed anxiety and depressive disorder, after adjusting for treatment, age, gender, and baseline depressive symptom severity the depressive symptom scale scores at 6-8 months were lower on average: difference in mean = -0.14(95%CI: -0.24 to -0.04).

The baseline factor with the largest magnitude of association with prognosis at 6-8 months was once again depressive symptom severity: for every standard deviation increase in scores at baseline the score at 6-8 months was higher on average (mean difference =0.38(95%CI: 0.29 to 0.48)), after adjusting for treatment, age and gender. Heterogeneity in this effect was high as the effect in IPCRESS was considerably higher than in the other studies (mean difference =0.62(95%CI: 0.47 to 0.76), all other studies had effect estimates between 0.18 and 0.46). However, the impact of this was not particularly large as it was the study with the lowest weighting of all those included due to sample size, and there was little impact on the pooled effect estimate when this study was removed, see Sensitivity Analysis section below. As with the primary outcomes, most of the CIS-R subscale scores and duration items were associated with prognosis at 6-8 months after adjusting for treatment, age, gender, symptom severity, and covariates. Those with the largest magnitude of effects were: the duration of depression (mean difference =0.07(95%CI: 0.02 to 0.12)), the average duration of anxiety (mean difference =0.11(95%CI: 0.04 to 0.18)), panic disorder (mean difference = 0.24(95%CI: 0.02 to 0.47)), and the history of depression variables relating to antidepressant treatment (mean difference =0.11(95%CI: 0.03 to 0.19)) and any type of past treatment (mean difference = 0.10(95%CI: 0.01 to 0.19)). As with the primary outcomes, there was limited evidence for the association with prognosis when the log outcome was used with these history of depression variables, with confidence intervals overlapping zero.

Table 3.6. Outcomes at 6-8 months post-baseline (difference in z-score of depressive symptoms and percentage difference in depressive symptoms) per unit increase in baseline prognostic indicators.

	Difference in z-score of depressive symptoms or % difference in depressive symptoms at 6-8 months post-baseline per unit increase in baseline prognostic indicator											
	Adjusted for Treatment, Age and Gender[ρ]				Depressive symptom severity adjusted*				Depressive symptom severity and covariate adjusted* or [‡]			
Prognostic Indicator	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**
Depressive symptom severity	0.38(0.29 to 0.48)	85	33.41(23.00 to 44.70)	75	0.38(0.29 to 0.48)	85	33.41(23.00 to 44.7)	75	0.38(0.29 to 0.48)	85	33.41(23.00 to 44.7)	75
CIS-R Total Score	0.04(0.03 to 0.04)	0	13.61(11.13 to 16.15)	0	0.02(0.01 to 0.03)	60	5.98(2.28 to 9.81)	38	0.02(0.01 to 0.03)	60	5.98(2.28 to 9.81)	38
Depressive Subscales Total[‡1]	0.08(0.06 to 0.11)	77	27.87(16.30 to 40.59)	64	0.04(0.01 to 0.06)	70	8.77(-3.79 to 22.97)	73	0.03(0.00 to 0.05)	62	7.67(-3.69 to 20.38)	67
Anxiety Subscales Total[‡2]	0.04(0.02 to 0.05)	83	11.22(4.12 to 18.82)	78	0.02(0.00 to 0.03)	81	3.40(-2.56 to 9.73)	66	0.01(0.00 to 0.03)	78	3.57(-2.42 to 9.92)	66
Compulsions Score	0.10(0.05 to 0.16)	54	6.35(2.95 to 9.86)	1	0.04(-0.03 to 0.11)	72	0.81(-3.84 to 5.69)	50	0.04(-0.03 to 0.11)	72	0.81(-3.84 to 5.69)	50
Compulsions Duration[‡1]	0.05(0.03 to 0.08)	0	3.41(1.41 to 5.46)	0	0.02(0.00 to 0.05)	4	1.23(-1.19 to 3.71)	31	0.02(0.00 to 0.05)	14	1.27(-1.01 to 3.60)	25
Concentration Score	0.16(0.10 to 0.21)	75	9.74(5.76 to 13.87)	31	0.07(0.02 to 0.13)	73	3.30(0.18 to 6.51)	0	0.07(0.02 to 0.13)	73	3.30(0.18 to 6.51)	0
Concentration Duration	0.10(0.04 to 0.17)	79	8.09(1.87 to 14.70)	79	0.06(0.01 to 0.10)	62	4.61(-0.01 to 9.45)	65	0.06(0.01 to 0.10)	62	4.61(-0.01 to 9.45)	65
Depression Score[‡3]	0.17(0.11 to 0.22)	60	13.33(7.30 to 19.70)	49	0.04(-0.02 to 0.10)	62	2.59(-4.80 to 10.56)	68	0.02(-0.06 to 0.10)	65	2.49(-4.36 to 9.83)	63
Depressive Thoughts Score[‡1]	0.20(0.14 to 0.25)	56	15.07(10.93 to 19.37)	16	0.08(0.04 to 0.13)	38	4.20(-1.77 to 10.53)	56	0.06(0.01 to 0.11)	29	3.62(-1.83 to 9.38)	48
Depression Duration[‡1]	0.12(0.07 to 0.18)	56	10.08(4.39 to 16.09)	55	0.08(0.03 to 0.13)	55	6.37(0.82 to 12.23)	57	0.07(0.02 to 0.12)	49	5.61(0.52 to 10.95)	50
Fatigue Score[‡1]	0.08(0.02 to 0.14)	52	4.83(-0.89 to 10.87)	30	0.05(-0.01 to 0.10)	43	2.91(-2.28 to 8.39)	23	0.03(-0.02 to 0.08)	25	3.15(-1.81 to 8.35)	17
Fatigue Duration	0.11(0.06 to 0.15)	40	8.31(4.84 to 11.89)	3	0.08(0.04 to 0.11)	22	6.04(2.79 to 9.39)	0	0.08(0.04 to 0.11)	22	6.04(2.79 to 9.39)	0
Generalised Anxiety Score[‡1]	0.09(0.06 to 0.12)	0	5.00(2.28 to 7.79)	0	0.03(0.00 to 0.06)	0	0.62(-1.93 to 3.24)	0	0.03(0.00 to 0.05)	0	0.75(-1.77 to 3.34)	0
Generalised Anxiety Duration	0.05(0.00 to 0.10)	50	3.65(-0.42 to 7.89)	52	0.03(-0.01 to 0.07)	45	2.11(-1.6 to 5.96)	50	0.03(-0.01 to 0.07)	45	2.11(-1.6 to 5.96)	50
Health Anxiety Score[‡1]	0.13(0.06 to 0.20)	74	9.75(5.07 to 14.64)	48	0.06(0.02 to 0.11)	48	4.47(1.27 to 7.77)	0	0.04(0.00 to 0.09)	42	3.67(0.51 to 6.93)	0
Health Anxiety Duration[‡3]	0.07(0.03 to 0.11)	63	5.67(2.76 to 8.66)	46	0.04(0.02 to 0.06)	0	3.44(1.47 to 5.45)	0	0.03(0.01 to 0.05)	0	2.79(0.82 to 4.79)	0
Irritability Score[‡3]	0.11(0.06 to 0.16)	52	6.55(2.99 to 10.24)	16	0.04(0.00 to 0.09)	50	1.37(-1.68 to 4.51)	0	0.03(0.00 to 0.07)	10	1.93(-1.21 to 5.16)	0
Irritability Duration[‡2]	0.06(0.00 to 0.12)	74	4.41(-0.30 to 9.33)	67	0.02(-0.03 to 0.07)	65	1.82(-2.41 to 6.22)	62	0.03(-0.03 to 0.08)	66	1.96(-2.36 to 6.47)	63
Obsessions Score	0.03(-0.01 to 0.07)	53	0.87(-3.38 to 5.3)	66	-0.01(-0.05 to 0.03)	51	-2.25(-5.55 to 1.18)	48	-0.01(-0.05 to 0.03)	51	-2.25(-5.55 to 1.18)	48
Obsessions Duration	0.03(0.00 to 0.05)	0	1.86(-0.06 to 3.82)	0	-0.01(-0.03 to 0.02)	19	-0.34(-2.17 to 1.53)	0	-0.01(-0.03 to 0.02)	19	-0.34(-2.17 to 1.53)	0
Panic Score[‡2]	0.13(0.05 to 0.21)	81	8.26(1.58 to 15.37)	72	0.06(0.00 to 0.13)	73	3.15(-2.33 to 8.94)	60	0.06(-0.02 to 0.14)	78	3.14(-2.26 to 8.85)	59
Panic Duration[‡1]	0.10(0.06 to 0.15)	60	7.53(4.70 to 10.43)	32	0.06(0.01 to 0.10)	58	4.15(1.62 to 6.75)	21	0.05(0.01 to 0.09)	52	3.75(1.43 to 6.12)	11
Phobias Score[‡1]	0.15(0.07 to 0.22)	75	9.36(4.31 to 14.66)	54	0.08(0.01 to 0.15)	76	4.2(-0.10 to 8.69)	40	0.05(0.00 to 0.10)	35	3.67(-0.14 to 7.63)	26
Phobias Duration[‡3]	0.07(0.03 to 0.11)	59	5.49(2.95 to 8.09)	37	0.03(0.00 to 0.06)	42	2.71(0.42 to 5.05)	28	0.03(0.00 to 0.06)	31	2.29(0.29 to 4.32)	10

Sleep Score‡3	0.12(0.09 to 0.15)	46	9.27(5.86 to 12.78)	60	0.05(0.02 to 0.08)	30	3.52(0.31 to 6.83)	61	0.04(0.00 to 0.08)	35	2.65(-0.50 to 5.91)	50
Sleep Duration‡1	0.09(0.05 to 0.14)	69	7.36(2.26 to 12.71)	74	0.06(0.01 to 0.11)	73	4.78(-0.22 to 10.03)	76	0.05(0.01 to 0.09)	66	4.17(-0.36 to 8.9)	71
Somatic Score‡3	0.11(0.07 to 0.14)	32	7.99(5.25 to 10.80)	0	0.07(0.04 to 0.11)	28	5.43(2.81 to 8.12)	0	0.06(0.03 to 0.09)	0	4.97(2.37 to 7.63)	0
Somatic Duration‡3	0.09(0.06 to 0.11)	2	7.58(5.38 to 9.83)	0	0.06(0.04 to 0.09)	0	5.74(3.63 to 7.91)	0	0.05(0.03 to 0.08)	0	5.13(3.04 to 7.27)	0
Worry Score‡1	0.07(0.03 to 0.11)	37	5.16(2.05 to 8.36)	1	0.01(-0.02 to 0.04)	0	-0.15(-3.10 to 2.88)	0	0.01(-0.02 to 0.04)	0	0.46(-2.47 to 3.47)	0
Worry Duration	0.07(0.02 to 0.11)	49	6.17(1.94 to 10.58)	55	0.04(-0.01 to 0.09)	59	4.41(-0.11 to 9.14)	64	0.04(-0.01 to 0.09)	59	4.41(-0.11 to 9.14)	64
Average Duration of Anxiety‡3	0.23(0.14 to 0.32)	71	19.00(11.45 to 27.07)	58	0.12(0.04 to 0.21)	63	10.77(3.37 to 18.71)	58	0.11(0.04 to 0.18)	54	9.69(3.29 to 16.49)	45
Number of Comorbid CMDs‡3	0.20(0.07 to 0.33)	86	14.70(3.33 to 27.33)	85	0.08(-0.03 to 0.19)	80	5.75(-4.34 to 16.89)	82	0.07(-0.03 to 0.17)	78	4.89(-4.42 to 15.11)	79
Agoraphobia‡1	0.32(0.12 to 0.51)	28	26.20(11.50 to 42.83)	0	0.15(-0.01 to 0.32)	13	11.71(-0.85 to 25.87)	0	0.09(-0.06 to 0.24)	0	6.87(-5.21 to 20.48)	0
CFS‡2	0.27(0.15 to 0.38)	28	21.45(9.97 to 34.13)	26	0.08(-0.05 to 0.21)	49	7.17(-4.88 to 20.76)	48	0.09(-0.05 to 0.22)	48	7.62(-4.26 to 20.97)	46
GAD	0.21(0.10 to 0.33)	0	13.22(3.42 to 23.94)	0	0.03(-0.08 to 0.13)	0	0.44(-8.03 to 9.69)	0	0.03(-0.08 to 0.13)	0	0.44(-8.03 to 9.69)	0
MADD	-0.32(-0.42 to -0.22)	0	-21.44(-28.22 to -14.02)	15	-0.14(-0.24 to -0.04)	5	-9.41(-17.29 to -0.78)	43	-0.14(-0.24 to -0.04)	5	-9.41(-17.29 to -0.78)	43
OCD	0.38(0.10 to 0.66)	72	29.43(6.63 to 57.10)	62	0.06(-0.17 to 0.29)	58	3.35(-15.24 to 26.02)	61	0.06(-0.17 to 0.29)	58	3.35(-15.24 to 26.02)	61
Panic Disorder	0.39(0.01 to 0.77)	73	33.81(8.53 to 64.97)	45	0.24(0.02 to 0.47)	33	17.22(1.93 to 34.81)	0	0.24(0.02 to 0.47)	33	17.22(1.93 to 34.81)	0
Social Phobia	0.22(0.06 to 0.38)	0	14.82(1.91 to 29.36)	0	0.11(-0.04 to 0.26)	0	7.08(-4.77 to 20.41)	4	0.11(-0.04 to 0.26)	0	7.08(-4.77 to 20.41)	4
Specific Phobias‡1	0.06(-0.08 to 0.21)	0	8.67(-2.85 to 21.55)	0	0.00(-0.14 to 0.14)	0	3.52(-7.25 to 15.54)	0	0.03(-0.10 to 0.17)	0	6.43(-4.55 to 18.68)	0
History of Depression‡3	0.19(0.10 to 0.28)	43	9.28(-0.60 to 20.13)	62	0.11(0.03 to 0.20)	8	2.88(-6.18 to 12.80)	47	0.08(-0.02 to 0.18)	12	2.81(-6.20 to 12.67)	43
History of ADM Treatment	0.19(0.11 to 0.28)	0	12.00(-2.56 to 28.73)	54	0.11(0.03 to 0.19)	0	4.11(-9.35 to 19.57)	56	0.11(0.03 to 0.19)	0	4.11(-9.35 to 19.57)	56
Any past Treatment	0.20(0.10 to 0.30)	10	11.73(-2.71 to 28.32)	53	0.10(0.01 to 0.19)	1	3.54(-9.48 to 18.43)	52	0.10(0.01 to 0.19)	1	3.54(-9.48 to 18.43)	52
Functional Impairment‡3	0.30(0.15 to 0.45)	61	24.5(13.46 to 36.60)	24	0.09(-0.02 to 0.19)	22	7.22(-1.24 to 16.39)	0	0.08(-0.02 to 0.17)	6	6.25(-2.14 to 15.35)	0
Hazardous Alcohol misuse‡2	0.03(-0.19 to 0.25)	49	-1.80(-15.11 to 13.6)	23	0.00(-0.19 to 0.19)	42	-2.43(-14.38 to 11.19)	15	0.00(-0.19 to 0.19)	42	-2.51(-14.97 to 11.78)	21

ρ adjusted for treatment allocation, age, and gender; *additionally adjusted for baseline depression scale z-score; additionally adjusted for: ‡1employment status; ‡2 marital status; ‡3 employment status and marital status; Z per 1 z-score increase.

Many depressive 'disorder severity' factors were missing in two studies (AHEAD (119) & HEALTHLINES (123)). The difference when including or excluding those studies on the effects of variables which were not systematically missing in any study were negligible, see the section on sensitivity analyses below. These studies were therefore removed from further primary analyses, although they were included in sensitivity analyses. The associations between prognostic factors present in all the primary analyses with the score on the depressive symptom scales used across the studies at 3-4 months post-baseline after adjusting for treatment, depressive symptom severity, and each of the depressive 'disorder severity' factors displayed as well as covariates, are shown in Figure 3.1 for the primary analyses, and Figure 3.2 for the sensitivity analyses including all studies.

Only four variables were significantly associated with the primary outcomes independent of treatment and covariates when included in a full 'disorder severity' model in addition to depressive symptom severity (see Tables 12 and 13, note for ease of interpretation the duration items are split into durations greater than one year compared to one year or less). These were: the duration of depression, the average duration of anxiety problems, the presence/absence of comorbid panic disorder, and a history of treatment with antidepressants. The latter was only significantly associated with prognosis when using the z-score outcome, not the log outcome, although when removing two studies with little variability in this factor due to their inclusion criteria specifying patients had a to have a history of treatment resistant depression (COBALT) or to have been on antidepressants at an adequate dose for at least six weeks pre-baseline (MIR), there was greater evidence for an effect 6.34%(95%CI: 0.34 to 12.70). The sum of the anxiety subscale scores on CIS-R, and a history of any previous treatment for depression could be included in the model in place of the average duration of anxiety and a history of antidepressants respectively, though had weaker associations with the primary outcomes than those retained in the model and displayed in Table 3.7.

In order to estimate the possible clinical importance of the prognostic factors the degree of difference in symptoms relative to the mean was compared to the proportional minimal clinically important difference (MCID), previously reported to be approximately 17.5% on the BDI-II (171). The question here is not whether a given patient experienced a change in symptoms that would be considered clinically

important (as is often the case when assessing differences against the MCID), instead this additional analysis was asking the question of "what degree of difference would there be in the average endpoint depressive symptoms scores between groups of patients in one category of a given variable compared to those in other categories of that same variable?". Of the factors listed above only a one standard deviation increase (or larger) in depressive symptom severity at baseline was associated with a difference in depressive symptoms at 3-4 months post-baseline of a magnitude greater than one MCID relative to those with baseline depressive symptom scores. This was true after adjusting for treatment, age, gender, marital status, and employment status (percentage difference in endpoint symptom scores = 29.94%(95%CI: 22.74 to 53.82)), and after additionally adjusting for the four 'disorder severity' factors: 26.27%(95%CI: 19.90 to 32.99)). There was considerable heterogeneity in both of these effects though removing the study contributing most to this led to larger effects and narrower confidence intervals, see the Sensitivity Analyses section below.

The durations of both depressive symptoms and anxiety symptoms were close to reaching the threshold of one proportional MCID after adjusting for depressive symptom severity. When comparing those with durations of greater than one year at baseline to those with durations one year or less, the endpoint depressive symptom scores were approximately 14% higher, but were not particularly close to the threshold (10-11% higher) when additionally adjusting for the other 'disorder severity' factors, see Table 3.7. Considering the association of 'disorder severity' factors with prognosis independent of treatment and independent of depressive symptom severity, in general patients with combinations of depression and anxiety 'disorder severity' factors had endpoint symptoms that maybe considered higher to a clinically important degree. For example, patients that had both depression and anxiety problems lasting longer than one year, comorbid panic disorder, and a history of antidepressant treatment, on average had 36.25%(95%CI: 12.35 to 65.23, n=220) higher scores at 3-4 months post-baseline than patients with none of the above. In addition, adding all four 'disorder severity' factors to models of outcome led to gains in the variance explained in the endpoint symptom scores (adjusted R^2), which increased with each factor added, going from 16% to 27% for the z-score outcome and from 10% to 21% for the log outcome (see Table 3.8).

Table 3.7. Association of prognostic indicators with outcomes (mean difference in z-score of depressive symptoms and percentage difference in depressive symptoms) after adjusting for disorder severity.

Prognostic Indicator		Independent of treatment and depressive symptom severity‡				Independent of treatment and depressive disorder severity†			
	High on Factor/Present N(%)	Mean difference (95%CI)*	I^2	% difference in depressive symptom scale score^	I^2	Mean difference (95%CI)*	I^2	% difference in depressive symptom scale score^	I^2
Depressive symptom severity and covariates	2364(55.10)	0.42(0.38 to 0.77)	17	29.94(22.74 to 37.58)	78	0.38(0.34 to 0.41)	0	26.27(19.9 to 32.99)	69
Depression Duration◊	2005(46.80)	0.18(0.10 to 0.23)	29	14.63(5.88 to 24.10)	49	0.15(0.08 to 0.22)	3	11.51(3.51 to 20.13)	36
Average Anxiety Duration◊	2780(64.70)	0.18(0.09 to 0.22)	54	14.16(6.52 to 22.34)	47	0.12(0.04 to 0.21)	37	9.78(3.57 to 16.36)	17
Panic Disorder	399(9.30)	0.18(0.06 to 0.18)	22	12.48(4.76 to 20.77)	0	0.15(0.03 to 0.26)	10	9.06(1.50 to 17.17)	0
History of antidepressants	2787(65.00)	0.09(0.03 to 0.09)	0	4.53(-0.65 to 9.98)	0	0.08(0.01 to 0.14)	0	3.25(-1.85 to 8.61)	0

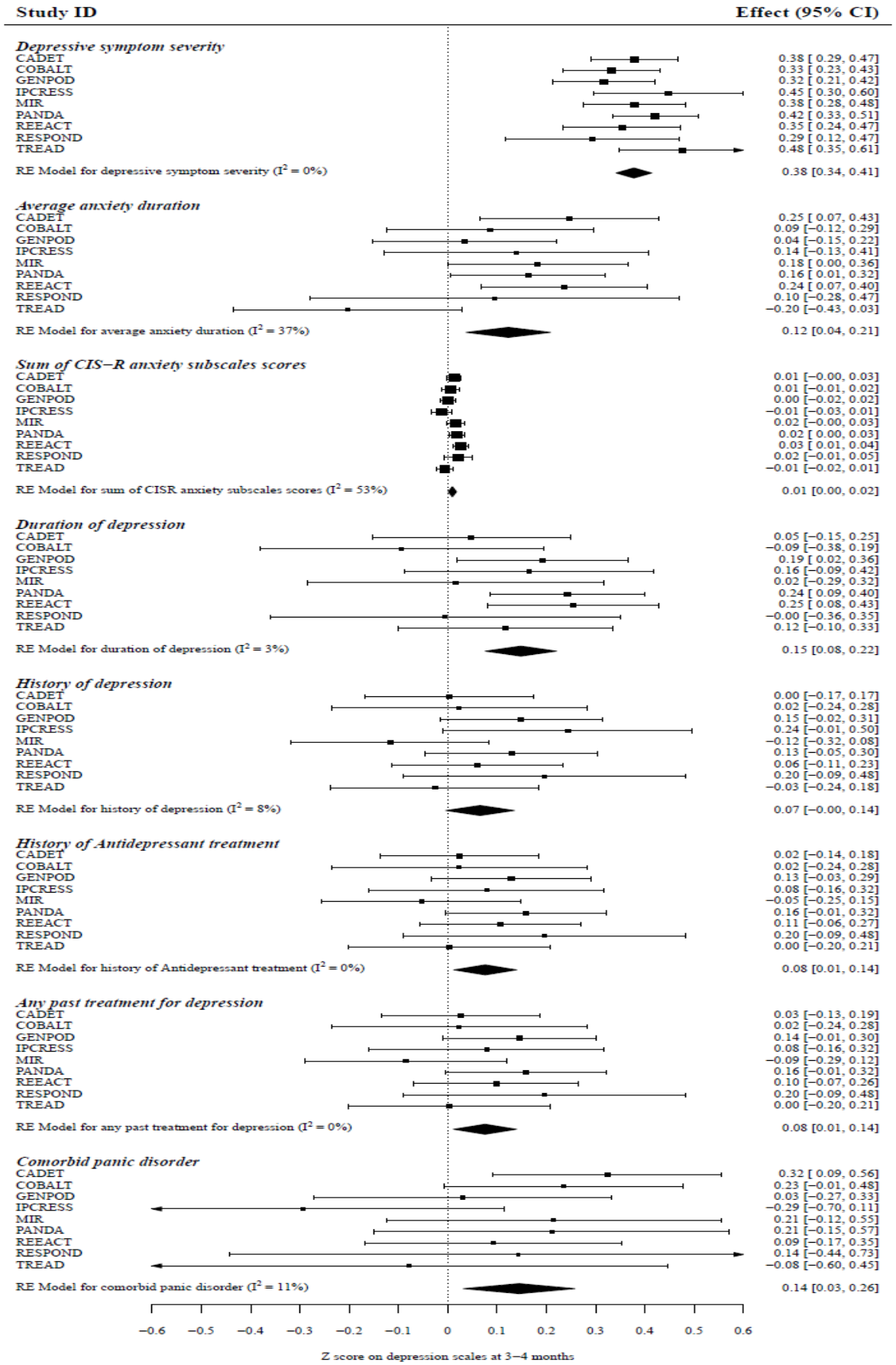
◊ dichotomised to less than or equal to 1-year, and greater than 1-year duration; *using z-score at 3-4 months as the outcome; ^ using the natural log of the depressive symptom scale scores at 3-4 months; ‡ adjusted for depressive symptom severity, treatment allocation, age, gender, employment status, and marital status. †adjusted for depressive symptom severity, depression duration, average anxiety duration, panic disorder, history of antidepressants, treatment allocation, age, gender, employment status, and marital status. All models excluded data from AHEAD & HEALTHLINES

Table 3.8. Impact on amount of variance explained in depressive symptom scale scores at 3-4 post-baseline, modelled with the z-score and natural logarithm outcomes, when adding each variable in turn.

Models, adding each variable one at a time	Cumulative impact adding each variable one at a time	
	z-score of depressive symptom scale scores adjusted R^2	log of depressive symptom scale scores adjusted R^2
Depressive symptom severity and covariates	0.16	0.10
Average anxiety duration	0.22	0.14
Depression Duration	0.25	0.19
Panic Disorder	0.27	0.20
History of antidepressants	0.27	0.21
Final model	0.27	0.21

Final model adjusted for depressive symptom severity, depression duration, average anxiety duration, panic disorder, history of antidepressants, treatment allocation, age, gender, employment status, and marital status.
All models excluded data from AHEAD & HEALTHLINES

Figure 3.1. Forest plot of associations between baseline severity factors and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors present in the included studies, and covariates.



Association between prognostic indicators and attrition at 3-4 months

Table 3.9. Difference in odds of attrition at three-to-four months post-baseline per unit increase in baseline prognostic indicators.

Prognostic Indicator	Adjusted for Treatment, Age, and Gender[ρ]		Symptom severity adjusted*		Symptom severity and covariate adjusted* or ‡	
	OR(95%CI)	I^2	OR(95%CI)	I^2	OR(95%CI)	I^2
Depressive symptom severity	1.12(0.98 to 1.28)	49	1.12(0.98 to 1.28)	49	1.12(0.98 to 1.28)	49
CIS-R Total Score	1.01(1.00 to 1.02)	43	1.00(0.99 to 1.02)	10	1.00(0.99 to 1.02)	10
Depressive Subscales Total[‡1]	1.03(1.00 to 1.06)	21	1.00(0.97 to 1.04)	16	1.00(0.97 to 1.04)	16
Anxiety Subscales Total[‡2]	1.02(1.00 to 1.04)	38	1.01(1.00 to 1.03)	1	1.01(1.00 to 1.02)	0
Compulsions Score	1.06(0.99 to 1.14)	9	1.04(0.97 to 1.11)	0	1.04(0.97 to 1.11)	0
Compulsions Duration[‡1]	1.00(0.93 to 1.07)	52	0.99(0.93 to 1.06)	41	1.00(0.95 to 1.05)	40
Concentration Score	1.01(0.92 to 1.10)	29	0.99(0.91 to 1.07)	4	0.99(0.91 to 1.07)	4
Concentration Duration	0.97(0.89 to 1.06)	51	0.94(0.86 to 1.04)	53	0.94(0.86 to 1.04)	53
Depression Score[‡3]	1.04(0.96 to 1.12)	4	0.97(0.89 to 1.05)	0	0.97(0.89 to 1.05)	0
Depressive Thoughts Score[‡1]	1.03(0.94 to 1.12)	25	0.97(0.90 to 1.05)	0	0.97(0.90 to 1.05)	0
Depression Duration[‡1]	0.97(0.91 to 1.03)	0	0.95(0.89 to 1.01)	0	0.95(0.89 to 1.01)	0
Fatigue Score[‡1]	1.00(0.92 to 1.08)	0	0.97(0.88 to 1.06)	12	0.97(0.88 to 1.06)	12
Fatigue Duration	1.03(0.95 to 1.12)	30	1.02(0.93 to 1.11)	36	1.02(0.93 to 1.11)	36
Generalised Anxiety Score[‡1]	1.04(0.96 to 1.14)	50	1.02(0.93 to 1.11)	51	1.02(0.95 to 1.09)	54
Generalised Anxiety Duration	0.98(0.90 to 1.08)	68	0.98(0.89 to 1.08)	68	0.98(0.89 to 1.08)	68
Health Anxiety Score[‡1]	1.14(1.06 to 1.23)	21	1.11(1.04 to 1.20)	8	1.11(1.02 to 1.21)	57
Health Anxiety Duration[‡3]	1.05(0.99 to 1.11)	38	1.03(0.97 to 1.10)	44	1.02(0.97 to 1.08)	45
Irritability Score[‡3]	1.00(0.93 to 1.08)	35	1.03(0.95 to 1.11)	26	1.01(0.97 to 1.06)	8
Irritability Duration[‡2]	1.02(0.97 to 1.07)	0	0.99(0.92 to 1.06)	30	0.99(0.96 to 1.01)	0
Obsessions Score	1.01(0.97 to 1.06)	0	1.00(0.96 to 1.06)	0	1.00(0.96 to 1.06)	0
Obsessions Duration	1.06(0.99 to 1.13)	0	1.00(0.96 to 1.05)	0	1.00(0.96 to 1.05)	0
Panic Score[‡2]	1.04(0.99 to 1.09)	0	1.02(0.96 to 1.09)	0	0.99(0.96 to 1.02)	0
Panic Duration[‡1]	1.09(1.02 to 1.17)	6	1.02(0.97 to 1.07)	0	0.99(0.97 to 1.02)	0
Phobias Score[‡1]	1.02(0.98 to 1.06)	0	1.06(0.99 to 1.14)	4	1.04(0.97 to 1.11)	30
Phobias Duration[‡3]	1.13(1.02 to 1.24)	41	1.00(0.96 to 1.05)	0	0.99(0.97 to 1.01)	0
Sleep Score[‡3]	1.00(0.93 to 1.08)	43	1.11(1.00 to 1.24)	49	1.11(1.00 to 1.24)	49
Sleep Duration[‡1]	1.00(0.94 to 1.07)	10	1.00(0.91 to 1.09)	54	1.00(0.91 to 1.09)	54
Somatic Score[‡3]	1.00(0.93 to 1.07)	41	0.98(0.92 to 1.04)	0	0.99(0.96 to 1.01)	0
Somatic Duration[‡3]	1.03(0.97 to 1.10)	4	0.99(0.92 to 1.05)	38	0.99(0.95 to 1.04)	29
Worry Score[‡1]	0.96(0.90 to 1.03)	21	1.00(0.93 to 1.06)	0	1.00(0.93 to 1.06)	0
Worry Duration	1.02(0.88 to 1.18)	64	0.95(0.89 to 1.01)	16	0.95(0.89 to 1.01)	16
Average Duration of Anxiety[‡3]	1.09(1.01 to 1.18)	14	0.97(0.82 to 1.14)	69	0.99(0.86 to 1.13)	67
Number of Comorbid CMDs[‡3]	1.11(0.88 to 1.40)	26	1.03(0.94 to 1.14)	20	1.03(0.95 to 1.12)	32
Agoraphobia[‡1]	1.00(0.89 to 1.14)	0	1.06(0.84 to 1.32)	22	1.06(0.86 to 1.30)	20
CFS[‡2]	1.05(0.85 to 1.29)	47	0.97(0.85 to 1.10)	0	0.98(0.96 to 1.01)	0
GAD	0.86(0.69 to 1.07)	27	0.98(0.80 to 1.22)	46	0.98(0.80 to 1.22)	46
MADD	1.19(0.96 to 1.48)	8	0.94(0.74 to 1.18)	32	0.94(0.74 to 1.18)	32
OCD	1.30(0.99 to 1.70)	0	1.10(0.89 to 1.37)	0	1.10(0.89 to 1.37)	0
Panic Disorder	1.01(0.79 to 1.29)	0	1.15(0.87 to 1.51)	0	1.15(0.87 to 1.51)	0
Social Phobia	1.18(0.96 to 1.45)	0	0.95(0.74 to 1.22)	0	0.95(0.74 to 1.22)	0
Specific Phobias[‡1]	0.99(0.82 to 1.19)	0	1.14(0.92 to 1.40)	0	1.04(0.92 to 1.17)	8
History of Depression[‡3]	0.99(0.83 to 1.18)	0	0.96(0.80 to 1.16)	0	0.96(0.80 to 1.16)	0
History of ADM Treatment	0.94(0.79 to 1.12)	0	0.97(0.81 to 1.16)	0	0.97(0.81 to 1.16)	0
Any past Treatment	1.22(1.03 to 1.44)	0	0.92(0.77 to 1.10)	0	0.92(0.77 to 1.10)	0
Functional Impairment[‡3]	0.91(0.71 to 1.18)	0	1.13(0.95 to 1.34)	0	1.05(0.91 to 1.21)	21
Hazardous Alcohol misuse[‡2]	0.91(0.70 to 1.17)	0	0.89(0.69 to 1.16)	0	0.98(0.95 to 1.01)	0

[ρ] adjusted for treatment allocation, age, and gender; *additionally adjusted for baseline depression scale z-score; additionally adjusted for: [‡1]employment status; [‡2] marital status; [‡3] employment status and marital status

A number of severity factors were associated with attrition independent of treatment. Higher severity of depressive symptoms at baseline was associated with greater odds of attrition, but unlike with the other outcomes assessed, there was limited evidence for the association (OR=1.12(95%CI: 0.98 to 1.28)). Higher durations of problems measured on a number of CIS-R anxiety subscales (panic, phobias and the mean duration across all CIS-R anxiety subscales) were associated with greater odds of attrition independent of treatment. The strongest evidence for associations with higher odds of attrition independent of treatment came from higher scores on the CIS-R health anxiety subscale (OR=1.14(95%CI: 1.06 to 1.23)), and for those with a history of any previous treatment for depression (OR=1.22(95%CI: 1.03 to 1.44)). After additionally adjusting for depressive symptom severity and covariates most associations were not significant with the exception of the health anxiety score (OR=1.11(95%CI: 1.02 to 1.21)), see Table 3.9. As with the primary outcomes, there was little evidence for associations between hazardous alcohol misuse (OR=0.91(95%CI: 0.70 to 1.17)) or functional impairment (OR=0.91(95%CI: 0.71 to 1.18)) and attrition.

Sensitivity Analyses

In sensitivity analyses using variables available in all studies, the indicators of 'disorder severity' that were associated with prognosis in the final models were: the duration of depression; anxiety symptom severity; and a history of antidepressant treatment, all independent of treatment, depressive symptom severity, covariates, and each other, see Tables 3.10 and 3.11, and Figure 3.2. As with the primary analyses, symptom severity was most strongly associated with prognosis, with each standard deviation increase at baseline associated with approximately 26% higher depressive symptom scale scores at 3-4 months post-baseline, see Table 3.10. Again similar to the primary analyses, the association between a history of antidepressant use and prognosis was significant when using the z-score outcome but not with the log outcome, however again it contributed to increases in the amount of variance explained with both outcomes, see Table 3.11. Overall, the amount of variance explained with these four factors (three indicators of depressive 'disorder severity' and depressive symptom severity) was marginally lower than with the five factors from the primary analyses (24% for the z-score outcome and 18% for the log outcome; compared to 27% and 21% respectively in the primary analyses).

Heterogeneity was somewhat higher with the factors assessed in all studies than in the primary analyses, particularly regarding the anxiety symptom severity variable, further sensitivity analyses removing studies to reduce the heterogeneity in this and other effects are presented in Table 3.12.

In all analyses removing studies due to heterogeneity resulted in very small differences in magnitudes of effect and had no impact on the direction of effects or conclusions that might be drawn from the effects found in the primary analyses, see Table 3.12. In addition, there were only very small differences in the magnitude (0.02 of a standard deviation in effect) of association of depressive symptom severity and prognosis when comparing a univariate meta-analysis of the z-score outcome at 3-4 months with a bivariate meta-analysis using both the 3-4 and 6-8 months z-score outcomes. There was an even smaller degree of difference (0.01 of a standard deviation) in associations when including or excluding the two studies that were systematically missing many ‘disorder severity’ variables, and a similarly small difference in associations when using observed data (not imputed) compared to the main analyses using imputed data.

Table 3.10. Association of prognostic indicators with outcomes adjusted for disorder severity, and impact in accuracy of models after adding each variable in turn. All variables in all studies.

Models, adding each variable one at a time	High on Factor/Present N(%)	Independent of treatment and depressive symptom severity‡				Independent of treatment and depressive disorder severity†			
		Mean difference (95%CI)*	I^2	% difference in depressive symptom scale score^	I^2	Mean difference (95%CI)*	I^2	% difference in depressive symptom scale score^	I^2
Depressive symptom severity and covariates	2759(52.8)	0.44(0.41 to 0.47)	16	30.74(24.94 to 36.82)	78	0.37(0.33 to 0.42)	38	26.03(19.65 to 32.74)	76
Z-score of main anxiety scale	2575(49.3)	0.12(0.09 to 0.15)	72	7.27(3.77 to 10.9)	52	0.11(0.05 to 0.17)	73	7.36(3.83 to 11.00)	53
History of antidepressants	3436(65.8)	0.10(0.04 to 0.16)	24	3.73(-2.12 to 9.92)	40	0.08(0.01 to 0.15)	27	2.93(-2.84 to 9.05)	40

*using z-score at 3-4 months as the outcome; ^ using the natural log of the depressive symptom scale scores at 3-4 months; ‡ adjusted for depressive symptom severity, treatment allocation, age, gender, employment status, and marital status. † adjusted for depressive symptom severity, z-score of baseline anxiety scale scores, history of antidepressants age, gender, and treatment allocation

Table 3.11. Impact in accuracy of models after adding each 'disorder severity' variable in turn. All variables in all studies.

Models, adding each variable one at a time	Cumulative impact adding each variable one at a time	
	z-score of depressive symptom scale scores adjusted R^2	log of depressive symptom scale scores adjusted R^2
Depressive symptom severity and covariates	0.16	0.10
Z-score of main anxiety scale	0.23	0.17
History of antidepressants	0.24	0.18
Final model†	0.24	0.18

†Depressive symptom severity, z-score of baseline anxiety scale scores, history of depression, adjusted for treatment allocation, age, and gender

Figure 2.2. Forest plot of associations between baseline severity factors and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors present in all studies, and covariates.

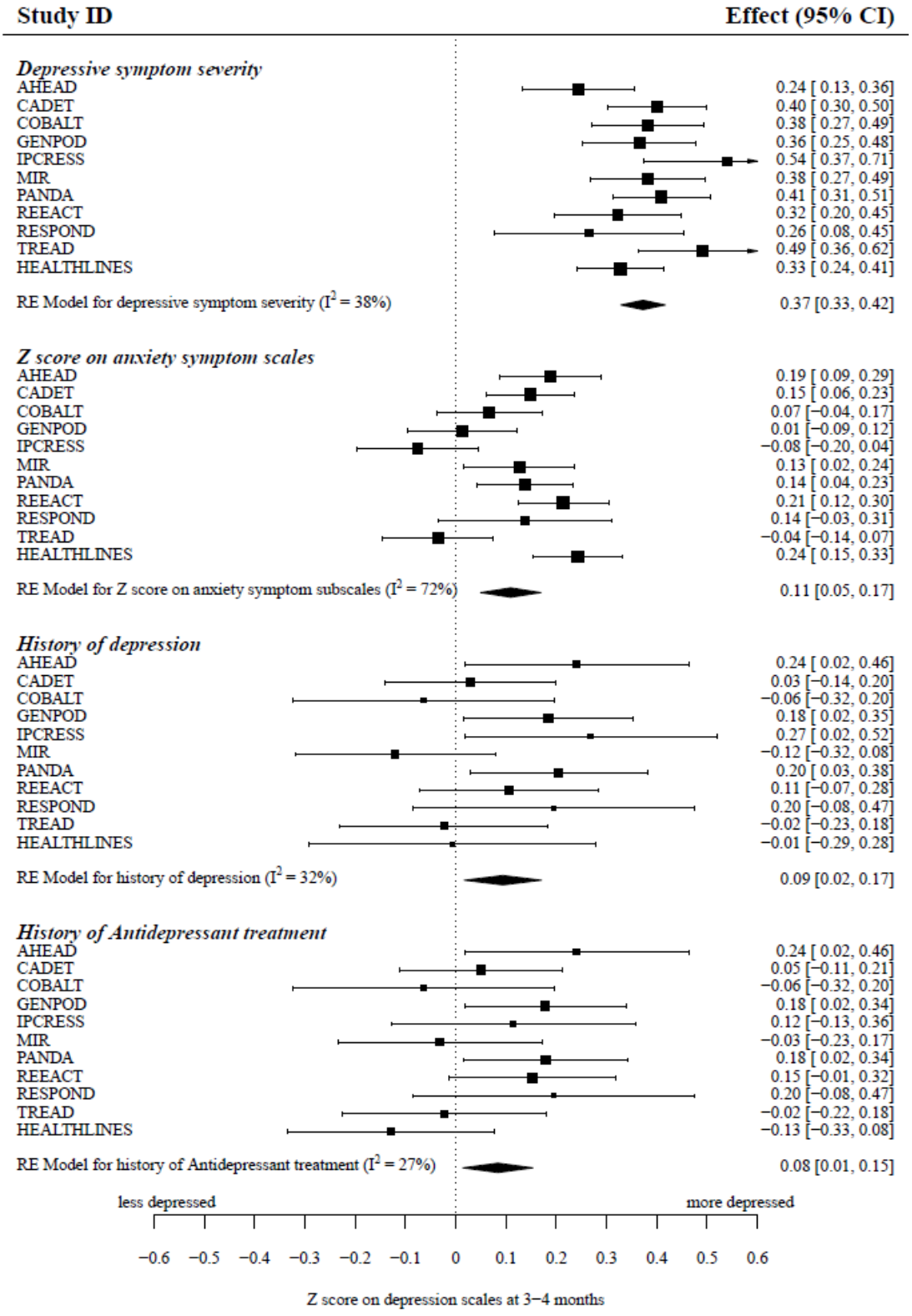


Table 3.12. Results of original analyses and corresponding sensitivity analyses.

Analysis	Change for Sensitivity Analysis	Pooled Effect Estimate
z-score 3-4 month outcome		**mean difference (95%CI)**
Depressive symptom severity[ρ]	Original Analysis using z-score of depressive symptoms at 3-4 months	0.44(0.41 to 0.47)
	bi-variate meta-analysis using both 3-4 month and 6-8 month	0.42(0.36 to 0.48)
Depressive symptom severity[ρ]	Analysis using all 11 studies irrespective of systematically missing data	0.44(0.41 to 0.47)
	Removing two studies with systematically missing data on many 'disorder severity factors' (AHEAD & HEALTHLINES)	0.45(0.42 to 0.49)
Depressive symptom severity[ρ]	Original Analysis using z-score of depressive symptoms at 3-4 months in imputed data	0.44(0.41 to 0.47)
	Analysis using observed 'un-imputed' data	0.43(0.39 to 0.47)
Anxiety Subscales Total[‡2]	Original Analysis using z-score of depressive symptoms at 3-4 months	0.04(0.03 to 0.05)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.04(0.04 to 0.05)
Panic Score[ρ]	Original Analysis with all studies	0.13(0.07 to 0.19)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.15(0.10 to 0.20)
Phobias Score[ρ]	Original Analysis with all studies	0.15(0.10 to 0.20)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.17(0.14 to 0.20)
Number of Comorbid CMDs[ρ]	Original Analysis with all studies	0.21(0.12 to 0.29)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.24(0.19 to 0.30)
log outcome at 3-4 month		**%(95%CI)**
Depressive symptom severity[ρ]	Original Analysis using z-score of depressive symptoms at 3-4 months	30.74(24.94 to 36.82)
	Analysis removing two studies contributing most to heterogeneity (RESPOND and PANDA)	30.57(25.18 to 36.19)
z-score 6-8 month outcome		**mean difference (95%CI)**
Depressive symptom severity[ρ]	Original Analysis with all studies	0.38(0.29 to 0.48)
	Analysis removing study contributing most to heterogeneity (IPCRESS)	0.35(0.26 to 0.44)
Depressive Subscales Total[ρ]	Original Analysis with all studies	0.04(0.01 to 0.06)
	Analysis removing study contributing most to heterogeneity (IPCRESS)	0.05(0.03 to 0.07)
Anxiety Subscales Total*	Original Analysis with all studies	0.02(0.00 to 0.03)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.02(0.01 to 0.04)
Concentration Score	Original Analysis with all studies	0.16(0.10 to 0.21)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.18(0.13 to 0.22)
Concentration Duration	Original Analysis with all studies	0.10(0.04 to 0.17)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.12(0.05 to 0.19)
Panic Score[‡]	Original Analysis with all studies	0.06(-0.02 to 0.14)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.09(0.01 to 0.17)
Phobias Score*	Original Analysis with all studies	0.05(0.00 to 0.10)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.07(0.03 to 0.11)
Number of Comorbid CMDs[‡]	Original Analysis with all studies	0.07(-0.03 to 0.17)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.10(0.00 to 0.20)
log outcome at 6-8 month		**%(95%CI)**
Depressive symptom severity[ρ]	Original Analysis with all studies	33.41(23.00 to 44.70)
	Analysis removing study contributing most to heterogeneity (IPCRESS)	30.01(20.34 to 40.47)
Anxiety Subscales Total[ρ]	Original Analysis with all studies	2.27(0.94 to 3.63)
	Analysis removing study contributing most to heterogeneity (TREAD)	2.94(2.29 to 3.60)
Sleep Duration*	Original Analysis with all studies	6.24(1.58 to 11.11)
	Analysis removing study contributing most to heterogeneity (TREAD)	7.61(2.79 to 12.66)
Number of Comorbid CMDs[‡]	Original Analysis with all studies	12.89(2.51 to 24.32)
	Analysis removing study contributing most to heterogeneity (TREAD)	17.92(10.33 to 26.02)
Remission outcome at 3-4 month		**OR(95%CI)**
Number of Comorbid CMDs[ρ]	Original Analysis with all studies	0.72(0.61 to 0.83)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.67(0.60 to 0.75)
Final Model Variables		
Depressive symptom severity[‡]	Original Analysis using z-score of depressive symptoms at 3-4 months	29.94(22.74 to 37.58)
	Analysis removing study contributing most to heterogeneity (RESPOND)	32.07(24.82 to 39.74)
Depressive symptom severity[†]	Original Analysis using log of depressive symptoms at 3-4 months	25.24(19.01 to 31.79)
	Analysis removing study contributing most to heterogeneity (RESPOND)	27.11(21.18 to 33.33)
Average Anxiety Duration[‡]	Original Analysis with all studies	0.23(0.18 to 0.28)
	Analysis removing study contributing most to heterogeneity (TREAD)	0.25(0.21 to 0.29)
Depressive symptom severity[‡]	Original Analysis using log of depressive symptoms at 3-4 months in all 11 studies	29.94(22.74 to 37.58)
	Analysis removing study contributing most to heterogeneity (RESPOND)	32.07(24.82 to 39.74)
Depressive symptom severity[†]	Original Analysis using log of depressive symptoms at 3-4 months in all 11 studies	25.24(19.01 to 31.79)
	Analysis removing study contributing most to heterogeneity (RESPOND)	27.11(21.18 to 33.33)
Z-score of main anxiety scale[‡]	Original Analysis using z-score of depressive symptoms at 3-4 months in all 11 studies	0.08(0.02 to 0.14)
	Analysis removing study contributing most to heterogeneity (PANDA)	0.07(0.04 to 0.70)
Z-score of main anxiety scale[†]	Original Analysis using z-score of depressive symptoms at 3-4 months in all 11 studies	0.08(0.02 to 0.14)
	Analysis removing study contributing most to heterogeneity (PANDA)	0.07(0.00 to 0.14)
Z-score of main anxiety scale[‡]	Original Analysis using log of depressive symptoms at 3-4 months in all 11 studies	5.50(1.62 to 9.53)
	Analysis removing study contributing most to heterogeneity (PANDA)	4.90(0.65 to 9.33)
Z-score of main anxiety scale[†]	Original Analysis using log of depressive symptoms at 3-4 months in all 11 studies	5.53(1.67 to 9.54)
	Analysis removing study contributing most to heterogeneity (PANDA)	4.91(0.63 to 9.37)

[ρ] adjusted for treatment allocation, age and gender only; *additionally adjusted for depressive symptom severity; ‡additionally adjusted for covariates (employment status and/or marital status); [†] additionally adjusting for disorder severity factors

Discussion

Depressive symptom severity was strongly associated with prognosis independent of treatment. Depressive symptom scale scores were on average 30% higher at 3-4 months post-baseline, 33% higher at 6-8 months post-baseline, and 31% fewer patients reached remission at 3-4 months post-baseline, for every standard deviation increase in baseline depressive symptoms. Absolute differences were also assessed: for every 11-point increase in BDI-II scores at baseline, on average BDI-II scores were about seven points higher at 3-4 months. For the studies that used the PHQ-9, the differences were scores about five points higher at 3-4 months for every five points higher at baseline.

Nearly all 'disorder severity' factors were also associated with prognosis independent of treatment but only a handful of these were associated with prognosis independent of depressive symptom severity. This illustrates the importance of adjusting for baseline depression symptom severity when investigating prognosis of depression. The factors independently associated with prognosis were: the duration of depression; average duration of anxiety (or severity of anxiety symptoms); comorbid panic disorder; and a history of treatment with antidepressants (or a history of any past treatment for depression irrespective of treatment type). A history of antidepressant treatment was associated with an approximate 5% difference in symptoms at 3-4 months post-baseline, the other three 'disorder severity' factors were all associated with greater magnitudes of difference independent of depressive symptom severity (between 12-15%). These factors were all associated with lower odds of remission and the proportion of patients reaching remission at 3-4 months was considerably lower when these factors were high or present, independent of treatment and depressive symptom severity. For example, patients with comorbid panic disorder were approximately 35% less likely to remit compared to those without panic disorder, and among those with a history of antidepressant medications approximately 11% fewer patients remitted compared to patients without such a history. Similarly, the same four factors were associated with prognosis at 6-8 months post-baseline independent of treatment and depressive symptom severity. Considering combinations of these disorder severity factors may give rise to clinically important differences in prognosis independent of the information gained from

assessing depressive symptom severity. For example, in the small subgroup of patients that had both durations longer than one year, a history of antidepressants, and comorbid panic disorder, relative to those participants with none of these features, the depressive symptom scale scores at 3-4 months were approximately 36% higher. In contrast to the extant systematic review and IPD literature, the associations here have been investigated independent of a wide range of treatments in primary care.

There were also some important negative findings. There was no evidence of an association between functional impairment and prognosis, independent of treatment and depressive symptom severity with any of the prognostic outcomes assessed here. Functional impairment has been found to be indicative of treatment response for people with either depression or anxiety concerns in the past (173–175). So, it could be that the single item used to capture functional impairment here (via the CIS-R) is not sufficient, although the association had not previously been assessed independent of treatment, nor in a sample of participants seeking treatment just for depression. Hazardous alcohol misuse was also not associated with any of the prognostic outcomes assessed here, independent of treatment, adding further weight to previous studies which have found that it is not related to treatment outcomes after adjusting for baseline depressive and anxiety symptoms (99). However, in comparison to a recent study in a primary care mental health population (99), the proportion of the present study sample that were misusing alcohol was relatively low. Further, previous studies have found that for those misusing alcohol, large reductions in use typically take place between initially presenting to a health professional and starting treatment (176). Therefore, it could be that in the studies included in Dep-GP, those misusing alcohol were already reducing their use at the point of their baseline assessments, and this may have mitigated the effect of alcohol misuse on prognosis. It is noteworthy though that the sample size for the analysis of hazardous alcohol use was much lower than most other factors assessed here (n=3026), and if the true effect were nearer the upper confidence interval found here, then it would have been considered to be associated with prognosis. In addition, with the exception of the score on the health anxiety subscale of the CIS-R, there was limited evidence for an association between the markers of severity considered here

and attrition independent of treatment, after adjusting for baseline depressive symptom severity.

Findings in Context

This study provides confirmation that depressive symptom severity is the strongest indicator of prognosis independent of treatment. A number of other studies have found symptom severity to be associated with outcomes but none have considered the association independent treatment, nor across a broad range of available treatments in primary care settings. This study was also the first to investigate a multitude of 'disorder severity' factors (that is those factors related to the severity of depression beyond depressive symptoms) associated with prognosis independent of treatment, and consider the benefit of assessing them over and above depressive symptom severity. There had been some suggestion from past studies that the duration of depression might be associated with prognosis although there were inconsistencies and contradictory findings in past reviews. In addition, there was limited evidence that comorbid anxiety and a history of antidepressant use maybe associated with outcomes from antidepressant treatments, but perhaps not other types of treatment. Here these were found to be associated with prognosis independent of treatment type, and evidence for two novel associations with prognosis were also found: the average duration of anxiety problems, and comorbid panic disorder.

Strengths and Limitations

There are a number of strengths to this study. It utilised a large dataset with over 6000 participants, all of whom were assessed with the same assessment measure: the CIS-R, and all of whom were recruited while seeking treatment in naturalistic settings. This allowed for the examination of a variety of the features of depressive 'disorder severity' and prognosis. The studies included in the Dep-GP IPD dataset also have a range of treatments, including commonly prescribed antidepressants, cognitive behavioural therapy of high and low intensities, and structured treatments such as physical activity and supportive counselling. As such, we can be confident that the associations here might generalise to other settings in which similar treatments are offered. The study was not concerned with causal relationships, so

confounding is less important here, but adjustments for a number of baseline covariates were made adding robustness to the findings.

A major limitation is that the population studied here had been recruited to participate in randomised controlled trials. This may therefore be a biased sample of all patients with depression and could limit the generalisability of the findings. However, 11 of the 12 studies were pragmatic trials so the participants here should be more representative of other depressed patients in primary care than would be the case if the included studies were not designed to recruit in such a way (117).

The data on durations were self-reported and relied upon a retrospective judgement about duration that is likely to have increased measurement error. In general, random measurement error would tend to bias associations towards the null (153). Although it is possible that those with more depressive symptoms reported longer durations of illness because of a negative cognitive bias (177). Adjustments were made for baseline depressive symptoms, minimising such bias. In any case, knowing that self-reported duration is a prognostic factor is of clinical value even if this might be partly influenced by the severity of symptoms.

Heterogeneity in some of the associations was high when considering the I^2 statistic, in the study protocol it was specified that sensitivity analyses would be run where I^2 was above 75% for all factors or above 50% for factors included in the final models, or if there were clear differences between the effects across the studies included in the IPD. More conservative limits for heterogeneity could have been set, but given that none of the sensitivity analyses substantively changed the findings related to any of the prognostic indicators and given that all models were run with random effects for study, it seems unlikely that this would have had a meaningful impact on the results presented here. In addition, assessments of attrition were limited by a lack of data on the reasons for attrition and for details of precisely when attrition occurred or the amount of treatment received/taken at the point of attrition.

Implications

It is difficult to objectively assess the clinical importance of prognostic factors. One approach is to compare the differences observed here with estimates for the minimal

clinically important difference (MCID). Previous work has suggested this is about 17.5% for the BDI-II (171). By this criterion, increases in depressive symptom severity by one standard deviation at baseline were associated with clinically important differences in outcomes. The 'disorder severity' factors found to be associated with prognosis might not be associated with clinically important differences in outcomes in their own right. However, when considered in combination, these factors could be useful for clinicians: there were differences of approximately 36% in symptoms at the 3-4 month endpoint relative to the mean when participants were in the high severity category of all four factors. These differences would almost certainly be considered clinically important though only a small proportion of the patients in this study sample were in that subgroup. When modelling prognosis, the inclusion of these four factors might also lead to improvements in the ability to predict prognosis independent of treatment, relative to using depressive symptom severity alone. However, most of the variance in primary outcomes was unexplained in this study and although unexplained variance includes measurement error, this could be concerning. The final 'disorder severity' model of prognosis here (adjusted R^2 of 27% for the z-score outcome and 21% for the log outcome) compares favourably with the approximately 9% of variance explained in other studies using similar constructs to consider prognosis in primary care (101). That notwithstanding, if these findings were applied clinically, caution would be required as there are likely to be a number of unknown factors that could also have an impact on prognosis.

In terms of attrition, it is not surprising that patients with higher degrees of concern about their health, comorbid to depression, would be more likely to leave the study or stop taking their randomised treatment, for example if they experienced more side effects from treatment. However, it is somewhat surprising that other markers of severity were not similarly associated with attrition (178). These findings may imply that greater granularity in attrition related data is required in order to better understand these associations, or that for some patients mitigation of their health anxiety on attrition may require other interventions. Data were not available on side effects from treatment, or on patient's expectations of their own outcomes with their randomised treatment. Such information may be particularly useful in better understanding the implications of the above association (101).

Conclusions

Depressive symptom severity had a strong association with prognosis independent of treatment. The duration of depressive symptoms, duration of anxiety symptoms, panic disorder and past antidepressant use were all also associated with prognosis independent of depressive symptom severity and treatment, and health anxiety symptom severity was associated with attrition. Consideration of these factors could be clinically important for determining prognosis and informing patients and clinicians about likely outcomes in the treatment of depression.

Chapter 4. The association between social support and prognosis.

Overview

In Chapter 3 I described the findings of an investigation into the associations between a number of factors that can be considered as part of the overall picture of the severity of a patient's depressive disorder (which I termed depressive 'disorder severity') and their prognosis independent of treatment, and independent of depressive symptom severity. I found that several such factors can contribute meaningfully to considerations of prognosis in the above context, however there was no assessment of factors that might not be considered to be part of the construct of 'disorder severity'. Social support has long been thought to be important in the onset and maintenance of depression, and has been highlighted by depressed patients as being influential to their ability to seek, engage with, adhere to, and ultimately benefit from treatment for their depression. It was also found to be associated with prognosis in the review presented in Chapter 1. This chapter therefore focusses on the associations of social support and individual sub-types of it, with prognosis independent from treatment, and independent firstly of depressive symptom severity, and secondly from the depressive 'disorder severity' factors found to be associated with prognosis in Chapter 3 (sub-aim 3); and with attrition (sub-aim 4).

Abstract

Background

People with depression consider social support to be important to their ability to engage in treatment, recover from depression and for their longer-term prognosis. High levels of social support are associated with good outcomes from certain treatments but the association has not been tested independent of treatment and independent of a variety of markers of the severity of depression. Whether there are individual items measuring some specific aspect of social support that are more strongly associated with prognosis independent of treatment and attrition than others, is also unknown.

Methods

Data from all individual participants of six RCTs (n=2858) in the Dep-GP IPD dataset were included in the analyses. All studies included adults randomised to any treatment following presentation to a GP with depression. Participants all completed the same baseline assessments of social support, depressive symptom severity, and depressive 'disorder severity' factors. Data were analysed with two-stage random effects meta-analyses.

Results

Social support was associated with prognosis and with attrition, independent of treatment. Adjusting for depressive severity reduced the magnitude of effects but higher levels of social support were still associated with lower odds of attrition and better prognoses. There was no clear evidence that individual social support scale items were differentially associated with the outcomes. However, there were differences in the magnitudes of effects; the items most strongly associated with prognostic outcomes were feeling accepted by others for who one is, feeling supported or encouraged by family or friends, and feeling cared about. The latter was also most strongly associated with attrition. Risk of bias was low in all studies, quality to determine prognostic effects of social support was high, and heterogeneity in effects was low.

Conclusions

Overall, social support was significantly associated with prognosis and attrition but small differences in social support may not be associated with clinically important differences in outcomes, after accounting for depressive symptoms. The Social Support Scale used here or single questions from it may be added to routine clinical

assessments and may be informative for prognosis and attrition, aiding in the clinical management of depression.

Introduction

The study outlined in Chapter 3 found that the most impactful predictor of outcome independent of treatment was depressive symptom severity pre-treatment. A number of other factors related to the severity of a patient's experience of depression, such as anxiety symptoms, the chronicity of anxiety problems, chronicity of depression, a comorbid diagnosis of panic disorder, and potentially a history of antidepressant treatment for depression, which I termed depressive 'disorder severity' factors were also associated with prognosis independent of treatment and independent of depressive symptom severity. However, when adding these disorder severity factors to a model of prognosis containing depressive symptom severity, the improvements in terms of variance explained by the model were modest and most of the variance remained unexplained. Therefore, it might be important to consider other factors that could be associated with prognosis for adults with depression, independent of treatment, and independent of depressive 'disorder severity'. This could be particularly true for those factors that may be modifiable such that whatever the treatment there is an opportunity to affect engagement in treatment and outcomes from treatment.

Soicial Support and Depression

From the review presented in Chapter 1, four study-level systematic reviews investigated the association between social support and prognosis (either in the "natural course" or in response to antidepressant medications). The studies broadly found evidence that lower social support was associated with poorer prognosis. However, each review contained very few primary studies which investigated the associations, there were a number of methodological problems with the studies including the combining of prognostic outcomes (such as treatment response and relapse), over varying time points (from two weeks to two years), and a combination of different ways of measuring and quantifying social support. Thus making it difficult to interpret sources of heterogeneity. There was also a lack of clarity on the setting and context of recruitment of participants, and combinations of some treated patients with mainly non-treated/community based samples, making it difficult to interpret the generalisability of the findings. Further, as stated previously, none of the studies investigated prognosis independent of treatment. So, the question remains as to whether social support is associated with prognosis in that context, and whether

such an association would hold after adjusting for depressive symptom severity and depressive ‘disorder severity’.

Defining social support

There is no universally accepted definition of social support but it is usually thought of as encompassing a subjective sense of the quality or value in the structure and function of one’s social contacts (179). For the purposes of this thesis I propose a working definition of social support as: an individual’s perception that they are cared for, esteemed, loved, or valued by their peers, friends or family, and are part of a social network that can be mobilised when needed (180). A number of authors have suggested that such a definition would necessarily include a number of potential dimensions of social support (179–181). Three dimensions have been commonly considered in past research: emotional support (demonstrations of love, caring, encouragement, or sympathy); informational support (the provision of support to solve problems or cope with them), and instrumental support (the provision of practical support in terms of behaviour, goods, or finances) (179–181). There is some evidence to suggest that emotional or instrumental support are associated with a lower likelihood of experiencing depression among non-patient or general population samples (82). However, reverse causality cannot be ruled out as many studies of these associations have been cross-sectional, and only one study has assessed the effect of informational support on such an outcome (82). So, there is a lack of evidence that the subtypes or dimenions have differential associations with prognosis. Further, there is a lack of clear evidence to suggest that the three subtypes of social support are clearly separable by or within individuals (179,182). That being the case, the three above subtypes will not be used to inform the anlayses presented here. However, it might be the case that certain aspects of social support ascertained by individual items on a social support measure may have differential associations with prognosis (137). It might also be the case that any association between social support and prognosis for depressed patients may be encompassed by only a small number of items from a scale measuring social support, if so, then assessments might be shortened by using just those items, saving time without losing important information. Such information might have utility in both research and clinical practice, and as such, the associations between

individual social support scale items and prognosis are of relevance to the aims of this thesis.

Theoretical links between social support and depression, and proposed mechanisms

The association between social support and the onset of depression or prognosis for depressed patients fits with biopsychosocial conceptualisations of depression and other mental health problems (183). As such, there are a number of potential mechanisms by which social support might affect prognosis for people with depression. Close ties with a supportive spouse, partner, friends or colleagues will often result in the recognition of changes in patterns of behaviour, outward signs of health or ill-health, and observable aspects of low mood (180). Those with greater social support might therefore be more able to rely on members of their social network to identify, highlight or discuss such observations at the point at which they become depressed, and to make suggestions or support them in seeking treatment for the depression (180). Further, those with higher degrees of social support or closer social ties, may have others within their social network whom they can observe acting in health promoting ways when faced with difficulties similar to those which the patient is themselves going through, and this modelling can in turn engender help seeking or normative health related behaviours (82,180).

The role one plays in the lives of others is also directly related to one's levels of social support and to the likelihood that one will act in ways that promote health in order maintain those roles. As such, this might impact upon treatment seeking and adherence to or engagement with treatment, and subsequently therefore to treatment outcomes (179,180). Linked to the impact of social roles on prognosis is the association between social roles and self-esteem, the latter is sometimes considered a symptom of depression (as discussed in Chapter 1) or as a construct strongly associated with the degree of severity of depression (184). Those with higher self-esteem and higher social support are also more likely to experience greater degrees of mastery which in turn is related to overcoming difficulties (185), expectations of improvement with treatment (101), and thus to treatment outcomes (101,180).

Another proposed mechanism derives from the perception of one's place in a supportive network, for example a sense of belonging or companionship, which is

more likely to be felt by those with greater self-reported levels of social support (180). These perceptions have been found to be associated with positive health outcomes of many sorts, and the lack of such companionship (or self-reported loneliness) is now a major focus of research and public health initiatives as it's role in negative health outcomes is being more widely recognised and more thoroughly researched (82,186–189). One further proposed mechanism through which social support is considered to act on health is by one's close social ties and social network acting as a buffer against stress (190), whether that be because members of the social network help solve stress related problems, stop one from directly facing the impact of certain stressful situations, or by mitigating the impact of stress, this buffering against stress might reduce the probability of depression occurring (180).

Social support, loneliness, and social isolation

The exact mechanism by which social support might affect prognosis is unclear. Whether any one or more of the above might effect the likelihood of engaging in treatment or of particular prognostic outcomes for adults presenting for treatment for depression, in primary care, is unclear too. However, there are now calls for a greater focus on social support and related social factors such as loneliness and social isolation (191) as potential indicators of prognosis (192). Loneliness is sometimes defined as the gap between desired social contacts (both the amount of them and perceived quality of them) and the social contacts one experiences (68) and is thought to encompass components divided into social loneliness (a lack of the desired level of one's social network) and emotional loneliness (the lack of desired intimate social relationships). In contrast social isolation is often defined as the objective rather than subjective rating of the quantitiy and mobilisation of one's social network (68).

As with social support there is a paucity of research assessing the links between loneliness or social isolation and prognosis for people with depression. Only one systematic review of those included in Chapter 1 reported on loneliness, and that finding was based on just a single primary study. That primary study was a cohort study of 285 dutch older adults diagnosed with MDD (193). The study authors reported that loneliness at baseline was associated with depressive symptoms at two-years post-baseline, and that the odds of remission for severely lonely

participants were approximately a quarter of the odds of remission for those reporting no loneliness (193). The generalisability of these findings for adults presenting to their GPs with depression is somewhat limited as just 15% were primary care patients, approximately 20% of respondents had missing data on either the lonelieness measure or the depressive symptom measures in the study, and 28% were deceased at follow-up.

A large retrospective cohort study has reported that participants who joined more social groups and so became less socially isolated over the course of the study were less likely to experience relapses of depression, less likely to develop depression over the course of the study, and more likely to recover from their depressive episodes (192). However, this study did not involve continuous means of measurement of depression so relapses may have been missed over the course of follow-up. Furthermore, the sample were largely non-treatment seeking and the majority had very low symptoms of depression at baseline (approximately 7% of the sample met the cut-off for probable depression at baseline), limiting the sample size available for the analysis of prognostic outcomes for the depressed participants throughout the study.

Patients seeking or receiving treatment reportedly consider social support, loneliness, and social isolation to be particularly important to their recovery from depressive episodes and in preventing relapses (82,186–189). The research evidence to support this notion has been strengthened in recent years (68,82,194). However, previous studies have not investigated prognosis both independent of treatment and independent of depressive symptom severity, and they have not investigated the association of social support with attrition. Therefore, the ability to identify independent prognostic associations which could inform the clinical management of depression for patients presenting for treatment, is somewhat limited from previous studies.

Aims

This study aimed to address sub-aims three and four of this thesis, and more specifically aimed to: 1) investigate whether social support is associated with prognosis for adults with depression, independent of treatment and independent of

depressive symptom severity and ‘disorder severity’; 2) to investigate which individual items from the Social Support Scale are associated with prognosis independent of treatment and depressive ‘disorder severity’; and 3) to investigate whether social support or individual items from the Social Support Scale of it are associated with attrition from treatment, independent of treatment, depressive symptom severity and ‘disorder severity’.

Methods & Materials

In order to meet the above aims an individual patient data meta-analyses using a subset of studies from Dep-GP that utilised a measure of social support is indicated. The Social Support Scale measure was used in six of the 12 Dep-GP studies (described in Chapter 2 and detailed in Table 2.2), this measure has been widely used in randomised controlled trials and largescale observational studies over the last 25-30 years e.g. (23,126). However, despite the widespread use of the measure (or more accurately, the first seven questions of the measure) and despite it being based on the Health and Lifestyles Survey measure of Social Support (137), there are no published validation studies of the eight-item scale used in the Dep-GP studies. There are insufficient data in Dep-GP to produce a full validation of the scale. However, before using data based on this measure to investigate the aims above, I will present an assessment of the reliability, internal consistency and discriminant validity of the measure currently in use, and consider any latent factors and the dimensionality of the scale, as these may inform the analyses to meet the above aims.

The methods for formation the Dep-GP IPD dataset and analysing data have been described in Chapters 2 and 3, and elsewhere (168). Below I give a brief outline of these methods and details of any differences specific to this study.

Identification and Selection of Studies:

Studies were included in the IPD if they were randomized controlled trials (RCTs) of adults seeking treatment for depression from a general practitioner, with unipolar depression confirmed via the revised clinical interview schedule (CIS-R) (112) at baseline. Studies in the present analyses also had to use the Social Support Scale (195) at baseline. Details of the measures including the Social Support Scale are in Table 2.2.

Six studies met these inclusion criteria (COBALT, GENPOD, IPCRESS, MIR, PANDA, and TREAD). See Table 4.1 for details.

Table 4.1. Description of the six Dep-GP IPD studies with Social Support Scale data.

Study	Sample Size	Inclusion criteria	Age	Gender	Baseline Depressive Symptom Severity	Baseline CIS-R Total Score	T0 Social Support Total Score	Remission at 3-4 months	Interventions	Outcome Measure at 3-4 months
			Mean (SD)	% Female	Mean(SD)	Mean(SD)	Mean(SD)			**Primary (additional)**
COBALT	N=469	Adults 18-75 with treatment resistant depression, scoring ≥14 BDI-II	49.6(11.7)	72%	BDI-II=31.8(10.7)	30.1(8.9)	20.0(3.8)	34%	CBT+TAU vs TAU	PHQ-9 (BDI-II)
GENPOD	N=601	Adults 18-74 with depressive episode	38.8(12.4)	68%	BDI-II=33.7(9.7)	30.8(8.0)	20.0(3.8)	41%	Citalopram vs Reboxetine	BDI-II (HADS)
IPCRESS	N=295	Adults scoring ≥14 BDI-II and GP confirmed diagnosis of depression	34.9(11.6)	68%	BDI-II=33.2(8.8)	29.6(8.7)	20.0(3.8)	34%	iCBT+TAU vs TAU + waiting list for iCBT	BDI-II
MIR	N=480	Adults ≥18 taking SSRIs or SNRIs at adequate dose for≥ 6 weeks, and scored ≥14 on BDI-II	50.7(13.2)	69%	BDI-II=31.1(9.9)	27.7(8.3)	20.5(4.1)	30%	Mirtazapine vs Placebo	BDI-II (PHQ-9)
PANDA	N=652	Adults presenting with low mood or depression to GP in last 2 years, free of ADM for 8 weeks up to baseline	39.7(15.0)	59%	BDI-II=23.9(10.3)	21.3(10.1)	20.6(3.8)	69%	Sertraline vs Placebo	PHQ-9 (BDI-II)
TREAD	N=361	Adults 18-69 who met diagnostic criteria for MDD and scored ≥14 on BDI-II	39.8(12.6)	66%	BDI-II=32.1(9.2)	28.1(7.8)	20.1(3.8)	35%	Physical Activity + TAU vs TAU	BDI-II

Abbreviations: ADM – Antidepressant medication; BDI-II – Beck Depression Inventory; HADS – Hospital Anxiety and Depression Scale; iCBT (internet based therapist delivered cognitive behavioural therapy); MDD – Major Depressive Disorder; TAU – treatment as usual.

Data Analysis Plan

Outcomes

The outcomes for this study were the same as those presented in Chapter 3 and described in Chapter 2 with two small differences: 1) as five of the six studies meeting inclusion criteria for this study used the BDI-II at the 3-4 month endpoint, the scores on that measure in those five studies were used as a further sensitivity analysis. 2) As the analyses in Chapter 3 showed that there were few differences in the prognostic factors found to be associated with outcomes when using both the standardized mean (z-score) and natural logarithm (log outcome) of depressive symptom scale scores at 3-4 months post-baseline, the log outcome was only used for the outcome at 3-4 months in the current study. If large differences were apparent between the social support items found to be associated with prognosis at 3-4 using the z-score and the log outcomes, then both would have been used for secondary and sensitivity analyses too. As is outlined in the Results section below, this was not the case, so only the z-score outcome was used for secondary and sensitivity outcomes of relevance (e.g. depressive symptom scale scores at 6-8 months post-baseline, BDI-II scores, and PROMIS-T scores).

Predictors under consideration

Potential baseline predictors of outcome were the total score and the eight individual items of the Social Support Scale.

Symptom Severity and 'Disorder Severity' factors under consideration

Previous studies have shown that social support is often rated lower by people experiencing higher levels of depression, and as outlined above: social support can impact on the severity of depressive symptoms (82,179,180), so understanding the association between social support and prognosis for depressed patients it is important to consider these independent from the severity of depression. In line with the aims and the analyses outlined in Chapter 3 the BDI-II score at baseline (as this depressive symptom scale measure was used in all studies at baseline) was adjusted for in models of the associations between the social support variables and outcomes independent of treatment and depressive symptom severity. In addition, for the models adjusting for depressive 'disorder severity' factors, those variables associated with prognosis independent of depressive symptom severity and of each

other (in Chapter 3) were also adjusted for. These were: the duration of anxiety, the duration of depression, comorbid panic disorder, and a history of treatment with antidepressant medications.

Adjusting for Covariates

As outlined in the preceding chapters, for the present analyses different covariates were considered in relation to each prognostic factor under investigation. Treatment allocation, age and self-reported gender at baseline were controlled for in all models. Based on analyses in Chapter 3 which showed marital status and employment status were associated with prognostic outcomes, and previous studies which have suggested strong associations between these variables and social support (68,82,179,180), marital status and employment status were also adjusted for in all models. Other covariates under consideration were the experience of stressful life-events in the six months prior to baseline, the highest level of educational attainment, financial wellbeing, and housing status. There is evidence from prior studies that social support is both particularly important when stressful life events or problems in socio-economic functioning are experienced, and that perceived levels of social support may be affected by the same socioeconomic markers and markers of stressful life events (68,82,180,194).

Assessing properties of the Social Support Scale

Before analysing the data to meet the above aims, I conducted analyses to explore the structure of the social support questionnaire. This involved an exploratory principal components analysis to identify any distinct underlying components within the scale that may inform later analyses, and analyses of the internal consistency, split-half reliability, discriminant validity and latent structure of the Social Support Scale measure, using an Item Response Theory (IRT) based analysis. IRT is particularly useful as an explanatory tool to assess the way in which respondents answer individual items and groups of items on a scale or questionnaire measure (196). It allows specification of the relationship between underlying or latent levels of the construct(s) measured in the given questionnaire, and respondent's answers to the individual items. In so doing, IRT analyses can separate item parameters and the characteristics of the sample of respondents from the manifest data, so that each can be understood and studied separately (196). IRT can be performed specifying a single underlying or latent trait, in which case we might call it unidimensional, or with

multiple latent traits, in which case we might call it multidimensional (MIRT). The model fit for unidimensional and multidimensional IRT models can be compared so that the model that provides the best fit to the data can be retained, and the best fitting model can help explain the latent structure of the questionnaire measure (196). The best fitting model can then be used in much the same way as a factor analytic model might be, modelling the probabilities of responses to individual items as a function of the single or multiple latent factors (depending on the dimensionality of the best fitting model) and can form the basis of classical tests of reliability such as internal consistency, split-half reliability and discriminant validity (196). For the present analyses an IRT analysis was conducted using the multidimensional IRT package 'mirt' (196), in R (160). This assumed social support is a latent factor with an unknown number of dimensions, initially fitting a model with the assumption of a single dimension and comparing this model to one with one more dimension, and doing this continually until adding dimensions did not improve the fit of the model. Model parameters were estimated using the marginal maximum likelihood method which is considered most useful when there are likely to be one or only a few latent factors identified in an exploratory IRT analysis (197).

Primary analyses

The methods for the primary analyses have been described in Chapter 2. In brief, four models were built for each social support item and for the total score, for each outcome at 3-4 months post-baseline (z-score of depressive symptom scales scores; the natural logarithm of depressive symptom scale scores; withdrawal/attrition; remission; and BDI-II score in five studies), and at 6-8 months post-baseline (z-score of depressive symptom scales scores). It is important to note that in order for estimates across the social support variables to be comparable, each item was scaled to a score between zero and one by dividing the variable by the maximum score available, so the total score was divided by 24 and each individual item was divided by three. The four models were:

1. The social support item/total score adjusted for treatment allocation, gender, age, employment status, and marital status.
2. As in 1 but with the addition of baseline depressive symptom severity.

3. As in 2 with the addition of all depressive 'disorder severity' factors that were significant or otherwise important in 2, and then removing factors that are no longer significant (at the 5% level).
4. As in 3 with the addition of covariates specific to the social support item (e.g. life events, financial wellbeing, housing status, or highest level of educational attainment).

Meta-analyses were conducted in line with the protocol for these analyses (168) and as outlined in Chapter 2 above. In brief, two-stage meta-analyses were conducted using DerSimonian and Laird random effects models. The degree of heterogeneity was assessed using prediction intervals and its impact assessed using the I^2 statistic (163).

Sensitivity analyses

As per the protocol, sensitivity analyses were planned to be conducted where heterogeneity might be considered problematic (e.g. with I^2 above 75%), removing the study contributing most to the heterogeneity or where any studies were rated as having 1) moderate or high risks of bias or 2) rated as offering a low quality of evidence for the effects investigated (see Risk of Bias section below). Further sensitivity analyses were conducted using the BDI-II score at 3-4 months as an outcome, excluding the one study without scores on this measure at the primary endpoint, and in addition, the analyses were run using the multidimensional IRT conversion of BDI-II scores and PHQ-9 scores at 3-4 months post-baseline to the PROMIS T-score.

Data handling and data management

Details of the pre-processing stages and handling of missing data including specifications for multiple imputation performed in each study can be found in the study protocol (168) and in Chapter 2.

Risk of Bias and Evidence quality

The risk of bias in each study was presented in Chapter 3 (Table 3.1), although the quality of evidence for each prognostic indicator using the Grading Recommendations, Assessment, Development and Evaluations (GRADE)

framework (165) was conducted again in relation to the prognostic indicators assessed in the present study as these were different to those assessed in Chapter 3 and might therefore have led to different ratings.

Results

Characteristics of the included studies

In total, six RCTs were identified as meeting inclusion criteria and were able to provide IPD. A description of each included study can be found in Table 4.1, descriptive statistics are presented in Table 4.2. Further details on the distributions of baseline and outcome variables for each of the six studies can be found in Appendix 2 (168).

Table4.2. Baseline Characteristics of the Study Sample.

Self-reported Baseline Characteristics	Factor	N(%), or Mean(SD), or Median (IQR)
Social Support	Total score	Median (IQR) =21 (16 to 24)
	Accepted	Mean(SD) =2.56(0.60)
	Cared about	Mean(SD) =2.75(0.48)
	Supported or Encouraged	Mean(SD) =2.53(0.61)
	Made to feel happy	Mean(SD) =2.42(0.64)
	Made to feel important	Mean(SD) =2.46(0.66)
	Made to feel loved	Mean(SD) =2.60(0.58)
	Can rely on others	Mean(SD) =2.59(0.61)
	Can talk to others	Mean(SD) =2.34(0.71)
Age		Mean(SD) =42.52(14.12)
Gender	Female	1900(66.53)
	Male	956(33.47)
	Other	0
Ethnicity	White	2698(94.43)
	Non-White	159(5.57)
Employment status	Employed	1639(57.39)
	Not seeking employment	685(23.98)
	Unemployed	532(18.63)
Marital Status	Married/cohabiting	1379(48.25)
	Single	911(31.88)
	No longer married	568(19.87)
Educational Attainment	Degree or higher	609(27.05)
	A-level or Diplomas	669(29.72)
	GCSE	589(26.17)
	None or Other	384(17.06)
Financial status	Doing OK	1184(41.47)
	Just about getting by	914(32.01)
	Struggling	757(26.51)
Housing status	Home owner	1096(45.88)
	Tenant	948(39.68)
	Other	345(14.44)
Long-term conditions	No	1873(78.40)
	Yes	516(21.60)
Number of recent life events		Mean(SD) =1.35(1.24)
Past Antidepressant use	No	908(31.77)
	Yes	1950(68.23)
CIS-R durations	Depression	3.42(1.37)
	Average Anxiety Duration	2.14(0.99)
Comorbid panic disorder	No	2623(91.78)
	Yes	235(8.22)
Baseline BDI-II score		Mean(SD) =30.44(10.53)
3-4 month BDI-II score		Mean(SD) =16.07(11.99)
6-8 month BDI-II score		Mean(SD) =18.64(13.44)
Remission at 3-4 months	No	1363(57.56)
	Yes	1005(42.44)
Attrition at 3-4 months	No	2382(83.34)
	Yes	476(16.66)

Quality assessments and Risk of Bias

The risk of bias in each study was assessed using QUIPS, see Table 4.3, and the quality of evidence for each prognostic indicator was assessed using the GRADE framework (165), see Table 4.3. Two reviewers (JB and a collaborator) independently assessed each study and judged the quality of evidence for each prognostic factor (interrater reliability: (Cohen's Kappa k=0.96 for QUIPS and k=1.00 for GRADE). Disagreements were resolved in consensus meetings with two further reviewers (SP and GL, the PhD supervisors).

Table4.3. Risk of Bias and Quality Ratings for each of the six included studies.

	QUIPS Risk of Bias Ratings						GRADE Quality Assessment
Study	**Study Participation**	**Study Attrition**	**Prognostic Factor Measurement**	**Outcome Measurement**	**Study Confounding**	**Statistical Analysis and Reporting**	**As a Prognostic Indicator**
COBALT	Low	Low	Low	Moderate	Low	Low	High
GENPOD	Low	Low	Low	Low	Low	Low	High
IPCRESS	Low	High	Low	Low	Low	Low	High
MIR	Low	Moderate	Low	Low	Low	Low	High
PANDA	Low	Low	Low	Low	Low	Low	High
TREAD	Low	Low	Low	Low	Low	Low	High
Overall	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**High**

Descriptive Statistics

Approximately 67% of the participants in the dataset were female, with a mean (standard deviation) age of 42.5(14.1). Over 94% of the sample were from white ethnic backgrounds. Just over two-thirds of the sample had a history of depression, and one third had been depressed for at least one year at the point of their baseline assessments. The mean and standard deviation of BDI-II scores at baseline was 30.4(10.5) so most participants scored in the severely depressed range. The median (and interquartile range) of scores on the social support scale was 21(16 to 24) so the largest group of participants had a moderate lack of social support (a score between 19 and 23), with approximately 30% reporting a severe lack of social support (a score below 19) and approximately 29% reporting no lack of social support (a score of 24/the maximum score), see Table 4.2.

Across the six included studies, there were significant differences between those with lower and higher levels of social support at baseline on nearly all demographic

and symptom related variables. Only gender, ethnicity, educational attainment, whether or not participants had a self-reported long-term health condition, and a history of treatment for depression did not significantly differ between those with higher compared to lower baseline social support. Despite these differences the correlation between the total social support score at baseline and baseline depressive symptom severity was weak (r=-0.29) and the correlation between the total social support score and the z-score of depressive symptoms at 3-4 months post baseline was very weak (r=-0.18) by conventional standards (198).

Properties of the Social Support Scale

In order to consider the utility in assessing the associations between individual items of social support and prognostic outcomes an exploratory principal components analysis was conducted. All eight items were highly correlated with one and other r=(0.60 to 0.82), and a single component solution explained approximately 76% of the variance attributable to the social support scale measure with all individual items highly correlated with the principal component (Pearson's ρ = 0.79 to 0.87), see Table 4.4.

In the IRT analysis to assess the reliability of the social support scale the measure displayed excellent model fit with a single dimension ($M2(12) = 149.6$, $p < .0001$, root mean squared error of approximation (RMSEA)= 0.06, comparative fit index (CFI)= 0.99 and Tucker-Lewis Index (TLI)= 0.99). Individual item loadings on the unidimensional latent variable ranged between 0.78 and 0.92. There was good internal consistency (Empirical reliability = 0.84; Cronbach's α = 0.91; Guttman's λ_6 = 0.91), split-half reliability (Revelle's β = 0.87), and discriminative validity (corrected item-total correlations: 0.68-0.76). The ability to use individual items of the scale to discriminate between respondents with different levels of social support (taken from the slope of the Item Characteristic Curve (also known as the discrimination value) of a series of mixed effect linear regression models fitted as part of the multidimensional IRT package) was moderately strong, with probabilities of such discrimination ranging between 0.24 and 0.42 for each item.

Table 4.4. Correlation Matrix of Social Support Items and Principal Component.

	Correlation between items (Pearson's ρ)								
	Accepted	**Cared about**	**Made to feel Happy**	**Made to feel Important**	**Made to feel Loved**	**Can rely on others**	**Supported and Encouraged**	**Can talk to others**	**Principal Component**
Accepted	1								
Cared about	0.74	1							
Made to feel Happy	0.64	0.78	1						
Made to feel Important	0.69	0.82	0.79	1					
Made to feel Loved	0.69	0.77	0.69	0.76	1				
Can rely on others	0.74	0.75	0.67	0.71	0.76	1			
Supported or Encouraged	0.74	0.78	0.76	0.76	0.73	0.76	1		
Can talk to others	0.66	0.70	0.71	0.70	0.60	0.69	0.75	1	
Principal Component	0.80	0.87	0.83	0.85	0.81	0.82	0.86	0.79	1

The association between Social Support and Prognosis

The total score on the social support scale was associated with the severity of depressive symptoms at 3-4 months post-baseline, independent of treatment and independent of age, gender, marital status and employment status (the *a priori* named covariates). The difference in the z-score at 3-4 months per unit increase in social support at baseline when scaled to between 0-1 was: -0.96(95%CI: -1.26 to -0.67), and using the natural logarithm outcome depressive symptom scores were on average 51.19%(95%CI: 37.86 to 61.66) lower at 3-4 months post-baseline, see Table 4.5. Once additionally controlling for depressive symptom severity the magnitude of effect was reduced (-0.26(-0.53 to -0.04), and 22.83%(7.39 to 35.69); the magnitude of effect was very marginally affected by additionally adjusting for depressive 'disorder severity' factors (-0.26(-0.51 to -0.01), and 22.83%(7.39 to 35.69). No additional covariates were retained in the models after adjusting for the above factors. Both ways of capturing the endpoint depressive symptom scale scores (using the z-score across all measures, and using the natural logarithm of the scores across all measures) gave very similar results in terms of the direction and magnitudes of associations between the social support variables and prognosis. Therefore, it was decided that models giving proportional outcomes would not be used for the secondary or sensitivity outcomes/analyses.

Table 4.5. Outcomes at 3-4 months (difference in z-score of depressive symptoms and % difference in depressive symptoms) per unit increase in baseline Social Support indicator.

	Difference or percentage difference in outcomes at 3-4 months post-baseline per unit increase in baseline Social Support Indicator											
	Effect independent of treatment				Effect independent of symptom severity and treatment				Effect independent of 'disorder severity' and treatment			
Social Support Indicator	**Z-score of depressive symptoms ^**		**% difference in depressive symptoms^**		**Z-score of depressive symptoms ***		**% difference in depressive symptoms***		**Z-score of depressive symptoms ‡**		**% difference in depressive symptoms‡**	
	Mean difference (95%CI)	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**	**%(95%CI)**	**I^2**
Total Score	-0.96(-1.26 to -0.67)	25	51.19(37.86 to 61.66)	35	-0.29(-0.53 to -0.04)	0	22.83(7.39 to 35.69)	0	-0.26(-0.51 to -0.01)	0	22.83(7.39 to 35.69)	0
Accepted	-0.77(-1.03 to -0.51)	38	43.45(28.76 to 55.12)	57	-0.29(-0.48 to -0.09)	0	20.76(7.80 to 31.91)	9	-0.25(-0.45 to -0.06)	0	20.76(7.80 to 31.91)	9
Cared about	-0.76(-1.01 to -0.50)	0	41.51(30.58 to 50.72)	0	-0.25(-0.49 to -0.01)	0	20.12(5.41 to 32.54)	0	-0.24(-0.48 to -0.01)	0	20.12(5.41 to 32.54)	0
Made to feel happy	-0.50(-0.68 to -0.31)	0	33.29(23.65 to 41.71)	0	-0.09(-0.27 to 0.08)	0	15.12(3.28 to 25.51)	0	-0.09(-0.27 to 0.08)	0	15.12(3.28 to 25.51)	0
Made to feel important	-0.45(-0.64 to -0.27)	0	29.33(19.28 to 38.12)	0	-0.03(-0.21 to 0.14)	0	6.89(-6.07 to 18.27)	0	-0.02(-0.20 to 0.16)	0	6.89(-6.07 to 18.27)	0
Made to feel loved	-0.65(-0.86 to -0.44)	0	34.98(22.12 to 45.72)	26	-0.19(-0.39 to 0.01)	0	11.70(-5.75 to 26.27)	28	-0.17(-0.37 to 0.03)	0	11.70(-5.75 to 26.27)	28
Can rely on others	-0.54(-0.78 to -0.31)	30	32.57(18.62 to 44.14)	39	-0.19(-0.38 to 0.01)	11	12.17(-0.59 to 23.31)	0	-0.16(-0.35 to 0.02)	0	12.17(-0.59 to 23.31)	0
Supported or Encouraged	-0.65(-0.84 to -0.45)	0	37.86(28.27 to 46.16)	0	-0.24(-0.42 to -0.05)	0	18.70(6.35 to 29.41)	0	-0.21(-0.40 to -0.02)	0	18.70(6.35 to 29.41)	0
Can talk to others	-0.49(-0.72 to -0.26)	50	31.36(18.33 to 42.31)	45	-0.15(-0.32 to 0.01)	8	13.08(2.12 to 22.81)	0	-0.14(-0.30 to 0.02)	0	13.08(2.12 to 22.81)	0

^adjusted for allocated treatment, gender, age, marital status and employment status; * additionally adjusted for depressive symptom severity; ‡ additionally adjusted for average anxiety duration, depression duration, panic disorder, and history of treatment with antidepressants

The findings with the secondary and sensitivity outcomes at 3-4 months were similar to those with the primary outcome: for every unit increase in the total score on the Social Support Scale at baseline there was an increase in the odds of reaching remission at the primary endpoint (OR= 2.33(95%CI: 1.25 to 4.37)). In sensitivity analyses using the five studies that had BDI-II scores at the primary endpoint, for each unit increase in social support there was a -3.26(-6.52 to 0.00) point difference in the average BDI-II score at 3-4 months (see Table 4.6). Additional sensitivity analyses using the PROMIS T-score as the outcome are shown in Appendix 5. There were however, no significant associations between the total Social Support Scale score and outcome at 6-8 months post-baseline. There was little heterogeneity in the effects so no further sensitivity analyses were deemed necessary.

Table 4.6. Associations of Social Support with secondary prognostic outcomes, adjusted for treatment, 'disorder severity' and covariates.

	Secondary Outcomes				**Sensitivity Analysis**	
Social Support Domain	**Remission at 3-4 months**		**z-score at 6-8 months**		**BDI-II score at 3-4 months†**	
	OR(95%CI)	**I^2**	**Mean difference (95%CI)**	**I^2**	**Mean difference (95%CI)**	**I^2**
Total Score	2.33(1.25 to 4.37)	0	-0.21(-0.57 to 0.15)	0	-3.26(-6.52 to 0.00)	0
Accepted	1.69(1.02 to 2.78)	0	-0.21(-0.50 to 0.07)	0	-3.37(-5.99 to -0.75)	0
Cared about	2.24(1.20 to 4.15)	0	-0.17(-0.51 to 0.18)	0	-2.80(-5.99 to 0.38)	0
Made to feel happy	1.52(0.97 to 2.37)	0	-0.10(-0.35 to 0.15)	0	-0.94(-3.23 to 1.35)	0
Made to feel important	1.44(0.92 to 2.25)	0	0.02(-0.24 to 0.27)	0	-0.30(-2.60 to 1.99)	0
Made to feel loved	1.90(1.15 to 3.13)	0	-0.23(-0.59 to 0.13)	30	-2.98(-5.63 to -0.34)	0
Can rely on others	1.70(1.05 to 2.74)	0	-0.16(-0.44 to 0.12)	0	-2.60(-5.49 to 0.30)	25
Supported or Encouraged	1.87(1.17 to 2.97)	0	-0.17(-0.43 to 0.09)	0	-2.20(-4.66 to 0.26)	0
Can talk to others	1.35(0.91 to 2.01)	0	-0.09(-0.31 to 0.13)	0	-1.39(-3.47 to 0.70)	0

All models adjusted for random allocation in each study, depressive symptom severity, average anxiety duration, depression duration, panic disorder, and a history of antidepressant treatment, gender, age, marital status, and employment status †Only available for 5 studies, excludes COBALT

The associations between individual Social Support items and Prognosis

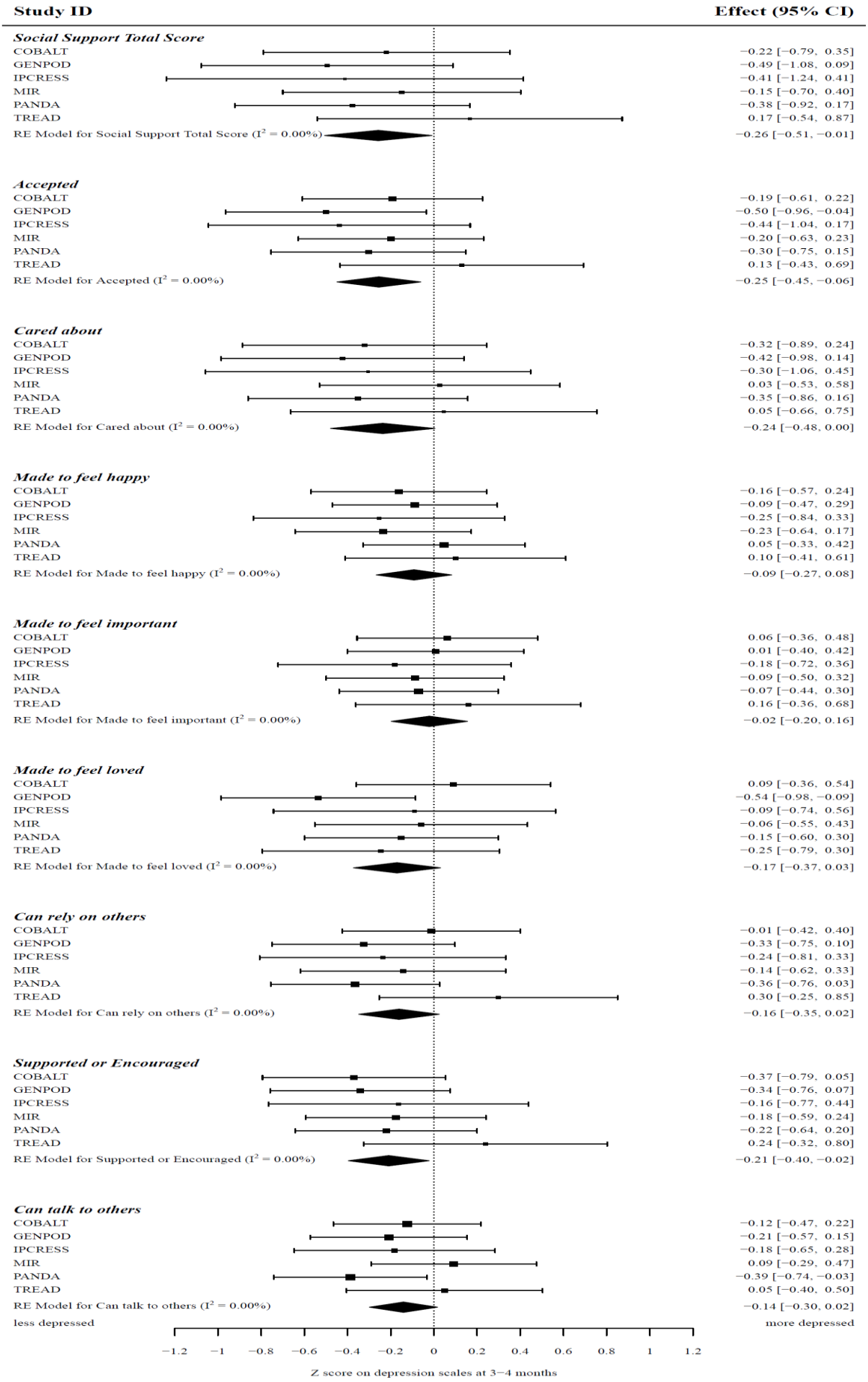
As described above, the psychometric properties of the Social Support Scale are such that all individual items adequately load onto a single latent factor, and that each item is able to be used to identify participants with different levels of social support, however there was variability in the factor loadings and item characteristic coefficients for each item. It might therefore be the case that certain items of the scale are able to be used to discriminate between participants with different levels of social support as well as the whole scale is, and certain items might therefore have greater utility if there are differential associations with prognosis between the items too.

All eight of the Social Support Scale items were significantly associated with prognosis independent of treatment and covariates, see Table 4.5. However, the magnitudes of association were different between the individual scale items, and only three items were significantly associated with the outcome after additionally adjusting for depressive symptom severity, 'disorder severity' factors, and covariates. The three items most strongly associated with the prognostic outcomes at 3-4 months post-baseline were: 1) whether or not one feels accepted for who one is, by family and friends; 2) whether or not one feels cared about by family and friends; and 3) whether or not one feels supported or encouraged by family and friends, see Tables 22 and 23, and Figure 4.1. For each unit increase in feeling accepted by family and friends there was an approximate 43% difference in depressive symptom scores at 3-4 months post-baseline, independent of treatment and all socio-demographic covariates (percentage difference =43.45%(95%CI: 28.76 to 55.12)). The difference was approximately 21% after additionally adjusting for depressive symptom severity and the depressive 'disorder severity' factors (the average duration of anxiety problems, the duration of depression, comorbid panic disorder and a history of treatment with antidepressants) (percentage difference =20.76%(95%CI: 7.80 to 31.91)). Similarly, the difference in the mean symptom score at 3-4 months per one-point increase in the "cared about" item was approximately 42% independent of treatment and socio-demographic covariates (percentage difference =41.51%(95%CI: 30.58 to 50.72)), and approximately 20% when also adjusting for depressive symptom severity and 'disorder severity' factors

(percentage difference =20.12%(95%CI: 5.41 to 32.54)). The difference in depressive symptoms at 3-4 months post-baseline was approximately 38% independent of treatment and socio-demographic covariates (percentage difference =37.86%(95%CI: 28.27 to 46.16)), and 19% after additionally adjusting for all the severity factors (percentage difference =18.70%(95%CI: 6.35 to 29.41)), see Figure 4.1 for details of heterogeneity.

In regards secondary outcomes, after adjusting for treatment, age, gender, marital status, employment status, depressive symptom severity and the four depressive 'disorder severity' factors, there was no clear evidence of an association between three social support items and remission at 3-4 months (being made to feel happy, being made to feel important, and feeling able to talk to family or friends whenever needed). These were also the items least strongly associated with the z-score of depressive symptoms at 6-8 months post-baseline, although the was no clear evidence of associations between any of the individual items and prognosis at 6-8 months post-baseline, see Table 4.6. In sensitivity analyses, the item relating to feeling accepted was most strongly associated with the score on the BDI-II at 3-4 months in the five studies that had these data (mean difference = -3.37(95%CI: -5.99 to -0.75)), this was followed by the item relating to being made to feel loved (mean difference = -2.98(95%CI: -5.63 to -0.34)). There was no clear evidence that any of the other items were significantly associated with this outcome. Similar to the primary outcome, in the sensitivity analysis using the PROMIS T-score outcome the items most strongly associated with prognosis at 3-4 months post-baseline were feeling accepted, being cared about, and being supported or encouraged. Further, being made to feel important was not significantly associated with the outcome, see Appendix 5.

Figure 3.1. Forest plot of associations between social support items and prognosis at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity'.



The association between Social Support and Attrition

The total score on the social support scale was associated with attrition independent of treatment such that higher scores were associated with considerably lower odds of attrition (OR(95%CI)= 0.35(0.13 to 0.93)), see Table 4.7. The more cared about patients felt the less likely they were to dropout or withdraw (OR(95%CI) = 0.32(0.17 to 0.61); percentage difference in attrition = 56.48%(95%CI: 30.23 to 72.86)). Likewise with the more they considered themselves to be made to feel happy by friends and family, the more supported or encouraged they felt, or the more they felt able to rely on others, see Table 4.7. After adjusting for depressive symptom severity all of the effects were reduced in magnitude to the point that most social support items and the total score were not significantly associated with attrition. The magnitude of effects was reduced very slightly again when additionally adjusting for 'disorder severity' factors. However, there was evidence that feeling cared about (OR(95%CI) = 0.37(0.17 to 0.81)), and feeling supported or encouraged (OR(95%CI) =0.54(0.32 to 0.92)) were associated with attrition after adjusting for the above factors, see Table 4.7. In terms of the sizes of effects, the item with the largest magnitude of association with attrition was the that related to feeling cared about, with an approximate 52% decrease in the probability of attrition by the 3-4 month end-point (51.68%(95%CI: 16.42 to 72.06)), independent of treatment, socio-demographic covariates and all severity factors adjusted for.

Table 4.7. Association of social support with attrition at 3-4 months post-baseline.

	Difference or percentage difference in outcomes at 3-4 months post-baseline per unit increase in baseline Social Support Indicator											
	Effect independent of treatment				Effect independent of symptom severity and treatment				Effect independent of 'disorder severity' and treatment			
Social Support Domain	**OR(95%CI)^**	**I^2**	**%(95%CI) difference in attrition^**	**I^2**	**OR(95%CI)***	**I^2**	**%(95%CI) difference in attrition***	**I^2**	**OR(95%CI)‡**	**I^2**	**%(95%CI) difference in attrition‡**	**I^2**
Total Score	0.35(0.13 to 0.93)	52	54.14(4.73 to 77.93)	53	0.43(0.15 to 1.22)	54	46.9(-15.05 to 75.49)	53	0.42(0.14 to 1.28)	58	47.2(-20.23 to 76.81)	56
Accepted	0.48(0.24 to 0.95)	44	43.47(4.28 to 66.61)	45	0.55(0.26 to 1.16)	49	37.15(-10.2 to 64.15)	47	0.54(0.24 to 1.21)	53	37.16(-14.51 to 65.51)	52
Cared about	0.32(0.17 to 0.61)	11	56.48(30.23 to 72.86)	17	0.37(0.18 to 0.75)	18	52.28(22.18 to 70.74)	16	0.37(0.17 to 0.81)	29	51.68(16.42 to 72.06)	27
Made to feel happy	0.49(0.27 to 0.89)	33	41.99(7.51 to 63.62)	36	0.55(0.30 to 1.02)	33	36.51(-0.99 to 60.08)	31	0.55(0.27 to 1.10)	44	36.13(-7.56 to 62.07)	43
Made to feel important	0.75(0.38 to 1.49)	47	19.37(-34.56 to 51.69)	47	0.89(0.43 to 1.85)	50	7.87(-59.1 to 46.64)	49	0.90(0.42 to 1.93)	52	6.95(-64.52 to 47.37)	51
Made to feel loved	0.66(0.35 to 1.22)	20	26.59(-18.78 to 54.63)	23	0.78(0.41 to 1.48)	20	16.62(-34.59 to 48.34)	19	0.79(0.37 to 1.69)	40	15.91(-50.21 to 52.92)	40
Can rely on others	0.40(0.19 to 0.87)	55	49.83(10.30 to 71.94)	57	0.44(0.20 to 0.97)	53	45.86(3.55 to 69.61)	53	0.44(0.19 to 1.01)	57	45.58(-0.06 to 70.41)	57
Supported or Encouraged	0.48(0.29 to 0.79)	0	42.24(16.61 to 59.99)	0	0.55(0.33 to 0.92)	0	36.52(7.21 to 56.56)	0	0.54(0.32 to 0.92)	0	36.57(6.53 to 56.96)	0
Can talk to others	0.70(0.45 to 1.08)	0	24.05(-5.30 to 45.22)	0	0.80(0.50 to 1.28)	6	16.14(-17.68 to 40.24)	2	0.76(0.48 to 1.20)	0	19.01(-14.08 to 42.51)	0

^adjusted for allocated treatment, gender, age, marital status and employment status; * additionally adjusted for depressive symptom severity; ‡ additionally adjusted for average anxiety duration, depression duration, panic disorder, and history of treatment with antidepressants

Discussion

Social support was associated with prognosis and with attrition from treatment. This was the case independent of treatment and independent of both depressive symptom severity and depressive 'disorder severity'. There was no clear evidence that the individual items of the social support scale were differentially associated with prognosis or attrition. However, three of the items had larger magnitudes of association than the others, and were consistently and significantly associated with the prognostic outcomes, this was not the case for the other five items. These three items were feeling accepted by family and friends for who one is, feeling cared about by family and friends, and feeling supported or encouraged by friends and family. The latter two items, and in particular the item related to feeling cared about, were also most strongly associated with attrition at 3-4 months post-baseline. Neither the total score nor any of the Social Support Scale items were significantly associated with prognosis at six-to-eight months post-baseline after adjusting for depressive symptom severity, in the four studies that had data at that time point. There was very little heterogeneity in these associations across the six included studies supporting the robustness of these findings.

Findings in Context

Recent study-level systematic reviews have suggested that social support is associated with the outcomes from antidepressant treatment (84) and is associated with the "natural course" of depression among general population non-patient samples (68,73,82), although there was limited data to support these findings. The present study has found the association to be present independent of a range of treatments offered to adults seeking treatment for depression from their GPs and also shown this to be the case independent of the severity of depression measured in a variety of ways. This latter point is particularly important as adjusting for depressive severity had a large impact on the magnitude of the associations between social support variables and outcomes. In addition, these associations were found independent of other factors considered to be potential confounders in the association between social support and prognosis including marital status and employment status (68,82,179,180). Other potential confounders were explored as covariates but were not associated with baseline social support here, including stressful life events and socio-economic status (68,82,180,194).

In this study, all items of the Social Support Scale were found to be highly correlated with a single principal component, and the IRT analysis found a single latent factor with each item adequately able to be used to discriminate those with different levels of social support. Prior studies have suggested a multidimensional nature to social support (179–181), proposing an important distinction between three sub-domains of social support: emotional support; informational support; and instrumental support, although there is somewhat limited evidence for the existence or utility of such sub-domains (179,182). In this study there was no clear evidence of such sub-domains within the scale used here. Further, the items most strongly associated with the prognostic outcomes could perhaps be considered to be part of an apparent emotional support sub-domain (being accepted, feeling cared about, and feeling supported or encouraged by family or friends) but other items that were considerably less strongly associated with the prognostic outcomes could also be considered to be part of the emotional sub-domain (being made to feel happy, and being made to feel important by friends or family). Two items that might be considered to be part of the instrumental sub-domain (being able to talk to others and being able to rely on others) were found to be less strongly associated with the outcomes than some items but more strongly than others, and again, confidence intervals for all of the items overlapped. None of the social support scale items could be considered to be part of the informational support sub-domain.

Strengths and Limitations

This appears to be the first study to use a large individual patient dataset, formed from a number of RCTs, to consider the associations of social support with attrition and prognosis independent of a range of treatments (ranging from antidepressant medications, to cognitive behaviour therapy, and structured exercise). Studies were selected on the basis of recruiting participants in primary care settings only, and studies had to use the same measures to assess baseline symptoms, ‘disorder severity’ factors, social support and a number of potentially important covariates. This means that findings may be generalizable to the largest proportion of adults seeking treatment for depression in the UK (23) and we can have confidence in the ability to use the same measures in clinical practice to inform patients and clinicians regarding the outcomes assessed here.

A number of authors have recently proposed that social support, loneliness, and social isolation are all particularly important to patients and to understanding their abilities to engage with and benefit from treatments for mental health problems (82,186–189). The Social Support Scale used in the trials that make up Dep-GP was for the most part limited to questions about the function of an individual's social support or satisfaction with perceived levels of support, rather than the structure of the individual's social support network (179). As such, loneliness and social isolation could not be studied here.

Social support is considered by some to be causally associated with mental ill health, depression included (180). The evidence for a causal association with prognosis or with attrition independent of treatment is very limited. Although it was not the focus of this study, there might have been the potential to consider the associations here as causal. Were this the case, it is unlikely that the associations reported as significant here were found by chance. If adjustments were made for multiple testing a number of associations would not have been significant, but such adjustments were not considered necessary, following Rothman (199). Further, while reverse causality has been considered a problem in prior studies reporting associations between social support and prognostic outcomes for people with depression (82), this was not the case with the present study where social support was measured prior to randomisation to treatment and measurement of the prognostic outcomes.

The biggest limit to considering any potential causal relationship between social support and prognosis or attrition independent of treatment in the current study is that confounding could not adequately be controlled for here. As discussed in Chapter 3 causal relationships were not the focus of this study, so confounding is not particularly relevant, but in order to better consider the mechanisms by which social support affects prognosis or attrition, investigating such causal relationships may be essential. Adjustments were made for a number of baseline covariates that may be confounders in any causal relationship between social support and outcomes such as employment status, marital status, the severity of depressive symptoms, chronicity of depression and of anxiety comorbidities, and a history of treatment for depression, adding robustness to the findings. However, as the data were combined

from different randomised studies the randomisation, and benefits of it to control for both known and unknown confounding factors, was lost.

There were also limitations in regards to bias which would affect any causal interpretations of the findings here. The social support scale used in this study had been used in a number of large-scale studies over the last 25-30 years without having been validated in the current eight-item format. However, it was found to be sufficiently reliable, and had very good internal consistency and convergent validity, suggesting it is an appropriate measure of social support, limiting measurement error due to the instrument. However, the data used here were self-reported and this may have led to other sources of measurement error; we might expect those with greater baseline depression severity to be least likely to rate their own social support to be high, adding a systematic (or unidirectional) bias into the results. However, adjustments were made for all depression severity factors associated with the outcomes and social support variables, not just the severity of depressive symptoms, minimising the potential for such bias here.

Using a standardised outcome is a method that has been criticised previously but the results using the z-score outcome were similar to those for natural logarithm outcome, the BDI-II scores in sensitivity analyses, and to the secondary outcomes (remission), suggesting the use of the standardized outcome metric did not unduly affect the results.

Finally, the total scores on the Social Support Scale reported by participants were generally quite high; just under half of the sample scored 21 or above, and approximately 29% of the sample had the maximum score of 24 on the Social Support Scale, suggesting a potential selection bias. That said, the authors of the original seven-item version of the measure considered scores under the maximum to be indicative of potential issues in the level of social support that respondents might have, so expect a highly skewed pattern of responses to the questions of the scale (137,200).

Implications

Large differences in social support appear to be associated with considerable differences in prognosis and attrition, so knowledge of a patient's levels of perceived social support prior to commencing treatment could be informative for the management of their depression.

Given the psychometric properties of the Social Support Scale it is reasonable to suggest that any of the eight scale items could be used to capture aspects of social support. As feeling 'accepted by family and friends', feeling 'supported or encouraged by family or friends', and feeling 'cared about by family or friends', were most strongly and consistently associated with the prognostic outcomes at 3-4 months post-baseline, and the latter was also the item most strongly associated with attrition, it might be the case that any one of these three items may usefully be added to assessments of adults presenting to their GPs with depression, where use of the whole scale is not practical.

Future studies might investigate the relationship between a sense of belonging or companionship and prognosis independent of treatment, as this is related to the perception of one's place in a social network, and has been posited as a potential mechanism by which social support might affect prognosis (180). Being made to feel happy or important by others might be more closely related to mechanisms involving self-esteem and mastery (180,184) than some of the other social support items. Low self-esteem might be considered a symptom of depression for some, and when adjusting for baseline depressive symptom severity the associations between the above two Social Support Scale items and prognosis were not significant. Although tests of multi-collinearity did not suggest any such problems when adjusting for depressive symptom severity, when investigating this potential mechanism it might be more informative to consider adjustments for individual symptoms of depression at baseline, excluding those related to self-esteem.

Nearly all social support items were associated with attrition independent of treatment, this is in keeping with patients' suggestion that social support is particularly important to their engagement with treatment (188,189). We might hypothesise that those with greater social support may be more likely to receive encouragement to stay in treatment even when doubting the effectiveness of it for

themselves (201), they may also have greater motivation or incentive to keep taking or engaging with treatment due to their perceived social roles (179,202), and the encouragement they receive may act as a buffer against stressors which might otherwise lead them to end treatment (180).

Despite the confirmation that the Social Support Scale used in the studies that form this IPD has adequate psychometric properties to reliably and consistently measure social support, the measure has yet to be validated formally in the eight-item format, and this would be a valuable contribution should it become more widely used in research or clinical practice. Other measures, particularly those assessing loneliness and social isolation could be important additions to assessments in future studies and could prove informative as potential targets for interventions whether in or outside of the therapy room (82,186–188,191). Future research should consider the unique contribution of loneliness and social isolation to determining prognosis for adults with depression and any interactions between them and social support, as well as assessing their impact on treatment engagement or attrition.

Conclusions

In conclusion this study has shown that social support is associated with prognosis independent of treatment for adults presenting to their GPs with depression. This was also the case for attrition from treatment independent of whatever treatment was given. The associations were weakened by adjusting for the severity of depressive symptoms and related ‘disorder severity’ factors, but nonetheless the associations were significant and may be clinically important. In addition, this study has shown that single items measuring different aspects of social support are all strongly associated with the wider construct of social support overall, and they each might be useful in determining prognosis or the likelihood of attrition from treatment. Social support has previously been shown to be considered of importance by patients themselves and is considered modifiable. So, adding a measure of social support to assessments of depressed adults in primary care may be informative for patients and clinicians, aiding in the clinical management of depression.

Chapter 5. The associations between stressful life events, socio-demographics, long-term health condition status, and prognosis.

Overview

In Chapter 4 I reported an analysis of the associations between social support and prognosis independent of treatment for adults treated for depression in primary care. These associations were also investigated independent of depressive symptoms and independent of the depressive 'disorder severity' factors found to be independently associated with prognosis in Chapter 3. There are a number of other potentially important factors which have been suggested to be associated with prognosis but have not been studied in the analyses described thus far. So, in this chapter I will report on a series of analyses investigating the associations of stressful life events, socio-demographic factors, and long-term health condition status with prognosis and attrition, independent of treatment and of depressive symptoms and 'disorder severity' factors, finalising the work to address sub-aims 3 and 4 of this thesis.

Abstract

Background

Stressful life events are associated with the onset of depression and have been found to moderate differential responses to antidepressants and psychological therapy. A number of demographic factors and long-term health conditions have also been found to be associated with outcomes from one type of treatment, but none of these associations have been investigated independent of treatment, or in large clinical samples. This study aimed to assess the associations of these factors with attrition from treatment and with prognosis independent of treatment, and independent of depressive symptom severity and 'disorder severity'.

Methods

Data were collated from the individual participants of six RCTs (n=2858) of adults seeking treatment for depression in primary care. All completed the same baseline assessments of stressful life events, socio-demographics (age, gender, ethnicity, marital status, employment status, financial wellbeing, housing tenancy status, and the highest level of educational attainment), and long-term physical health condition status, depressive symptom severity and depressive 'disorder severity' factors. Data were analysed with two-stage random effects meta-analyses.

Results

Being the victim of a violent crime and having problem debts were associated with prognosis independent of treatment and of depressive severity factors. Most of the socio-demographic factors and long-term health condition status were associated with prognosis independent of all factors adjusted for. But, in general age and gender were not. There were mixed findings regarding the highest level of educational attainment which was marginally associated with prognosis with one outcome (the z-score of depressive symptom scale scores at 3-4 months post-baseline), but not when using other outcomes. Similar associations were found with attrition.

Conclusions

Important information can be gained about prognosis independent of treatment for adults with depression when considering life events, socio-demographics, and long-term health condition status. The factors investigated are measureable with self-report questionnaires so may be captured in routine clinical assessments. Future studies might investigate subgroups of depressed patients based on combinations of

factors assessed here, and also assess the ability to use the prognostic factors identified here to prospectively predict prognosis for new patients presenting to GPs with depression.

Introduction

In Chapter 1 I gave a brief outline of a number of systematic reviews that reported associations between life events, socio-demographics, and long-term health conditions, with prognosis for adults with depression. Below I give further details of these reviews and of other studies that suggest investigating such factors independent of treatment may lead to a greater understanding of prognosis for depressed patients.

Stressful life events and Depression

It is well established that first episodes of depression are commonly preceded by the experience of stressful life events (203,204). Life events that can be considered to be severe and acute (such as losing one's job, being the victim of an attack, or divorce) are strongly associated with risk of a first depressive episode (205). If the same types of events are reported to moderately impact one's life (rather than severely impacting on life) there is some more limited evidence that they might also be associated with an increased risk of depression (206).

The review of systematic reviews (or 'meta-review') presented in Chapter 1 showed that there is some evidence that stressful life events are associated with prognosis for people with depression (73,83) but this was based on just two reviews. One of these reviews included just one primary study that investigated this association (73). That study was a cross-sectional follow-up to a prior cross-sectional study in a number of European community settings, and involved 347 adults with depressive disorders of varying types, of which 65 received problem-solving treatment or group psychoeducation, and 34 (it is unclear whether these people included some of the 65 noted above or not) were prescribed some antidepressants (207). The study authors reported that adults that had experienced a stressful life event and had not received support for it, were at greater risk of not recovering from their depression compared to those that had received some or a lot of support (207). Only 182 participants completed the follow-up questionnaire, of which only 75 were rated as being "in recovery". Further, there were relatively few participants that both experienced the life events and did not get support for them, limiting the sample size for the analysis. The second review was conducted in 1994 and it is unclear whether or not it was conducted systematically as limited information was provided on the review methods

(83). It included 29 studies assessing the association of both life events and social support with the "natural course" of depression. The review author reported that there was some evidence of an association between experiencing life events prior to having treatment and poorer course of depression (83). However, many of the included studies had very small samples and included adults with other mental health disorders (e.g. schizophrenia, or those with a recent suicide attempt), and most of those 29 studies were cross-sectional so reverse causality could not be ruled out (83).

In contrast to the review findings, a recent Danish case-register study of 301 adults with MDD reported that neither the experience of any stressful life events prior to the onset of depression, nor the total number of life events experienced, was associated with remission from either first or second line antidepressant medications (208). Similarly, in a randomised controlled trial of 60 patients randomised to cognitive therapy and 120 randomised to antidepressant medications, it was reported that stressful life events were not associated with prognosis with either treatment (209). However, there was a prescriptive effect such that those with more stressful life events were more likely to remit following cognitive therapy compared to antidepressants (209). That same trial also found that those with more life events at baseline were more likely to dropout (209). None of the above studies has assessed the association between the number or specific type of life events and prognosis for depressed patients independent of treatment.

Socio-demographics and Depression

It has been suggested that a number of socio-demographic factors (considered one-by-one, below) are associated with prognosis for adults with depression, but as described in Chapter 1, findings have either been largely mixed or there has been a lack of reviews and in some cases a lack of primary studies investigating these associations.

Age

From the reviews included in Chapter 1 the evidence for an association between age and prognosis was inconclusive as various studies found contradictory effects. For example, some study-level meta-analyses have found that there are larger effect

sizes with psychological therapies delivered to younger patients than in studies delivering the same therapies to older patients (85,86). An IPD meta-analysis reported similar findings for patients treated with antidepressant medications (81). Another study-level meta-analysis and an IPD meta-analysis have found no evidence for an effect of age on prognosis (70,79), and another IPD meta-analysis found that response to antidepressants and to pill placebo improved with increasing age (66). However, evidence from national evaluations of over 600,000 patients to complete psychological therapies in primary care mental health services in England (in Improving Access to Psychological Therapies (IAPT) services) give a contrasting picture, suggesting that older adults (65 years old or above) are considerably more likely to recover with treatment than adults under the age of 65 (210,211). These findings may be affected by selection biases as it has been suggested that older people with more severe problems might be less likely than those with less severe problems to be referred to psychological treatment, and that those with more severe problems might be less likely to attend psychological therapy (212). Investigating the association between age and prognosis for primary care patients independent of a range of treatments may therefore be helpful.

Gender

Depression is more frequently diagnosed in women than in men, and in the UK women are more likely to receive treatment for depression than men (213). However, the review presented in Chapter 1 found that there is little evidence that outcomes from treatment are associated with self-reported gender. Whilst some reviews found women to be less likely to reach remission (84), other reviews suggested an association between gender and treatment outcomes without specifying a direction of effect (66,69,70,72), and four further reviews found no association between gender and treatment outcomes (51,79,81,86). Evidence from the national IAPT programme in England would suggest that there is little difference in the proportions of men and women reaching recovery with psychological therapies: in 2016-17 which is the last year for which there are reported gender-specific results, 46% of women and 46.3% of men were in recovery at the end of their treatment (210). However, that report includes patients treated for anxiety disorders as well as those treated for depression. Therefore the question of whether gender is associated with prognosis

for adults with depression, independent of treatment received in primary care remains.

Ethnicity

The meta-review presented in Chapter 1 found only three reviews which reported on the associations between ethnicity and treatment outcomes for people with depression, with mixed findings. One review reported that patients from minority ethnic groups were less likely to remit with Duloxetine (72), but two other reviews reported no associations between ethnicity and outcomes (51,86). However, reports on IAPT patients have shown that those from Black, Asian or minority ethnic (BAME) backgrounds are less likely to recover following psychological therapies than patients from White ethnic backgrounds (211). Two other studies have found that the ethnicity of IAPT patients is a key variable in determining the probable treatment outcomes from psychological therapies overall (174,175), and from particular treatments, e.g. comparing counselling for depression to cognitive behaviour therapy for depression (175). Some authors have suggested that ethnicity is not directly linked to prognosis but instead factors that are often related to minority ethnic status such as perceived or actual discrimination (214–216), financial insecurity (174), or stressful life events (214), are reported to be associated with prognosis. These factors are known to be important in predictive models of treatment outcomes and are more likely to afflict people from minority ethnic groups (217).

Marital Status

In the meta-review presented in Chapter 1, three reviews investigated the association between marital status and prognosis for people with depression, with mixed findings. One large IPD meta-analysis of 13 RCTs of low-intensity CBT found no effect of marital status on treatment outcomes, although most of the studies drew community samples and the review focussed on prescriptive analyses instead of prognosis (79). A study-level systematic review found that women that were married had better odds of remission when treated with IPT than unmarried women, but this was limited to perinatal samples (85). A second study-level review reported good evidence for an association between being married or cohabiting and greater odds of response to antidepressant treatments, but only five primary studies reported on the

effect, one found no evidence of effect, and only two of the remaining four were rated as providing a high level of evidence for the effect (84).

One potential reason for the mixed findings may be that the degree of support in a spousal relationships is perhaps more important to treatment outcomes than having or not having a spouse (218,219). There is some evidence to suggest that among non-patient samples, single adults appear to have better prognoses than those in unhappy or struggling relationships, however those same studies have found associations between marital status and the course of depression too (220,221). Further, a lack of spousal support, lack of intimacy, and marital discord have all been associated with greater risk of depression onset and of relapse or recurrence of depression (219,222,223). A small cohort study of patients treated with Nefazadone also reported such an association with treatment outcomes (224). Marital status has also been found to be one of the strongest predictors of differential benefit of cognitive therapy over antidepressants, such that married or cohabiting patients were more likely to achieve remission with cognitive therapy than antidepressant medications, there was no prognostic effect of marital status on either treatment though (209).

Employment Status

The meta-review in Chapter 1 highlighted just three reviews that assessed the association between employment status and prognosis for adults with depression. A study-level systematic review reported that employment status was associated with antidepressant treatment response (84). However, this was based on just one primary study. That was a study of 542 depressed participants from a trial of quality improvement interventions at the service-level in six primary care centres in the USA, the authors therefore treated this as a cohort study (225). In that study, patients that were unemployed at baseline were considerably less likely to remit or to respond to antidepressant medications than employed participants. In a non-systematic IPD study employment status was among the top 25 predictors of response to antidepressant treatments (44) but this was based on just three (large) studies. The 25 predictors were selected based on a machine learning variable selection technique and then fed into a second machine learning model to predict outcomes, the nature of the effects in such models is uncertain (i.e. whether they are

main effects or interaction effects), so it is difficult to interpret the direction of the effects. In contrast to the above, employment status was not associated with outcomes from low-intensity CBT treatments in 3876 individual participants of 13 RCTs (79). However, that IPD study was primarily focussed on investigating prescriptive effects not prognostic ones, there were moderate-to-high degrees of heterogeneity in the effects, about a third of the cases had missing endpoint data, and three studies or approximately 16% of eligible participants were not able to be included as no IPD data could be collected from those studies. In addition, most of the 13 included RCTs recruited non-treated samples, limiting the generalisability of the findings to patients seeking treatment for depression. In addition to the reviews highlighted in Chapter 1, another study of 626 outpatients in the Netherlands reported that being in paid employment at baseline was associated with greater odds of remission irrespective of treatment type, and independent of baseline severity (226).

Socio-economic status

The meta-review in Chapter 1 also found three systematic reviews reporting on the association between one or more socio-economic factors and prognosis for adults with depression. A study-level systematic review reported that there was strong evidence for an association between socio-economic status and prognosis for those treated with antidepressants (84). However, that included just one high quality study and that was a trial of Nefazadone compared to cognitive behavioural analysis system of psychotherapy (CBASP) or the combination of both treatments (227). That study reported that those with lower incomes at baseline were more likely to drop-out of the study irrespective of treatment type. A second study-level systematic review of naturalistic cohort studies, included just one primary study that reported on socio-economic factors and their association with prognosis. That study found that owning a home was associated with better odds of recovery from depression in a non-treated sample over the course of nine years (73). A third IPD study found that the effect size of CBT may be higher in patients with disability benefits compared to those without disability benefits, but this was based on just 34 participants with the exposure of interest (88).

Although not reported by any of the reviews highlighted in Chapter 1, poor quality housing has also been linked to the onset of mental health problems, particularly when people live in such conditions for extended periods of time (228). The suggestion is that poor quality housing might also affect treatment outcomes, as failing to resolve issues with housing (which is not usually a primary treatment target for people with depression) may result in a greater likelihood of attrition from treatment and lack of response to treatment (217,228).

Educational Attainment

Two reviews included in Chapter 1 reported on the association between educational attainment and prognosis. One was a study-level systematic review (73) that included two primary studies that found an association between years of education and prognosis, one was the cross-sectional follow-up 9 years after the initial cross-sectional study of 347 adults with depressive disorders in community settings across Europe described above (207). The other was a large population survey that included just 95 adults with depression, it reported that each additional year of education at baseline was associated with a five week reduction in the length of depressive episodes (229). The second review was an IPD study that found educational attainment to be among the top 25 predictors of antidepressant treatment response, and the limitations of that study have been discussed in detail above (44). Another study not included in the meta-review in Chapter 1 was focussed on moderators of deterioration and reported that those with lower levels of educational attainment might be at greater risk of deterioration with internet delivered CBT compared to those with higher levels of educational attainment (230).

Long-term physical health conditions and Depression

Three reviews described in Chapter 1 reported on the association between long-term health conditions and prognosis for adults with depression. One of these was a study-level meta-analysis described in detail above in which poorer outcomes from antidepressants were reported for patients with a number of conditions including: atrial fibrillation, severe heart disease, obesity, and high cholesterol (77). A meta-review reported weak evidence of no association between long-term health conditions and prognosis with pill placebo but that was based on a single primary study in a single review (49). That study-level meta-analysis actually found higher

effect sizes for a variety of antidepressants in RCTs that had long-term health conditions as an inclusion criterion compared to studies that did not (85). However, the interpretation of the association between long-term condition status and prognosis in the above reviews in particularly challenging. This is because the findings come from comparisons of primary studies that for the most part recruited adults with particular comorbidities, with outcomes from trials of adults where comorbidity was not an inclusion criteria. For many of the latter studies, the long-term health condition comorbidities were not exclusion criteria, so the number of adults in those studies that had the same comorbid conditions as those in the studies they were compared to is unknown. No IPD meta-analysis to have assessed these associations were identified, but in a large clinical sample of IAPT patients, those self-reporting a long-term health condition have been found to have poorer outcomes than patients who did not report such conditions (162).

Overall, the reviews and primary studies noted above have not assessed the associations between the factors listed and prognosis for adults with depression independent of treatment. They have also not included large samples drawn from primary care, limiting their generalisability to a large proportion of patients seeking treatment for depression (23). Therefore, investigating such factors independent of treatment and markers of depressive severity using data from the Dep-GP IPD dataset may lead to a greater understanding of prognosis for depressed patients.

Aims

This study aimed to meet the remaining elements of sub-aims 3 and 4 not addressed in the previous chapters, and more specifically to: 1) determine whether life events (the number of events and individual life events), socio-demographic factors (gender, age, ethnicity, marital status, employment status, financial wellbeing, housing status, and the highest level of educational attainment), and long-term health condition status are associated with prognosis independent of treatment for adults with depression treated in primary care, and whether these factors are associated with prognosis independent of depressive symptom severity and depressive ‘disorder severity’; and 2) to determine whether those same factors are associated with attrition, independent of treatment and independent of depressive symptom severity.

Methods & Materials

Identification and Selection of Studies:

The formation of the Dep-GP IPD dataset has been described in Chapters 2-4, and elsewhere (168). To be included in the present analyses studies also had to use the Social Readjustment Rating Scale (a measure of recent stressful life events) (138), collect data on the socio-demographic factors at baseline, and assess the presence or absence of long-term physical health conditions. Details of the measures used in the included studies from Dep-GP that are of relevance to the analyses described here are in Table 2.2.

Characteristics of the included studies

Six studies from the Dep-GP (168) dataset used the necessary questionnaire measures so met inclusion criteria for the present study, these are the same six studies that were included in the analyses presented in Chapter 2.3.

Data Analysis Plan

The analyses for this study were conducted in line with the study protocol and as described in Chapters 2-4, there were some additional elements of the analysis specific to this study, these are outlined below.

Predictors under consideration

Potential baseline predictors of outcome were the total score and individual items of the life events measure, as well as binary variables created to capture patients that had experienced any of the listed life events in the six months prior to baseline (compared to none experienced), and those that had experienced two or more of the life events in the six months prior to baseline (compared to fewer than two). The socio-demographic factors consisted of age, gender, ethnicity, marital status, employment status, financial wellbeing, housing status, and the highest level of educational attainment. Long-term health condition status (capturing those with any self-reported long-term physical health condition comorbid to depression at baseline, compared to those with none).

Adjusting for Covariates

As per the study protocol, treatment allocation, i.e. the randomisation in each study, and factors *a priori* considered to be important covariates (age and self-reported

gender at baseline) were controlled for in all models. Other covariates were adjusted for if they were associated with both the prognostic factor and the outcome, affected the association between the prognostic factor and outcome, were not considered a potential mediator of the association between the prognostic factor and outcome, and were not multi-collinear with the prognostic factor in its association with outcome. Marital status, employment status, financial wellbeing, housing status, highest level of education, social support, and long-term health condition status, were all considered as potential covariates for some variables but not for those that violated the above considerations (e.g. employment status and financial wellbeing were not considered as covariates in the models including the life event related to debt, and marital status was not considered a covariate in the models including the life event related to divorce).

Results

Characteristics of the included studies

In total, six RCTs were identified as meeting inclusion criteria and were able to provide IPD. A description of each included study can be found in Table 4.1, descriptive statistics across the six studies are presented in Table 4.2.

Quality assessments and Risk of Bias

The risk of bias and the quality of evidence for each prognostic indicator was the same for the current study as for that presented in Chapter 4, despite the GRADE ratings being re-done for the prognostic factors under consideration in the present study, see Table 4.3. As before, there were very high levels of agreement between the two reviewers, interrater reliability: Cohen's Kappa k=0.96 for QUIPS and k=1.00 for GRADE.

Descriptive Statistics

The majority of the descriptive statistics for the sample included in the present study are described in Chapter 4 and outlined in Table 4.2. Of the results not shown in Table 4.2: most participants (71.4%) reported at least one life event within the six months prior to their baseline assessment, with the mean (SD) number of events being 1.35(1.24). The most commonly reported life events were suffering a serious illness or injury (33.7%) and problematic debt that could not be paid back if it had to be paid immediately (excluding mortgages or rent arrears) (33.4%). The least

common events were being sacked or losing one's job (6.2%) and being the victim of a violent crime or assault (6.9%). Having had serious disputes with close friends, family or neighbours ("arguments") was reported by 23.6% of the sample, being bereaved of a loved one was reported by 18.9%, separating or getting a divorce from a long-term partner or spouse was reported by 11.1%, and having trouble with the police, personally or having a close relative having to appear in court was reported by 8.2% of the sample. The mean baseline BDI-II scores across the studies were lower for participants with no life events in the six-months prior to baseline (mean(SD) = 28.08(10.17)), than for those with one life event in that time period (mean(SD) = 29.56(10.37)), and for those that had two or more life events in that time period (mean(SD) = 32.92(10.39)).

The majority of participants were female (66.5%), from white ethnic backgrounds (94.4%), and aged between 31 and 53 years old (50%). The majority of participants were employed (57.4%), most were married or cohabiting (48.3%), were doing OK financially (41.5%), had obtained A-levels or higher levels of educational attainment (56.8%), and most participants were home owners (including those with a mortgage) (45.9%). Data on the specific type of long-term conditions (LTCs) was missing at baseline for 847 patients, 516 (18.1%) had at least one LTC of which the most common were asthma or chronic obstructive pulmonary disease (n=182, 6.4%), arthritis (n=105, 3.7%), and diabetes (n=75, 2.6%). The least common LTCs were cancers (n=12, 0.4%), cardiac diseases (n=31, 1.1%), kidney disease (n=50, 1.7%), and stroke (n=61, 2.1%). Due to small numbers of some of these conditions and a relatively high degree of missing data on this variable at baseline, which could not reasonably be assumed to be missing at random due to the inclusion criteria of some of the studies (so imputation of specific types of LTCs was not considered appropriate), it was decided that only the presence or absence of an LTC would be assessed as a potential prognostic factor, rather than the individual types of LTC.

The association between life events, socio-demographics or long-term health condition status and prognosis

Life events

The majority of the life events measured in this study were associated with prognosis independent of treatment, having any stressful life event in the six months prior to baseline was associated with 15.96%(95%CI: 8.64 to 23.78) higher depressive

symptom scale scores at 3-4 months post-baseline independent of treatment. For every additional life event reported at baseline participants had on average 7.86%(95%CI: 4.68 to 11.14) higher depressive symptom scale scores at 3-4 months post-baseline, independent of treatment. There was evidence that being the victim of a violent crime or assault (mean difference = 0.29(95%CI: 0.12 to 0.46), percentage difference = 25.78%(95%CI: 12.22 to 40.98)), and having unpayable debts (mean difference = 0.29(95%CI: 0.17 to 0.41), percentage difference = 21.94%(95%CI: 9.21 to 36.15)), were strongly associated with prognosis independent of treatment. Whereas, being bereaved (mean difference = 0.01(95%CI: -0.09 to 0.11), percentage difference = 2.02%(95%CI: -5.13 to 9.72)), and being sacked (mean difference = -0.06(95%CI: -0.22 to 0.09), percentage difference = 2.37%(95%CI: -9.76 to 16.14)), were not significantly associated with prognosis independent of treatment, see Table 5.1. The associations were weaker when adjusting for depressive symptoms at baseline: only the total score on the life events measure (mean difference = 0.05(95%CI: 0.03 to 0.10), percentage difference = 3.62%(95%CI: 0.12 to 7.24)); having any life events (mean difference = 0.13(95%CI: 0.03 to 0.23), percentage difference = 7.95%(95%CI: -1.36 to 18.14)); having serious disputes/arguments (mean difference = 0.09(95%CI: 0.01 to 0.18), percentage difference = 7.61%(95%CI: 0.59 to 15.11)); debt (mean difference = 0.16(95%CI: 0.03 to 0.29), percentage difference = 11.54%(95%CI: 0.05 to 24.35)); and being the victim of a crime (mean difference = 0.19(95%CI: 0.04 to 0.35), percentage difference = 17.00%(95%CI: 5.23 to 30.09)), were significantly associated with prognosis. After adjusting for depressive ‘disorder severity’ factors too, of the individual life events there was only evidence that debt (mean difference = 0.14(95%CI: 0.01 to 0.27), percentage difference = 10.65%(95%CI: -0.65 to 23.22)), and being the victim of crime (mean difference = 0.15(95%CI: 0.01 to 0.29), percentage difference = 12.30%(95%CI: 1.10 to 24.75)), were associated with the standardised mean of the depressive symptom scales at 3-4 months post-baseline. Only being a victim of crime was associated with the natural logarithm of depressive symptom scale scores at 3-4 months post-baseline, see Table 5.1, and Figures 5.1 and 5.2 for study heterogeneity. Those patients that reported at least one life event had worse prognoses than those reporting none, independent of treatment, depressive symptom severity, ‘disorder severity’, and covariates: mean difference in

depressive symptom scores at 3-4 months post-baseline = 0.12(95%CI: 0.01 to 0.22).

When considering secondary outcomes, after adjusting for treatment, ‘disorder severity’, gender, age, and covariates specific to each prognostic factor: having two or more life events prior to baseline was associated with lower odds of remission at 3-4 months OR=0.82(95%CI: 0.67 to 0.99). There was no clear evidence that any individual life event items were associated with remission. Having had any life events in the six months prior to baseline 0.17(95%CI: 0.06 to 0.28), having debts 0.13(95%CI: 0.02 to 0.25), and being sacked/losing one’s job (0.29(95%CI: 0.03 to 0.56)) were all significantly associated with a higher mean depressive symptom scale scores at 6-8 months post-baseline. There was an absence of evidence that any of the other items were significantly associated with that outcome. None of the life events items were significantly associated with the BDI-II score at 3-4 months when removing the one study without that outcome in a sensitivity analysis, see Table 5.2, or in the additional sensitivity analyses using the PROMIS T-Score outcome (see Appendix 5 Table 2).

Figure 4.1. Forest plot of associations between Life events total score, any life events, and two or more life events, and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors, and covariates.

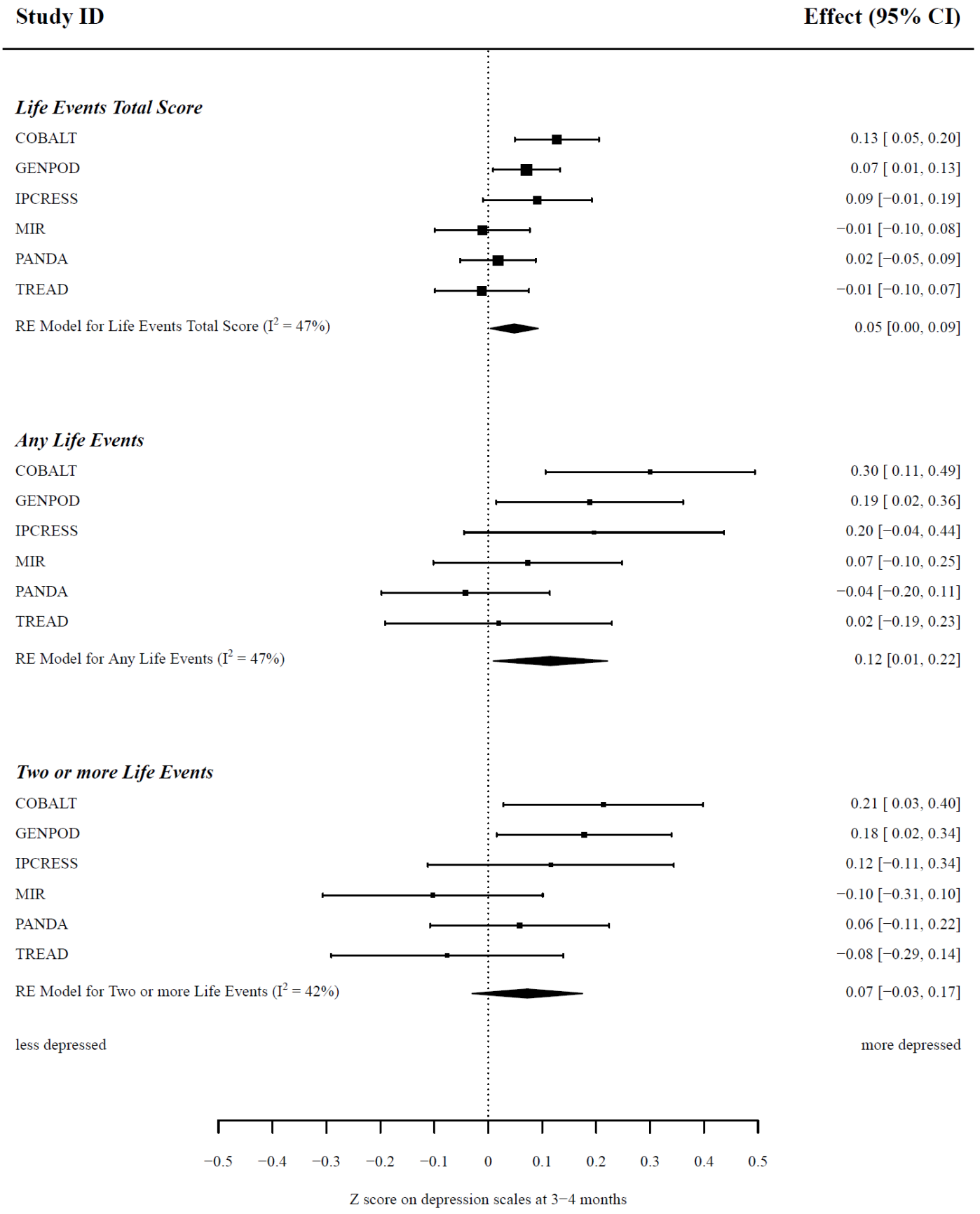
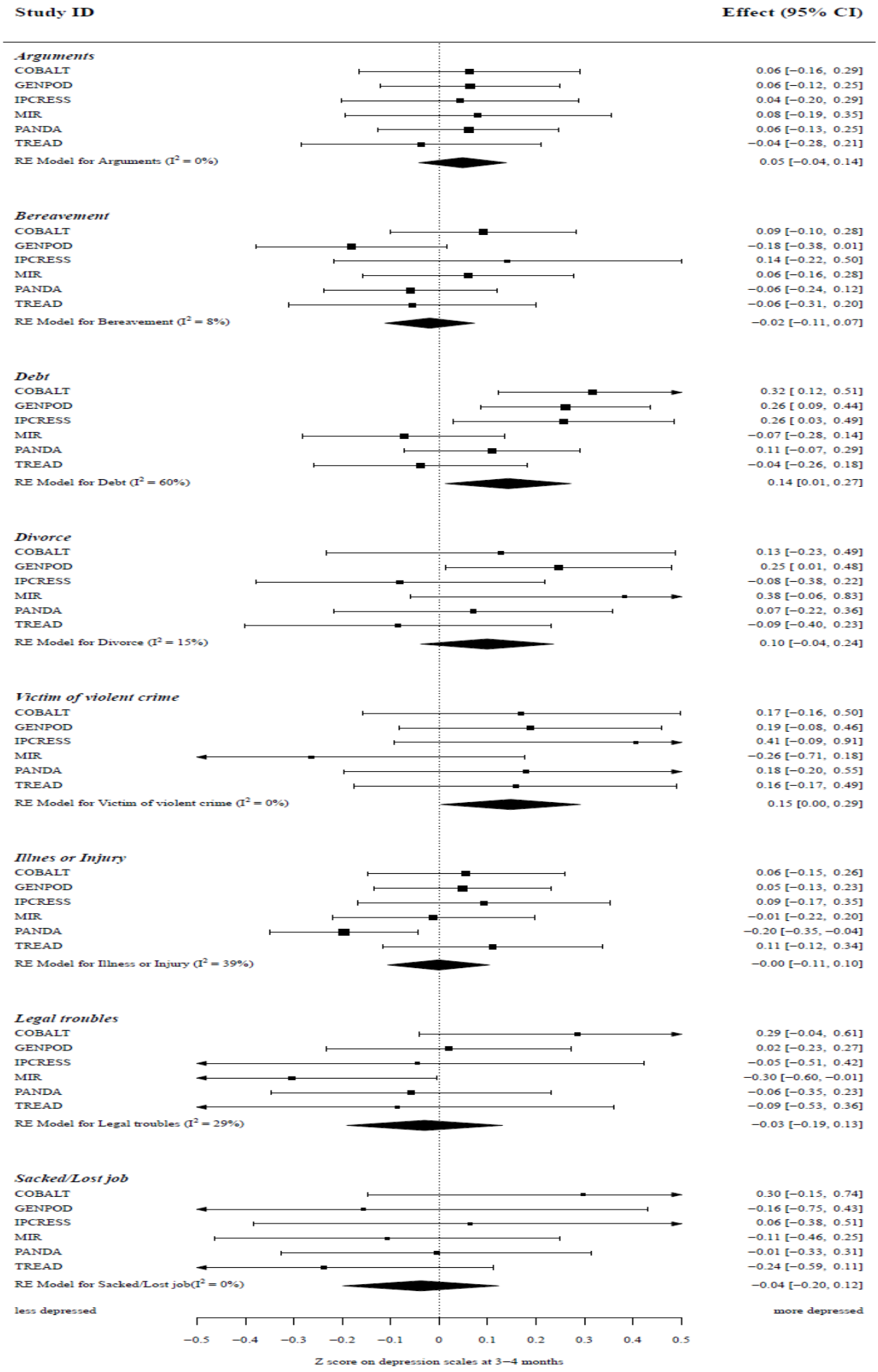


Figure 5.2. Forest plot of associations between Life events and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors, and covariates.



Socio-demographics

Age

There was no evidence that each year increase in age at baseline was associated with prognosis independent of treatment (mean difference = 0.00(95%CI: -0.01 to 0.00); percentage difference in symptoms =-0.31%(95%CI: -0.79 to 0.17)). Heterogeneity was high for the associations between age and prognosis at 3-4 months, particularly with the log outcome where I^2 was 76 for the fully adjusted analysis. When removing the one study that contributed most to this heterogeneity the effect was somewhat closer to the null, although the direction of the effect did not change: percentage difference = -0.15%(95%CI: -0.60 to 0.31).

Gender

There was also no evidence that gender was associated with any of the prognostic outcomes at 3-4 months post-baseline, independent of treatment (mean difference = 0.03(95%CI: -0.08 to 0.14); percentage difference in symptoms =-4.32%(95%CI: -4.39 to 13.82)).

There was some evidence that all of the remaining socio-demographic factors were associated with prognosis after adjusting for treatment, depressive symptom severity, depressive 'disorder severity', gender, age, and other covariates, see Table 5.1.

Ethnicity

Being from a White ethnic background was associated with lower levels of depressive symptoms at 3-4 months relative to being from a non-white ethnic background, although the effect was not apparent without adjusting for depressive symptom severity (independent of treatment mean difference = 0.16(95%CI: -0.05 to 0.38); additionally adjusted for depressive symptom severity and 'disorder severity' factors: mean difference = 0.21(95%C: 0.04 to 0.38), percentage difference =14.18%(95%CI: 0.70 to 29.46)).

Marital status

Being separated, divorced or widowed was associated with higher depressive symptom scale scores at 3-4 months post-baseline relative to being married or cohabiting, independent of treatment (mean difference = 0.15(95%CI: 0.10 to 0.19),

percentage difference = 10.04%(95%CI: 5.98 to 14.26)). The magnitude of association was reduced by adjusting for depressive symptom severity and further by additionally adjusting for 'disorder severity' factors and employment status (mean difference = 0.08(95%CI: 0.03 to 0.12), percentage difference =5.77%(95%CI: 1.55 to 10.16)).

Employment status

Being employed was associated with lower depressive symptom scale scores at 3-4 months independent of treatment, age and gender, relative to not being in employment (mean difference = 0.23(95%CI: 0.13 to 0.32), percentage difference = 18.23%(95%CI: 13.14 to 23.55)). The magnitude of the association was reduced when adjusting for depressive symptom severity and again when additionally adjusting for depressive 'disorder severity' factors and marital status (mean difference = 0.12(95%CI: 0.05 to 0.19), percentage difference = 9.85%(95%CI: 4.49 to 15.47)).

Socio-economic factors

In terms of socio-economic factors, struggling or just about getting by financially was associated with higher depressive symptom scores independent of treatment, age and gender relative to doing OK financially or living comfortably (mean difference = 0.22(95%CI: 0.16 to 0.28), percentage difference =14.60(95%CI: 8.09 to 21.51)). As above, the magnitude of the associations were smaller when adjusting for all factors: (mean difference = 0.08(95%CI: 0.03 to 0.13), percentage difference = 4.29%(95%CI: 0.27 to 8.47)). Not being a homeowner was also associated with higher depressive symptom scores at 3-4 months post-baseline relative to being a homeowner, independent of treatment, age, gender, depressive symptom severity, 'disorder severity' and employment status (mean difference = 0.12(95%CI: 0.05 to 0.19), percentage difference = 9.95%(95%CI: 3.73 to 16.54)).

Educational attainment

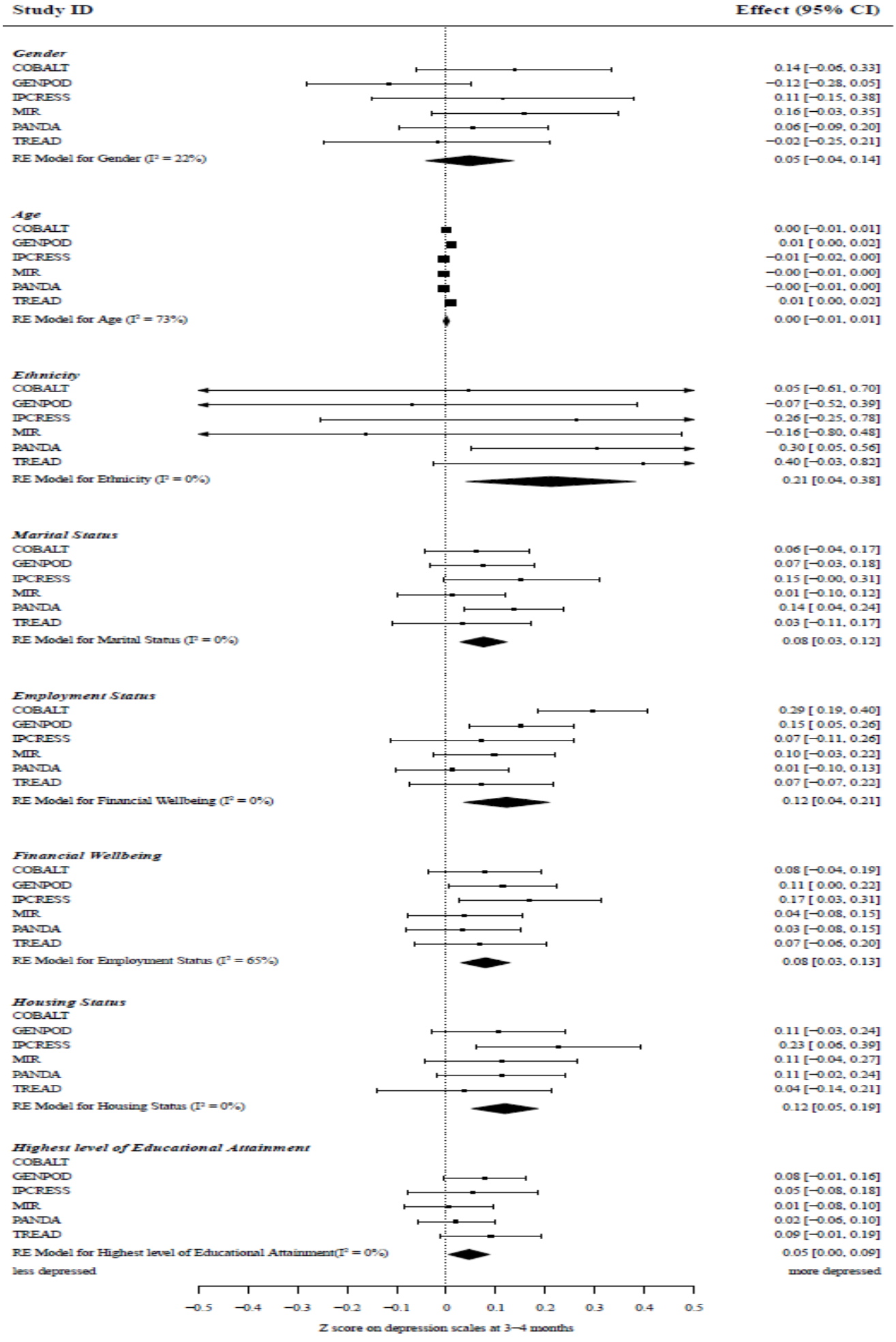
Each unit decrease in the highest level of educational attainment (going from degree level education or above, to A-levels or diplomas, to GCSEs, to no formal educational qualifications) was associated with higher depressive symptom scale scores at 3-4 months post-baseline independent of treatment, age and gender (mean difference = 0.10(95%CI: 0.06 to 0.15), percentage difference = 5.06(95%CI:

1.79 to 8.44)). However, after additionally adjusting for depressive symptom severity and ‘disorder severity’ factors, the association was only significant with the z-score outcome (mean difference = 0.05(95%CI: 0.01 to 0.10)), not the log outcome (percentage difference = 1.43(95%CI: -1.62 to 4.58)).

See Figure 5.3 for details of study heterogeneity for all socio-demographic factors.

There were similar patterns of results with secondary outcomes, although gender was significantly associated with prognosis at 6-8 months post-baseline independent of depressive symptom severity and ‘disorder severity factors’ (mean difference = 0.17(95%CI: 0.02 to 0.32)), ethnicity was not significantly associated with remission after making all adjustments (odds ratio for non-white compared to white ethnicities = 0.93(95%CI: 0.75 to 1.11)), the 6-8 month post-baseline outcome (mean difference = 0.13(95%CI: -0.28 to 0.
55)), or with the BDI-II score at 3-4 months in a sensitivity analysis (mean difference = 2.07(95%CI: -0.08 to 4.21)), see Table 5.2. In a further sensitivity analysis using the PROMIS T-score as the outcome: marital status, employment status, financial wellbeing and housing status were significantly associated with prognosis at 3-4 months post-baseline after adjusting for all variables, but other socio-demographics were not, see Appendix 5 Table 2.

Figure 5.3. Forest plot of associations between Socio-demographic factors and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors, and covariates.



Long-term physical health condition status

Having any long-term physical health condition comorbid to depression at baseline was associated with worse prognosis at 3-4 months post-baseline, independent of treatment, age and gender (mean difference = 0.15(95%CI: 0.05 to 0.25), percentage difference = 13.66(95%CI: 2.99 to 26.29)). The magnitude of the associations between LTC status and prognosis decreased as more variables were adjusted for. After adjusting for all factors above, having an LTC was associated with higher depressive symptom scale scores at 3-4 months post-baseline relative to not having an LTC (mean difference = 0.12(95%CI: 0.01 to 0.22), percentage difference = 11.13%(95%CI: 1.39 to 21.81)), see Table 5.1, and see Figure 5.4 for details of study heterogeneity. LTC status was associated with remission after adjusting for all variables as specified in Table 5.2 (odds ratio for having an LTC compared to not having one = 0.77(95%CI: 0.60 to 0.99)), but there was no clear evidence of association with prognosis at 6-8 months after making all adjustments (mean difference = -0.03(95%CI: -0.34 to 0.28)). There was limited evidence of an association between LTC status and the BDI-II score at 3-4 months in a sensitivity analysis, see Table 5.2. LTC status was associated with prognosis when using the PROMIS T-score outcome in a further sensitivity analysis, see Appendix 5 Table 2.

Heterogeneity was within acceptable limits in all analyses apart from age, so no further sensitivity analyses were necessary based on heterogeneity.

Figure 5.4. Forest plot of associations between LTC status and the z-score of depressive symptom scales at 3-4 months post-baseline independent of treatment, depressive symptom severity, depressive 'disorder severity' factors, and covariates.

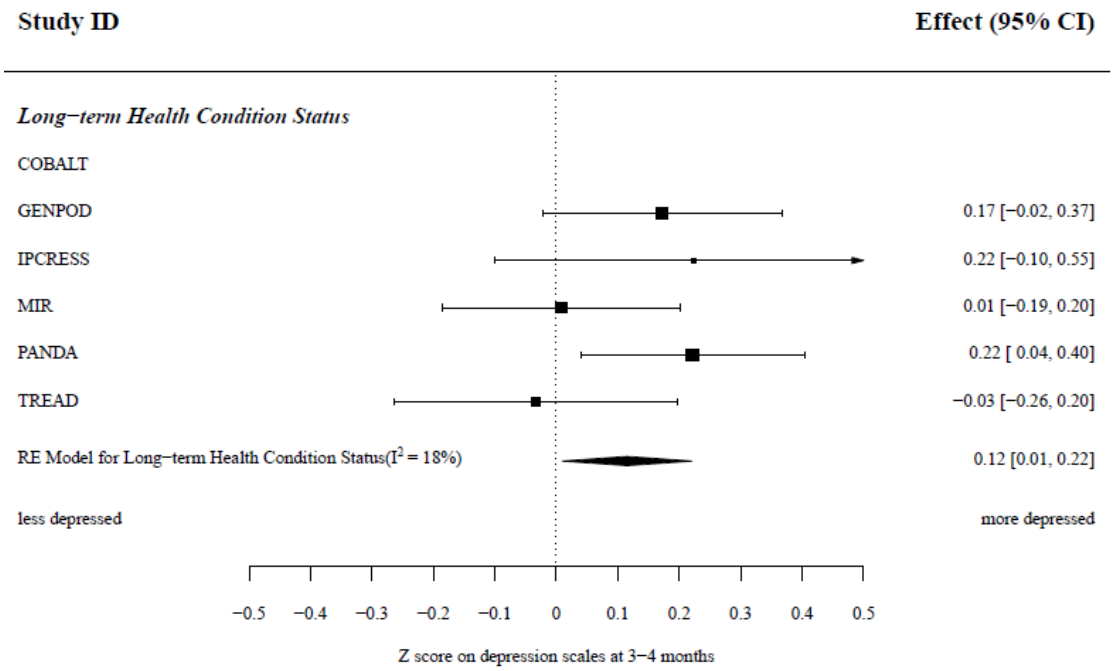


Table 5.1. Outcomes at 3-4 months (difference in z-score of depressive symptoms and % difference in depressive symptoms) per unit increase in baseline prognostic factors.

		Difference or percentage difference in outcomes at 3-4 months post-baseline per unit increase in baseline Prognostic Indicator											
Prognostic Factors		Effect independent of treatment				Effect independent of symptom severity and treatment				Effect independent of 'disorder severity' and treatment			
Type	Prognostic Indicator	Z-score of depressive symptoms^		% difference in depressive symptoms^		Z-score of depressive symptoms*		% difference in depressive symptoms*		Z-score of depressive symptoms ‡		% difference in depressive symptoms‡	
		Mean Difference (95%CI)	I^2	%(95%CI)	I^2	Mean Difference (95%CI)	I^2	%(95%CI)	I^2	Mean Difference (95%CI)	I^2	%(95%CI)	I^2
Life events	Life events total score	0.11(0.08 to 0.15)	23	7.86(4.68 to 11.14)	27	0.05(0.01 to 0.10)	50	3.62(0.12 to 7.24)	47	0.05(0.00 to 0.09)	47	3.26(-0.15 to 6.79)	37
	Any life events	0.23(0.15 to 0.31)	0	15.96(8.64 to 23.78)	0	0.13(0.03 to 0.23)	43	7.95(-1.36 to 18.14)	48	0.12(0.01 to 0.22)	47	6.77(-2.97 to 17.49)	50
	Two or more life events	0.22(0.13 to 0.32)	24	16.91(8.85 to 25.56)	20	0.08(-0.03 to 0.19)	50	7.28(-0.44 to 15.61)	31	0.07(-0.03 to 0.18)	42	6.41(-0.84 to 14.20)	46
	Arguments	0.23(0.13 to 0.32)	0	17.22(9.22 to 25.08)	0	0.09(0.01 to 0.18)	0	7.61(0.59 to 15.11)	0	0.05(-0.04 to 0.14)‡1	0	3.47(-3.42 to 10.86)‡1	0
	Bereavement	0.01(-0.09 to 0.11)	0	2.02(-5.13 to 9.72)	0	-0.01(-0.10 to 0.08)	0	0.28(-7.19 to 8.34)	16	-0.02(-0.11 to 0.07)	8	0.05(-7.48 to 8.19)	14
	Debt	0.29(0.17 to 0.41)	46	21.94(9.21 to 36.15)	62	0.16(0.03 to 0.29)	60	11.54(0.05 to 24.35)	64	0.14(0.01 to 0.27)	60	10.65(-0.65 to 23.22)	60
	Divorce	0.19(0.06 to 0.32)	0	12.92(1.94 to 25.08)	0	0.10(-0.03 to 0.24)	12	6.35(-3.81 to 17.57)	3	0.10(-0.04 to 0.24)‡2	15	6.76(-3.30 to 17.87)‡2	9
	Victim of violent crime/assault	0.29(0.12 to 0.46)	7	25.78(12.22 to 40.98)	0	0.19(0.04 to 0.35)	12	17.00(5.23 to 30.09)	0	0.15(0.01 to 0.29)‡3	0	12.30(1.10 to 24.75)‡3	0
	Illness or Injury	0.00(-0.16 to 0.15)	66	-2.48(-11.53 to 7.49)	48	0.00(-0.13 to 0.12)	58	-2.09(-9.96 to 6.46)	38	-0.00(-0.11 to 0.10)‡4	39	-2.03(-9.24 to 5.76)‡4	26
	Legal troubles	0.16(0.01 to 0.30)	3	13.36(1.85 to 26.17)	0	0.04(-0.14 to 0.22)	43	5.15(-4.87 to 16.24)	0	-0.03(-0.19 to 0.13)	30	0.33(-9.49 to 11.20)	0
	Sacked/Lost job	-0.06(-0.22 to 0.09)	1	2.37(-9.76 to 16.14)	0	-0.05(-0.20 to 0.16)	3	2.25(-9.15 to 15.09)	0	-0.06(-0.21 to 0.08)	0	0.86(-10.47 to 13.63)	0
Socio-demographics	Age†	0.00(-0.01 to 0.00)	0	-0.31(-0.79 to 0.17)	74	0.00(0.00 to 0.01)	73	0.15(-0.33 to 0.64)	76	0.00(-0.01 to 0.01)	73	0.03(-0.47 to 0.54)	78
	Gender†	0.03(-0.08 to 0.14)	6	4.32(-4.39 to 13.82)	42	0.07(-0.02 to 0.16)	22	7.23(-0.52 to 15.59)	31	0.05(-0.04 to 0.14)‡5	22	4.59(-3.29 to 13.11)	34
	Ethnicity	0.16(-0.05 to 0.38)	19	11.68(-3.93 to 29.83)	29	0.19(0.02 to 0.37)	0	13.35(-0.07 to 28.58)	0	0.21(0.04 to 0.38)	0	14.18(0.70 to 29.46)	0
	Marital Status	0.15(0.10 to 0.19)	30	10.04(5.98 to 14.26)	22	0.11(0.06 to 0.15)	0	7.76(3.99 to 11.67)	0	0.08(0.03 to 0.12)‡6	0	5.77(1.55 to 10.16)‡6	0
	Employment Status	0.23(0.13 to 0.32)	10	18.23(13.14 to 23.55)	54	0.15(0.06 to 0.24)	69	12.36(7.17 to 17.80)	35	0.12(0.04 to 0.21)‡3	64	9.85(4.49 to 15.47)‡3	33
	Financial Wellbeing	0.22(0.16 to 0.28)	3	14.60(8.09 to 21.51)	12	0.10(0.06 to 0.15)	0	6.14(1.59 to 10.90)	22	0.08(0.03 to 0.13)‡3	0	4.29(0.27 to 8.47)‡3	29
	Housing Status	0.18(0.10 to 0.26)	5	14.53(7.26 to 22.30)	0	0.12(0.05 to 0.19)	0	9.98(3.76 to 16.56)	0	0.12(0.05 to 0.19)‡6	0	9.95(3.73 to 16.54)‡6	0
	Highest level of Educational Attainment	0.10(0.06 to 0.15)	0	5.06(1.79 to 8.44)	0	0.05(0.01 to 0.09)	0	1.92(-1.09 to 5.03)	12	0.05(0.01 to 0.10)	21	1.43(-1.62 to 4.58)	0
	Long-term health condition status	0.15(0.05 to 0.25)	5	13.66(2.29 to 26.29)	24	0.12(0.03 to 0.22)	14	12.02(1.338 to 23.78)	26	0.12(0.01 to 0.22)	18	11.13(1.39 to 21.81)	19

Note: Total score is scaled 0-8 and unless otherwise stated, estimates are per one-point increase; individual items are scored 0-1 and estimates are per 1-point increase.
^adjusted for allocated treatment, gender, and age; * additionally adjusted for depressive symptom severity; ‡ additionally adjusted for average anxiety duration, depression duration, comorbid panic disorder, and history of treatment with antidepressants; † not adjusted for itself; ‡1 additionally adjusted for marital status and social support total score; ‡2 additionally adjusted for financial wellbeing; ‡3 additionally adjusted for marital status; ‡4 additionally adjusted for marital status, employment status, financial wellbeing, and social support total score; ‡5 additionally adjusted for social support total score; ‡6 additionally adjusted for employment status

Table 5.2. Associations of life events, socio-demographics, and long-term health condition status with secondary outcomes, adjusted for treatment, 'disorder severity', gender and age.

Prognostic Factors		Secondary Outcomes				Sensitivity Analysis	
Type	**Prognostic Indicator**	**Remission at 3-4 months^**		**z-score at 6-8 months**		**BDI-II score at 3-4 months†**	
		OR(95%CI)	**I^2**	**Mean Difference (95%CI)**	**I^2**	**Mean Difference (95%CI)**	**I^2**
Life events	Life events total score	0.92(0.82 to 1.02)	38.63	0.04(-0.01 to 0.09)	0.00	0.41(-0.04 to 0.87)	14.43
	Any life events	0.82(0.65 to 1.04)	39.00	0.17(0.06 to 0.28)	0.00	0.93(-0.13 to 1.98)	14.00
	Two or more life events	0.82(0.67 to 0.99)	2.00	-0.01(-.012 to 0.11)	8.00	0.56(-0.71 to 1.84)	35.00
	Arguments	0.84(0.67 to 1.06)‡1	0.00	0.08(-0.07 to 0.23)‡1	22.00	0.70(-0.46 to 1.87)‡1	0.00
	Bereavement	0.98(0.75 to 1.28)	24.45	-0.05(-0.18 to 0.09)	0.00	-0.27(-1.49 to 0.96)	7.50
	Debt	0.84(0.64 to 1.10)	41.87	0.13(0.02 to 0.25)	0.00	1.12(-0.55 to 2.79)	57.52
	Divorce	0.80(0.58 to 1.08)‡2	0.00	-0.13(-0.32 to 0.06)‡2	0.00	1.00(-0.92 to 2.92)‡2	31.81
	Victim of violent crime/assault	0.78(0.54 to 1.12)‡3	0.00	0.10(-0.12 to 0.32)‡3	0.00	1.75(-0.27 to 3.78)‡3	10.00
	Illness or Injury	1.17(0.96 to 1.44)‡4	0.00	0.02(-0.11 to 0.15)‡4	5.83	-0.04(-1.15 to 1.06)‡4	10.66
	Legal troubles	1.07(0.75 to 1.51)	3.00	-0.07(-0.38 to 0.24)	38.94	-1.31(-3.01 to 0.39)	0.00
	Sacked/Lost job	0.98(0.66 to 1.44)	0.00	0.29(0.03 to 0.56)	38.61	-1.21(-2.98 to 0.56)	31.05
Socio-demographics	Age†	1.00(0.99 to 1.02)	63.02	0.00(-0.01 to 0.00)	0.00	0.01(-0.07 to 0.08)	73.47
	Gender†	0.91(0.75 to 1.11)‡5	0.00	0.17(0.02 to 0.32) ‡5	39.57	0.03(-1.19 to 1.25)‡5	31.38
	Ethnicity	0.93(0.56 to 1.53)	25.40	0.13(-0.28 to 0.55)	56.83	2.07(-0.08 to 4.21)	0.00
	Marital Status	0.86(0.76 to 0.97)‡6	0.00	0.05(-0.01 to 0.12)‡6	0.00	0.85(0.23 to 1.47)‡6	0.00
	Employment Status	0.81(0.66 to 0.99)‡3	57.78	0.15(0.00 to 0.25)‡3	54.17	0.94(0.17 to 1.54)‡3	5.06
	Financial Wellbeing	0.88(0.78 to 1.00)‡3	0.00	0.09(0.00 to 0.17)‡3	33.62	0.86(0.17 to 2.12)‡3	0.00
	Housing Status	0.79(0.66 to 0.93)‡6	0.00	0.13(0.03 to 0.23)‡6	0.00	1.33(0.55 to 2.12)‡6	3.04
	Highest level of Educational Attainment	0.94(0.79 to 1.12)	61.00	0.06(0.00 to 0.11)	0.00	0.37(-0.20 to 0.94)	0.00
	Long-term health condition status	0.77(0.60 to 0.99)	0.00	-0.03(-0.34 to 0.28)	73.32	1.09(-0.02 to 2.19)	0.00

Note: Total score is scaled 0-8 and unless otherwise stated, estimates are per one-point increase; individual items are scored 0-1 and estimates are per 1-point increase.

All models are adjusted for allocated treatment, gender, age; depressive symptom severity, average anxiety duration, depression duration, comorbid panic disorder, and history of treatment with antidepressants; † not adjusted for itself; ‡1 additionally adjusted for marital status and social support total score; ‡2 additionally adjusted for financial wellbeing; ‡3 additionally adjusted for marital status; ‡4 additionally adjusted for marital status, employment status, financial wellbeing, and social support total score; ‡5 additionally adjusted for social support total score; ‡6 additionally adjusted for employment status

The association between life events, socio-demographics or long-term health condition status and attrition

Life events

Several of the life events items were associated with attrition at 3-4 months post-baseline, independent of treatment. The pattern of associations was similar to that with the prognostic outcomes: there was evidence that the total number of life events (odds ratio per life event increase =1.23(9%%CI: 1.12 to 1.35)); having had any life events (odds ratio for any compared to none = 1.75(95%CI: 1.36 to 2.25)); or having had two or more life events in the six months prior to baseline (odds ratio for two or more compared to one or fewer = 1.48(95%CI: 1.20 to 2.03)) were each associated with attrition independent of treatment. Further, having serious arguments (OR =1.56(95%CI: 1.20 to 2.03)), debt (OR = 1.45(95%CI: 1.17 to 1.80)), being the victim of violent crime (OR = 1.58(95%CI: 1.09 to 2.28)), and having legal troubles (OR = 1.71(95%CI: 1.22 to 2.39)), were also all significantly associated with attrition independent of treatment. Being bereaved (OR = 1.02(95%CI: 0.71 to 1.46)), getting a divorce (OR = 1.37(95%CI: 1.00 to 1.89)), having a serious illness (OR = 1.23(95%CI: 0.99 to 1.53)), and being sacked (OR = 1.49(9%CI: 0.96 to 2.30)), were not significantly associated with attrition independent of treatment, see Table 5.3. The magnitude of effect was reduced when adjustments were made for depressive symptom severity, depressive 'disorder severity' and covariates, but none of the significant associations became non-significant when making such adjustments. For some of the items displayed in Table 5.3 it was not possible to calculate the change in attrition. Within some studies when splitting the prognostic indicators into their component categories there was no attrition, so the confidence intervals would not be interpretable and therefore the corresponding cells of Table 5.3 have been left blank. This also resulted in only five studies being included in the analyses of life events related to being the victim of a violent crime, legal troubles, and being sacked, as there was no attrition with participants that had those life events in the COBALT study at the 3-4 month endpoint.

Socio-demographics

There was no evidence that the following were associated with attrition independent of treatment: gender (OR = 1.11(95%CI: 0.76 to 1.61)); ethnicity (OR = 1.39(95%CI: 0.76 to 2.52)); or employment status (OR = 0.98 (95%CI: 0.83 to 1.17). After

adjusting for depressive symptom severity, ‘disorder severity’ and covariates specific to each variable, the following were significantly associated with attrition at 3-4 months post-baseline: financial wellbeing was associated with higher odds of attrition (OR = 1.33(95%CI: 1.15 to 1.54)), as was not being a homeowner (OR = 1.24(95%CI: 1.03 to 1.86)), and lower levels of educational attainment (OR = 1.35(95%CI: 1.20 to 1.52)). For each year increase in age at baseline the odds of attrition were slightly lower after adjusting for treatment, depressive symptom severity and ‘disorder severity’ (OR = 0.98(95%CI: 0.97 to 0.99)), see Table 5.3.

Long-term physical health condition status

There was little evidence attrition was any more or less likely among those with self-reported long-term physical health conditions at baseline. This was the case independent of treatment, age, and gender: (OR = 0.85(95%CI: 0.64 to 1.12)), and likewise, independent of depressive symptom severity (OR = 0.84(95%CI: 0.64 to 1.11)), and depressive ‘disorder severity’ (OR = 0.85(95%CI: 0.64 to 1.13)).

Table 5.3. Association of life events, socio-demographics, and long-term health condition status with study attrition at 3-4 months post-baseline.

		Difference or percentage difference in outcomes at 3-4 months post-baseline per unit increase in baseline Prognostic Indicator								
		Effect independent of treatment			Effect independent of symptom severity and treatment			Effect independent of 'disorder severity' and treatment		
Type	**Prognostic Indicator**	**OR(95%CI)^**	**I^2**	**%(95%CI) difference in attrition^**	**OR(95%CI)***	**I^2**	**%(95%CI) difference in attrition ***	**OR(95%CI)‡**	**I^2**	**%(95%CI) difference in attrition ‡**
Life events	Life events total score	1.23(1.12 to 1.35)	31.08	17.22(10.32 to 24.55)	1.21(1.09 to 1.34)	39.28	16.01(8.73 to 23.77)	1.21(1.08 to 1.37)	50.10	16.00(7.68 to 24.96)
	Any life events	1.75(1.36 to 2.25)	0.00	55.73(26.25 to 92.10)	1.70(1.31 to 2.19)	0.00	52.00(23.18 to 87.55)	1.74(1.35 to 2.25)	0.00	54.60(25.29 to 90.76)
	Two or more life events	1.48(1.20 to 2.03)	0.00	34.57(14.42 to 58.28)	1.45(1.17 to 1.79)	0.00	32.11(12.21 to 55.53)	1.46(1.18 to 1.81)	0.00	32.70(12.61 to 56.37)
	Arguments	1.56(1.20 to 2.03)	21.23	40.44(14.57 to 72.15)	1.53(1.19 to 1.96)	14.21	37.83(13.91 to 66.78)	1.53(1.14 to 2.04)‡1	26.18	50.75(12.17 to 102.61)‡1
	Bereavement	1.02(0.71 to 1.46)	39.01	2.52(-23.22 to 36.91)	1.00(0.70 to 1.42)	36.99	0.87(-23.97 to 33.82)	0.98(0.69 to 1.37)	29.39	5.10(-22.67 to 42.83)
	Debt	1.45(1.17 to 1.80)	0.00	33.95(13.72 to 57.78)	1.41(1.14 to 1.75)	0.00	31.39(11.47 to 54.86)	1.46(1.17 to 1.83)	0.00	34.44(13.76 to 58.87)
	Divorce	1.37(1.00 to 1.89)	9.22	30.38(5.92 to 60.49)	1.35(0.97 to 1.88)	13.61	29.05(3.99 to 60.15)	1.34(0.95 to 1.88)‡2	10.84	27.56(-5.03 to 71.33)‡2
	Victim of violent crime/assault	1.58(1.09 to 2.28)	0.00		1.52(1.05 to 2.20)	0.00		1.45(1.01 to 2.11)‡3	0.00	
	Illness or Injury	1.23(0.99 to 1.53)	0.00	17.29(-0.68 to 38.51)	1.23(0.99 to 1.54)	0.00	17.37(-0.61 to 38.6)	1.21(0.96 to 1.51)	0.00	15.31(-2.49 to 36.35)
	Legal troubles	1.71(1.22 to 2.39)	0.00		1.68(1.20 to 2.36)	0.00		1.57(1.11 to 2.22)‡4	0.00	
	Sacked/Lost job	1.49(0.96 to 2.30)	11.03		1.49(0.97 to 2.30)	9.58		1.48(0.93 to 2.37)	18.21	
Socio-demographics	Age†	0.97(0.97 to 0.98)	17.48	-2.01(-2.81 to -1.20)	0.98(0.97 to 0.99)	0.00	-1.84(-2.53 to -1.13)	0.98(0.97 to 0.99)	0.00	-1.86(-2.56 to -1.16)
	Gender†	1.11(0.76 to 1.61)	61.77	9.01(-18.06 to 45.02)	1.13(0.78 to 1.63)	60.95	10.6(-16.62 to 46.71)	1.13(0.81 to 1.59)‡5	49.62	9.59(-15.73 to 42.51)
	Ethnicity	1.39(0.76 to 2.52)	39.75		1.37(0.76 to 2.47)	37.95		1.38(0.74 to 2.55)	39.89	
	Marital Status	1.15(1.00 to 1.32)	0.00	12.33(0.58 to 25.45)	1.13(0.99 to 1.30)	0.00	11.29(-0.37 to 24.31)	1.15(1.00 to 1.32)‡6	0.00	13.91(1.77 to 27.49)‡6
	Employment Status	0.98(0.83 to 1.17)	30.23	-1.34(-13.70 to 12.80)	0.95(0.80 to 1.13)	31.90	-3.73(-15.96 to 10.28)	0.93(0.79 to 1.11)‡3	26.28	-7.72(-18.52 to 4.52)‡3
	Financial Wellbeing	1.35(1.17 to 1.55)	13.24	26.29(13.73 to 40.22)	1.31(1.13 to 1.52)	16.94	24.19(11.97 to 37.74)	1.33(1.15 to 1.54)‡3	14.20	28.83(14.53 to 44.92)‡3
	Housing Status	0.22(0.05 to 0.39)	0.00	17.64(3.89 to 33.20)	0.22(0.05 to 0.39)	0.00	18.06(4.17 to 33.80)	1.24(1.03 to 1.86)‡6	25.34	19.05(4.95 to 35.03)‡6
	Highest level of Educational Attainment	0.32(0.19 to 0.44)	0.00	27.98(16.93 to 40.07)	0.30(0.18 to 0.42)	0.00	26.85(15.95 to 38.77)	1.35(1.20 to 1.52)	0.00	27.41(16.46 to 39.38)
	Long-term health condition status	0.85(0.64 to 1.12)	0.00	-12.25(-29.85 to 9.77)	0.84(0.64 to 1.11)	0.00	-12.64(-30.06 to 9.12)	0.85(0.64 to 1.13)	0.00	-6.29(-27.84 to 21.70)

Note: Total score is scaled 0-8 and unless otherwise stated, estimates are per one-point increase; individual items are score 1-3 and estimates are per 1-point increase.
^adjusted for allocated treatment, gender and age; * additionally adjusted for depressive symptom severity; ‡ additionally adjusted for average anxiety duration, depression duration, comorbid panic disorder, and history of treatment with antidepressants; † not adjusted for itself; ‡1 additionally adjusted for marital status and social support total score; ‡2 additionally adjusted for financial wellbeing; ‡3 additionally adjusted for marital status; ‡4 additionally adjusted for marital status, employment status, financial wellbeing, and social support total score; ‡5 additionally adjusted for social support total score; ‡6 additionally adjusted for employment status

Discussion

In this study I found that the number of life events experienced in the six months prior to starting treatment was associated with prognosis at 3-4 months post-baseline independent of treatment. Those that had experienced any severely stressful life event had higher depressive symptoms at 3-4 months post-baseline, on average, compared to patients that had not experienced such events in that time period. The associations between the number of events or having any compared to no life events and prognosis were considerably weaker after adjusting for baseline depressive symptoms, reflective of the fact that those with no life events had significantly fewer depressive symptoms at baseline. Further, most of the individual life events were associated with prognosis independent of treatment. When adjusting for depressive symptom severity there was limited evidence for an association with prognosis for several of the life events, and after adjusting for depressive 'disorder severity' factors too, there was good evidence of association with prognosis at 3-4 months post-baseline for only two life events items. Victims of violent crimes in the six-months prior to baseline and those with problematic levels of debt both had more depressive symptoms at 3-4 moths post-baseline. Two life events were associated with poorer prognosis at 6-8 months post-baseline: problematic debt and having been sacked/losing one's job. Attrition was more likely if participants had more life events prior to baseline, reported having serious arguments or disputes, had problematic debt, were victims of violent crime, or reported legal troubles.

There was good evidence that being from a non-white ethnic background, not being married, being unemployed, not doing OK financially, not being a homeowner, and having at least one long-term physical health condition comorbid to depression were all significantly associated with worse prognosis at 3-4 months post-baseline, but there was no evidence of such associations for age or gender. There were mixed findings regarding the highest level of educational attainment as those with lower levels of attainment had marginally higher average depressive symptom scale scores at 3-4 months post-baseline, relative to those with higher levels of educational attainment, independent of all variables adjusted for. However, the association when using the log outcome at 3-4 months post-baseline was not significant, neither were associations with secondary outcomes such as remission at

3-4 months, the z-score at 6-8 months, or the sensitivity analyses using the BDI-II score at 3-4 months or the PROMIS T-score at 3-4 months. For the other factors the effects were broadly the same across outcomes except that there was no evidence that ethnicity was associated with remission, and no evidence that either gender or long-term health condition status were associated with prognosis at 6-8 months post-baseline.

Several of the socio-demographic factors were also associated with attrition independent of treatment, depressive 'disorder severity', and covariates. Attrition was more likely among younger participants than older ones, those with worse financial wellbeing, those that were not homeowners, and those with lower levels of educational attainment.

Findings in Context

There was some evidence that life events are associated with prognosis from prior studies but there was a degree of inconsistency in those findings and no previous studies had assessed this independent of treatment. Finding that the total number of life events prior to starting treatment was not as strongly associated with prognosis after adjusting for 'disorder severity' as having any life events was, might lead us to consider whether there could be a 'ceiling effect' to the prognostic associations between life events and outcomes from treatment. This is in keeping with one small case register study that found having one life event was associated with worse odds of remission with antidepressants than having no prior life events, but that there were no differences between one event and two events, or three or more events (208). However, another study found that the association between life events and attrition from either cognitive therapy or antidepressants occurred with life events modelled as a z-score in which the mean was approximately seven (209). Whether or not there is a ceiling effect was not tested here. Despite the suggestion that the experience of any severely stressful life event may be associated with prognosis (83,205), here the magnitude of associations of the individual life events with the prognostic outcomes was not uniform. Being the victim of a violent crime was associated with about 12% higher depressive symptom scores at 3-4 months post-baseline. Having debts that could not be paid back if required (excluding mortgages

or rent arrears) was associated with approximately 11% higher scores at 3-4 months. The estimated percentage difference in depressive symptom scores for those that had legal troubles, were sacked, or were bereaved was between 0-1% despite these all being considered to be severely stressful events too.

The associations between both age and gender with prognosis had been well studied previously, but with mixed and inconsistent results. They had also not been studied independent of a range of treatments, and rarely in IPD studies which are most useful to study such associations. Here, no evidence was found that either age or gender were associated with any of the prognostic outcomes. This is out of keeping with studies of clinical cohorts from the English national IAPT programme, from which it has been found that older adults are considerably more likely to recover at the end of their psychological treatment than are adults under the age of 65 (210,211). However, a recent study of approximately 100,000 patients from eight IAPT services has shown that the above effect is driven by patients with anxiety disorders and that once controlling for symptom severity and a number of other covariates, there is only a small effect of age on outcomes for depressed patients (212). Further, that study also found that attrition was less likely in older adults than younger adults (212), in the present study each year increase at baseline was associated with an approximate 2% reduction in the odds of attrition.

There has been a lack of studies investigating the associations between the other socio-demographic factors and prognosis for depressed patients, particularly independent of treatment. Here there was good evidence that such factors are associated with prognosis at 3-4 months post-baseline, independent of treatment and a number of other factors that might have strong associations with prognosis, such as depressive symptom severity and the ‘disorder severity’ factors. Some of these socio-demographic factors were also strongly associated with the prognostic outcomes, for example: non-white participants had poorer prognoses than those from white backgrounds but only after adjusting for baseline depressive symptoms, as on average White patients had slightly higher symptom severity scores that non-white participants. After adjusting for all severity factors and covariates, non-white participants on average experienced an estimated 14% higher score on depressive symptoms scales at 3-4 months post-baseline after adjusting for treatment. The

differences for each category increase in employment status (from employed to not seeking employment, to unemployed) and housing status (from being a home owner, to being a tenant, to having some other less secure housing status such as living in a hostel or being homeless) were estimated to be approximately 9% each, after adjusting for treatment, all severity variables and covariates. However, unlike previous studies, in this study there was inconsistent evidence for an association between educational attainment and prognosis with significant associations with prognosis at 3-4 months using the z-score outcome, but non-significant associations with all other prognostic outcomes after adjusting for all specified variables. It is noteworthy though that associations between educational attainment and prognosis had not previously been studied independent of treatment and all variables adjusted for here. There was however evidence that those with higher levels of educational attainment, and those with better socio-economic circumstances, were less likely to experience attrition.

Having a long-term physical health condition comorbid to depression is considered to be related to poorer prognosis (212) but as with the other factors above, had not been tested independent of a range of treatments. In this study, those with any long-term physical health conditions were less likely to reach remission and on average had an estimated 11% higher score on the depressive symptom scales at 3-4 months post-baseline compared to those with no LTCs, independent of treatment and all severity factors adjusted for.

Strengths and Limitations

In addition to the strengths and limitations discussed in Chapter 4 where I used the same set of studies, there were a number of other strengths and limitations of the present study. This was the first study to consider patient's characteristics prior to commencing treatment for depression sought from a GP, to assess the associations between life events, socio-demographic factors and long-term health conditions, with prognosis independent of treatment. This study also investigated these associations independent of a range of indicators of the severity of depression at baseline rather than just using depressive symptoms. Many prior studies of these factors were based on differences in the means between groups receiving the same sort of treatment, others were based on small samples or largely non-treated samples

affecting the generalisability of the findings for depressed adults seeking treatment (208,209,231). In contrast, the findings here were based on individual participant data, drawing on studies with relatively large samples of adults that all sought treatment in primary care, so they may be generalisable to a large proportion of depressed patients.

The large number of tests conducted might have increased the chance of making Type 1 errors (232). However, following recommendations by Rothman (199) no adjustments were made here, and although there were no definitive hypotheses regarding the direction of effects prior to conducting these analyses, all analyses have been presented irrespective of statistical significance. This does not remove the possibility of some Type 1 errors but does mitigate some of the potential issues of mining data for associations discussed by other authors (232,233).

As discussed in Chapter 3 one limitation with the Dep-GP dataset is that the prognostic factors being investigated in this study were self-reported and this may have led to increased measurement error in some of those factors (this is more likely to be the case for life events than for the sociodemographic variables). It has previously been shown that depressed patients are more likely to exhibit cognitive biases which effect recall of negative events (177) and that those biases are associated with treatment outcomes for adults with depression treated in primary care (34), so it is possible that baseline depressive symptoms confounded the associations between life events and prognosis. However, as in previous analyses presented in this thesis, here adjustments were made for both depressive symptom severity and the broader concept of depressive ‘disorder severity’. While these adjustments ameliorated some of the associations between life events and prognosis, they did not do so for all. Knowing which factors are associated with prognosis after these adjustments might have clinical utility.

Some of the findings may have been subject to selection biases: all included studies sought to recruit adults in primary care but some limited their inclusion criteria to exclude some older adults, e.g. COBALT and GENPOD had upper age limits of 75 and 74 years old respectively, and TREAD had the upper age limit of 69 years old. In general, across the studies there were very few adults from BAME backgrounds, limiting the sample size available to analyse the effects of ethnicity on prognosis and

attrition. This probably reflects a degree of selection bias in the study populations as BAME adults were generally under-represented relative to expected prevalence of depression in BAME groups in the communities serving the study recruitment sites (23). However, all but one of the RCTs that make up the Dep-GP IPD dataset could be considered pragmatic trials, and as noted in Chapter 1, this should improve the representativeness of the study samples, reducing selection biases and potentially improving the chances of the findings generalising beyond the study sample (57).

Implications

As noted above, not all life events were associated with prognosis with the same magnitude and this was particularly the case after adjusting for depressive symptoms at baseline. The two items most strongly associated with prognosis at 3-4 months post-baseline after making such adjustments were being the victim of a violent crime and having problem debts. It might be suggested that for depressed treatment-seeking patients, clinicians should consider asking about such life events in the six-months prior to patients presenting at health services, and consider onward referrals or additional support specific to these events (such as to victim support organisations, or debt advice services) (234). It is possible that unresolved sequelae of these events may act as barriers to the potential benefits of treatments for depression (209,231). Similarly, although based on less evidence (as only four studies contributed data at the 6-8 month endpoint), the finding that becoming unemployed was associated with prognosis at 6-8 months might lead to a referral to employment support specialists (209,226). In addition, those with many different life events were more likely to drop out or withdraw from treatment and this might suggest that greater support could be required in order to mitigate against attrition for such patients, or that means of treatment that are more readily accessible or for which attrition is known to be lower, may be important to consider.

In regards to a patient's 'hierarchy of needs' (234) it may help to consider how likely it is that someone with such issues who is not getting additional support for them (whether facilitated/provided by their clinician or not), may be able to fully engage with and benefit from the treatment provided for their depression. Indeed, patients with worse financial or housing statuses were also more likely to dropout or withdraw from treatment (209,227,235).

The association between long-term health condition status and prognosis at 3-4 months post-baseline might also lead clinicians to consider additional support, but this may need to take a different form compared to support offered to mitigate against or resolve social issues. Instead, such support may mean referrals to other health teams or community support services are necessary, and it may be useful to consider the appropriateness and accessibility of certain treatments for patients depending on their particular needs in relation to their LTCs.

In terms of future research, it might be useful to measure life events and many of the socio-demographic factors here, particularly the socio-economic factors, for studies of prognosis. It is noteworthy that relatively few participants from BAME backgrounds participated in the studies included in the present analysis. To better understand the associations between ethnicity and prognosis, greater efforts might be required to recruit participants in all communities of patients, or stratified sampling techniques might be employed to ensure more representative samples are obtained. Further, there were insufficient data to consider the specific types of LTCs patients may have presented with here, partly because the variable used to collect this information was a count of LTCs, and partly because data on specific LTCs were missing for approximately 30% of patients here. To better understand the associations between LTCs and prognosis, not only should such information be collected, but it would be useful to collect data on the duration of the LTC, on patient's perceptions about the LTC(s) they have, and expectations of benefit (or lack of benefit) from treatment (236,237).

Conclusions

In conclusion, this study has shown that important information can be gained about prognosis independent of treatment for adults with depression when considering life events, socio-demographics, and long-term health condition status. The factors investigated here are all easily measured with self-report questionnaires and although their associations with prognosis may be biased by the severity of depression, particularly so for life events, the fact that these associations were present after adjusting for a broad range of markers of severity supports the potential robustness of the findings here.

Chapter 6. Discussion and implications of results for understanding and predicting prognosis for adults with depression in primary care

Overview

In this thesis I have presented a rationale for the compilation of an IPD dataset of adults with depression, set out a protocol for a series of analyses using the data, and presented the findings from three sets of analyses. In this final chapter I will summarise the main findings of the work presented in this thesis, and will consider the general strengths and limitations of the work conducted. I will discuss how the findings might inform further research and how the findings might be utilised clinically, prior to such further research, and afterwards.

Thesis summary

The focus of this thesis was on the assembly of an individual patient data (IPD) dataset, taking into consideration the methodological issues that can arise when doing so. This involved using a systematic approach to both identify an appropriate pool of randomised controlled trials of adults to form an IPD dataset, and a systematic approach to the analyses of the data. The research presented in this thesis led to the formation of the Depression in General Practice (Dep-GP) IPD dataset, and used these data to investigate the associations between a number of patient characteristics measured pre-treatment, and the prognosis for patients independent of treatment. In particular, this thesis was concerned with whether or not there was evidence for such associations after controlling for the effect of baseline depressive symptoms, and after also controlling for other markers of the severity of depression. Where appropriate, to aid considerations of the utility of any such associations, the clinical importance of differences in the prognosis of patients with different levels of each baseline characteristic were considered. The discussion that follows below outlines the key findings from the research presented in this thesis and considers the wider context of the results, including any potential clinical implications, and will finish with a discussion of further research that might build on the work presented here.

Key findings

The systematic searches of the literature outlined in Chapter 2 led to 13 studies being found to meet the inclusion criteria for the IPD dataset. 12 study teams were able to provide IPD leading to the formation of the Dep-GP IPD dataset from the 6271 patients included in those studies. One study could not be included as individual patient data were no longer available and the aggregate data that could be provided were not sufficient to conduct any of the planned analyses. That study (118) was considerably smaller than the others that were included, so overall those included in the Dep-GP IPD dataset represented approximately 98% of the patients across all eligible studies.

The main findings in regards to prognosis at 3-4 months and attrition at 3-4 months post-baseline from analyses in Chapters 3-5 are displayed in Figure 6.1. The first set

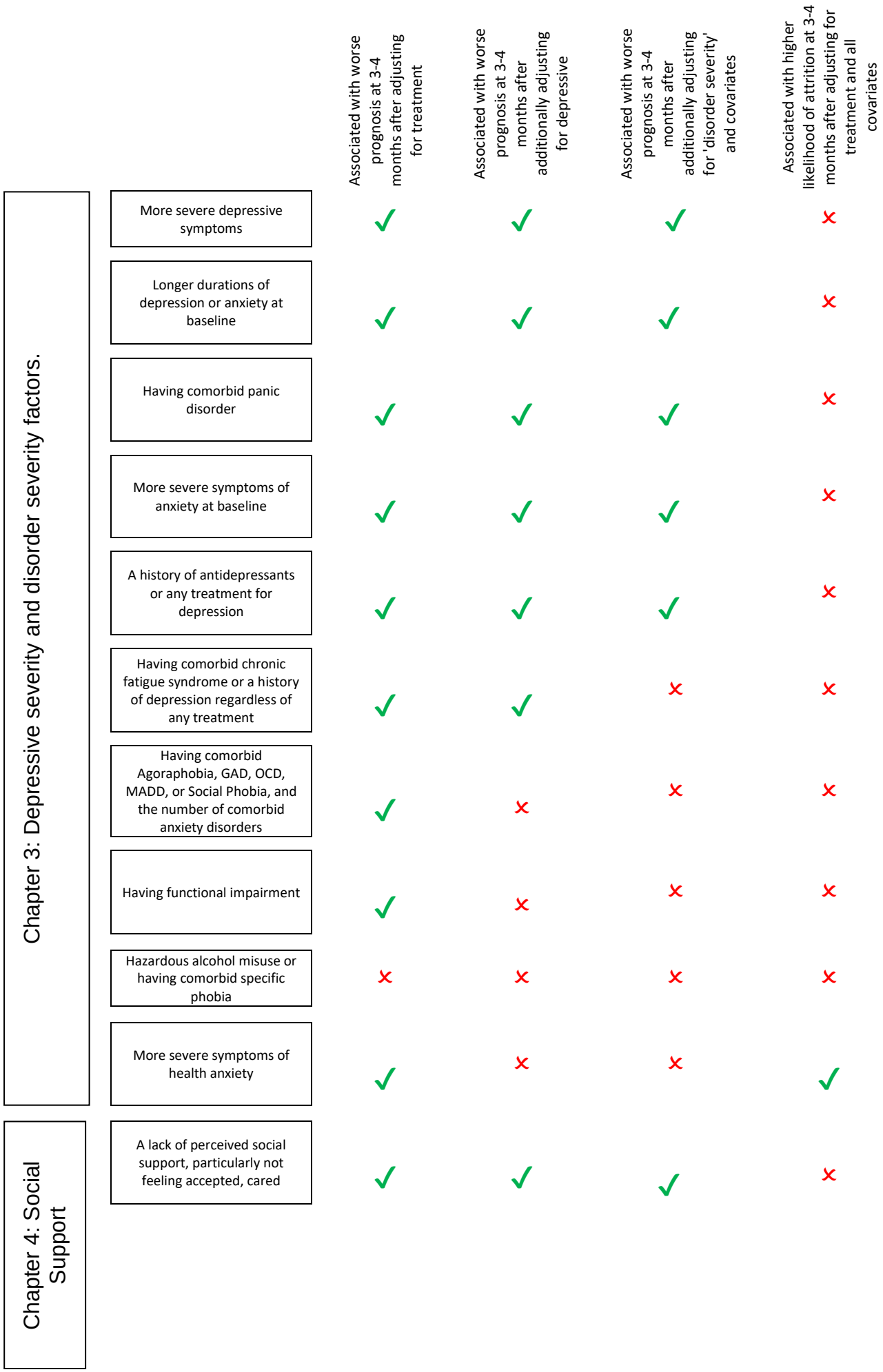
of analyses of the Dep-GP data demonstrated that differences in depressive symptom severity prior to seeking treatment can lead to clinically important variations in prognosis independent of treatment, and that symptom severity may account for approximately 16% of the variance in prognosis. Other markers of severity, separate from but associated with depressive symptom severity, which were termed indicators of depressive 'disorder severity' were also associated with prognosis independent of treatment, and independent of depressive symptom severity. These were: 1) the duration of depression; 2) the average duration of anxiety problems; 3) comorbid panic disorder; and (with less robust evidence) 4) a history of antidepressant medication. The duration of anxiety could be substituted by the severity of anxiety symptoms, and a history of antidepressants could be substituted by a history of any past treatment for depression (irrespective of the type of treatment) but these were less strongly associated with prognosis. When adjusting for depressive symptom severity and for each of the four 'disorder severity' variables, the amount of variance explained in prognosis independent of treatment rose to approximately 27%. There was no evidence for associations between alcohol misuse and prognosis, or functional impairment and prognosis, independent of depressive symptom severity. The only severity marker associated with attrition from treatment after adjusting for depressive symptoms at baseline was the severity of health anxiety related symptoms at baseline.

In the analyses presented in Chapter 4 there was evidence that social support was associated with prognosis independent of treatment and of both depressive symptom severity and depressive 'disorder severity'. Three individual items from the Social Support Scale measure (137) were more consistently (across different outcomes) associated with prognosis, these were feeling accepted for who one is, feeling supported and encouraged, and feeling cared about by family or friends. The latter two were also associated with attrition independent of all variables adjusted for (treatment, age, gender, employment status, marital status, depressive symptom severity, and the 'disorder severity' factors).

In Chapter 5 the associations between life events, socio-demographics (age, gender, marital status, employment status, financial wellbeing, housing status, and the highest level of educational attainment), and long-term health condition status with

prognosis were examined. In these analyses two particular life events were more strongly and consistently associated with prognosis independent of treatment and of ‘disorder severity’, than the other events, these were: being the victim of a violent crime and having problem debts. Ethnicity, marital status, employment status, financial wellbeing, housing status, and long-term health condition status were also associated with prognosis independent of all factors adjusted for, but age and gender were not. There were mixed findings regarding the highest level of educational attainment with limited evidence for an association with one prognostic outcome and no evidence for associations with other prognostic outcomes. Similar associations were found with attrition.

Figure 6.1 Summary of findings from IPD meta-analyses in Chapters 3-5.



	about*, or supported*[2] by friends or family				
	A perceived lack of being made to feel happy, important, that can rely on or talk to friends and family when needed	✓	✗	✗	✓
Chapter 5: Life events	Reporting any severe life events in six months prior to treatment, particularly debt or being victim to a violent crime or assault	✓	✓	✓	✓
	Reporting any being divorced, having serious arguments, or legal troubles in six months pre-baseline	✓	✓	✗	✓
	Reporting illness or injury, being sacked, or being bereaved in six months prior to baseline	✗	✗	✗	✗
Chapter 5: Socio-demographics	Being single or no longer married, being unemployed or not seeking employment	✓	✓	✓	✗
	Not doing OK financially or not being a homeowner	✓	✓	✓	✓
	Lower highest levels of educational attainment	✓	✓	✗	✓
	Non-White Ethnicity	✓	✓	✓	✗
	Female gender	✗	✗	✗	✗
	Younger age	✗	✗	✗	✓
Chapter 5: LTCs	Having comorbid long-term physical health conditions	✓	✓	✓	✗

[2] *feeling cared about, or supported or encouraged by family or friends were associated with attrition although the total social support score was not.

Overall strengths and limitations

There were a number of strengths to the research conducted as part of this thesis: the Dep-GP IPD dataset is the largest of its kind, giving rise to the opportunity to explore associations which may have small effects on prognosis, which single studies or aggregate level meta-analyses do not have the power to detect.

The studies included in Dep-GP all recruited participants from primary care settings giving clarity to the minimum population for whom results might be generalizable. Whether or not they could be generalizable beyond the UK primary care setting is discussed in the chapters above and no definitive answer on this can be drawn from the data assembled as part of Dep-GP. It is clear that restricting inclusion criteria for the IPD to studies that recruited in primary care led to a reduced number of otherwise potentially eligible trials. In order to ensure greater generalisability, an alternative might have been to not restrict inclusion criteria in such a way, and instead to conduct analyses on sub-groups of studies based on the setting for recruitment, or on whether or not the setting was clarified in the studies. However, many of the treatments utilised in the studies included in the Dep-GP are used in many countries (including a range of antidepressant treatments and both low-intensity and high-intensity psychological therapies), and findings regarding prognosis for adults with depression receiving these types of treatment, conducted across the world, are largely in keeping with the findings here (78,209,238). This might suggest that the results here could potentially be generalizable to other health care settings and systems, and suggests results here may be more robust than had a narrower range of treatments been used in the included studies, as past studies have shown some prognostic effects to be specific to one or other type of treatment (44,72,81). There was also a range of subtypes of depressed patients in the study samples, for example some studies specifically targeted patients for whom it was uncertain whether or not any treatment for depression would be required (e.g. PANDA (127)), and other studies specifically included patients with ‘treatment resistant depression’ (e.g. COBALT (121)) and those with depression during the perinatal period (e.g. RESPOND (129)). Again, this could mean that the results may be generalizable to a broad population of adults with depression.

Another strength is that the Dep-GP was set up such that there was a common assessment of severity factors across the studies, resulting in little need for the harmonization of the baseline data, thereby reducing the chance of additional error or bias to be introduced at that stage (46,63). This measure afforded a thorough consideration of an array of 'disorder severity' factors. However, that common measure: the CIS-R, has no measure of symptoms related to traumatic stress, which may be commonly comorbid with depression (239) and influential in the prognosis of both disorders (44), and in the likelihood of dropping out of treatment (240). In addition, there have been criticisms of the CIS-R as a fully structured and lay or self-administered screening tool, in particular with regards to the determination of primary diagnoses by the CIS-R scoring algorithm (241). The primary diagnoses determined by the CIS-R were not used in the any of the studies presented in this thesis, negating the impact of any issues with the validity of these diagnoses in the studies presented above. Perhaps of greater import though, the choice of CIS-R as an inclusion criterion undoubtedly limited the number of studies that were found to meet inclusion criteria for the Dep-GP. Although it was more commonly used than other clinical interviews, it might have been possible to include studies using those less commonly used interviews too, and then have conducted subgroup analyses per-measure to address issues of harmonizing biases. This is unlikely to have affected the findings in Chapter 3 to a great extent as most other available studies would not have had data on the majority of assessed factors. However, it might have been possible to include many more studies in Chapters 4 and 5 had I not used the same inclusion criteria. This would have been particularly relevant after conducting the analysis for Chapter 3. For example, findings from that chapter could have been used to reduce the number of 'disorder severity' factors that were deemed important to adjust for to the point that the CIS-R was no longer a necessary inclusion criteria when considering prognostic associations for the variables assessed in the later chapters.

In further consideration of harmonizing data across the studies, the initial intention when compiling the Dep-GP IPD was to use a composite measure that 'cross-walks' the scores on multiple outcome measures to a single score. However, no such composite measures exists for all of the depressive symptom measures used in the Dep-GP studies. It would nonetheless have been possible to use a composite

measure in the analyses presented here in Chapters 4 and 5 as they involved subsets of the Dep-GP studies which all used the BDI-II or PHQ-9 at the primary endpoint. One such composite measure exists and allows for the cross-walk of scores on the BDI-II and PHQ-9 (among other measures not used in any Dep-GP studies) to the PROMIS T-Score (146,242). However, applying the cross-walk directly to the total scores observed on the BDI-II and PHQ-9 led to discrepancies in the PROMIS T-scores given to patients that might be considered to have approximately equivalent levels of symptom severity. For example, scores of zero on the BDI-II are given a PROMIS T-score of 14.6 when using the PROMIS cross-walk tables (242), and for those scoring zero on the PHQ-9 the cross-walked score is 22.9. For those scoring the maximum on the BDI-II the cross-walked score is 137.7 and on the PHQ-9 it is 112.2. It was therefore determined that this method was not likely to give balanced results between studies using the different measures, and would introduce a potentially high degree of measurement error.

An alternative to the cross-walk tables is to use a newer, multidimensional item-response theory (IRT) informed version of the PROMIS which utilises all of the scores on the individual items of the measures rather than their total scores, and which assumes scores on a latent trait factor underlying the measures are more accurately comparable (243). However, on this multidimensional IRT composite measure scores of zero on the BDI-II and PHQ-9 were four points apart on the PROMIS T-score, and scores at the cut-off for remission (just meeting remission) (9 on PHQ-9 and 10 on BDI-II) varied depending on the pattern or responses across the items of each measure; they ranged from 57.9 to 65.2 on PHQ-9 and from 51.2 to 58.8 on BDI-II. It was therefore decided that analyses utilising the multidimensional IRT informed PROMIS T-score would be conducted for the purposes of sensitivity checks only, and instead a z-score of the depressive symptom scale scores across the different measures would be used for the primary outcome.

A further limitation of the outcome chosen in the studies presented here was the need to amalgamate across time points in different studies in order to get a feasible common endpoint for analysis. For example, the primary endpoint for the studies presented in this thesis was between three and four months post-baseline, this might

have resulted in comparisons across studies of patients with different doses of treatment having been received. However, large studies of antidepressants have shown that most have their maximal impact between eight and 12 weeks (32); for both high and low intensity psychological therapies most symptomatic change occurs within the first four weeks (33,244); and although a small subset, perhaps around 14% of patients might show significant symptomatic changes later in therapy, this happens around six weeks into treatment (33). So, it is likely that despite potential differences in doses between three and four months, the patients receiving the treatments would have made most of their symptomatic changes prior to the three-month endpoint and perhaps significant changes between three and four months would have been unlikely for most patients. Further, this chronological as opposed to treatment termination-based endpoint afforded two other potential benefits. Firstly, by using approximately the same endpoint across studies we can be more confident that findings here apply to prognosis with acute-phase treatment rather than amalgamating outcomes from maintenance or continuation phase treatments or combining symptom severity post-treatment with relapses or recurrences of depression (149). Secondly, as this endpoint was determined chronologically relative to baseline, it might facilitate more concrete conversations about prognosis with patients that are particularly useful when there are several treatment options of different durations being considered.

There was a pragmatic approach applied to the calculation of attrition as an outcome variable. Each study team that sent IPD for Dep-GP had a different way of measuring attrition: some studies collected specific information on all of the reasons for attrition occurring with free text boxes explaining patient's own reasons, although these were often left blank; other studies had a small set of categories for attrition including withdrawal by the patient, withdrawal by the study team, withdrawal by the patient's GP, and loss to follow-up; and other studies simply noted that a patient did not complete the assessments at a given endpoint. So, it was decided that attrition would be evaluated as an outcome irrespective of the reason for it occurring.

As noted above, Dep-GP contained 12 studies and a large sample size, however, a limitation of the studies presented here is that some of the main variables of interest were not collected in a number of those studies. This resulted in the main analyses

in Chapter 3 being conducted on the patients from nine studies (n=4290) and those in Chapters 4 and 5 being conducted on the patients from six studies (n=2858). The missing data could have been imputed using multi-level imputation methods as discussed in the Dep-GP protocol (168) and in Chapter 2. However, the sensitivity analyses presented in Chapter 3 demonstrated that this was most likely unnecessary as the exclusion of studies without the variables of interest made a negligible difference to the overall effect of the variables that were present in all studies, and there were very few differences in the overall prognostic effect of depressive symptom severity on prognosis when using bi-variate meta-analyses to include all 12 studies. Further, inspection of the forest plots presented in Figures 3.1 and 3.2 showed that for the other markers of severity assessed in Chapter 3, exclusion of the three studies that were not included in the analyses in Chapters 4 and 5 would not have greatly altered the main effects of those analyses. It is possible that their exclusion would nonetheless have affected the findings related to other prognostic factors (social support, life events, socio-demographics and long-term condition status).

In addition, there were a number of factors noted in the reviews I reviewed in Chapter 1 that might potentially be associated with prognosis independent of treatment, but there were no data available on these factors in any of the Dep-GP studies. These were temperament (75), positive and negative emotionality (91), connectivity between a number brain regions including the subgenual cortex, prefrontal cortex and midbrain regions (87), and metabolism in the right anterior insula (87). However, evidence for the association between prognosis and the above factors was extremely limited and none of the studies assessed these associations independent of treatment.

Overall implications & future directions

Forming and using IPD datasets

As outlined in Chapter 2, putting together an IPD dataset is a considerable amount of work, often best undertaken by a team of researchers, perhaps over several years, as it involves many different areas of work to ensure data are useable for analyses and that all necessary data security and information governance regulations are satisfied. The Dep-GP IPD could not have been formed without significant input from

the sponsors of this work and a number of collaborators all willing to contribute data and advice, both on solutions to problems that arose with the data and on methods of analysis (see the Acknowledgements section for specifics of these people and the roles they played). It is noteworthy that the grant gained to support the work presented in this thesis (and other related pieces of work) was awarded in April 2016, I delayed the start of it for a year until April 2017 in order to allow me to complete other work and also for the time to run systematic literature searches, contact study authors and request IPD from them. Despite the additional year, the final data for this IPD were not obtained until October 2019.

Preliminary analyses were run prior to the last dataset being received, allowing for the final analyses to be set up to run very quickly once those data were cleaned and harmonized with the previously received data. Yet as with every other study added to the IPD dataset, the final one had a number of data problems which meant analyses were far from straightforward. Some of the problems arose because data required to run the analyses presented here were not part of the main planned analyses in the individual RCTs that formed Dep-GP. The result was that there were a number of errors with some of these data, inconsistent recording of some of the data, and a loss of data that had not been recognised until data were requested for Dep-GP (see Chapter 2 for details). Therefore, a very large proportion of the work involved to conduct the analyses presented here was not related to data analysis, instead it primarily revolved around project management and learning how to identify and navigate potential problems due to various international, national, and local pieces of legislation and policy affecting the compilation of an anonymised secondary dataset. This included changes brought about during the process of work on this thesis, such as the introduction by the European Union of the ‘General Data Protection Regulation’ (GDPR) which necessitated a number of additional contracts and changes to the way data were shared, stored, and utilised. It also involved identifying and resolving problems within the data obtained from each study team.

Researchers wishing to conduct similar work would be well advised to consult with specialists in research contracts, university ethics and sponsorship, and undertake training in study management and information governance including modules on

ensuring compliance with legislation and local policies, and further training in research database management, prior to setting up an IPD dataset.

One of the key points in the process of determining how the Dep-GP IPD dataset would be set up was to consider in what context prognosis ought to be studied to meet the aims of this thesis and other work related to it as part of the wider grant. As outlined in Chapter 1, many prior IPD datasets have sought to assess prognosis in the context of: 1) response to one or other type of treatments; 2) to test prescriptive (or interaction) effects between baseline variables and response to two or more types of treatment; or 3) have used cohort or general population studies to assess the "natural course" of depression with no treatment, or irrespective of any treatment (whether or not treatment was measured). However, as argued in Chapter 1, I considered that the most useful context for understanding general prognostic factors is prognosis independent of treatment. In order for the Dep-GP to be used for studies in this context it was necessary to search for studies that included treatment, preferably ones that randomised to different treatments, and that those treatments should be widely available in primary care. There are of course many other sorts of treatment that are used in primary care and beyond, as detailed in Chapter 1, that were not part of any of the Dep-GP studies. It might therefore be beneficial to consider ways in which studies of other treatments could be added to Dep-GP or to consider the formation of a new IPD dataset not bound by the inclusion criteria for Dep-GP. The most pertinent of those criteria is probably that all studies had to use the CIS-R at baseline. As noted above, the CIS-R was necessary for the purpose of studying 'disorder severity' factors appropriately, ensuring they were assessed in a uniform fashion across studies and that there was a depth to the number of different factors assessed. However, on the basis of the evidence provided in the analyses in Chapter 3, it might potentially be appropriate to include studies that have captured such factors in other ways, perhaps even in different ways across studies. Indeed, the sensitivity analyses in Chapter 3 using the z-score of scores on the different scales of anxiety symptoms across the studies, found that this was associated with prognosis at 3-4 months independent of treatment, age, gender, depressive symptoms, and other 'disorder severity' factors available in all studies. This would allow for many more studies to be included in the Dep-GP and may further improve the generalisability of the findings.

Research and clinical implications

Taken together, the studies in this thesis have demonstrated that in addition to considering the severity of depressive symptoms for determinations of prognosis, it may be clinically important to also consider the four depressive 'disorder severity' factors (duration of anxiety, duration of depression, comorbid panic disorder, and a history antidepressant medication). Further, it may be clinically useful to consider social support, life events, socio-demographics and long-term health condition status as these are informative for prognosis independent of treatment and independent of both depressive symptom severity and depressive 'disorder severity'. Including these factors may help improve predictions of prognosis, the models of prognosis outlined in Chapter 3 compared favourably with other similar models, however, if these findings were applied clinically, caution would be required as there are likely to be a number of other, unknown factors that could also have an impact on prognosis.

Individual items representing some of the above constructs found to be associated with prognosis here were more strongly associated with the outcomes than others, for example: social support items related to feeling accepted for who one is, feeling supported or encouraged, and feeling cared for by friends and family; along with life events related to being the victim of a violent crime and having problematic debts. In addition, marital status, employment status and financial wellbeing or housing status, and long-term health condition status were all independently associated with prognosis and attrition. If more accuracy is to be gained in determining prognosis independent of treatment from routine assessments of adults with depression presenting in primary care, then these items might be borne in mind by clinicians during assessments of their patients. It is noteworthy that a number of the above factors have been found to be associated with the onset or development of depression, so they might themselves be targets for interventions (183,245). It is likely that a number of clinicians already consider the above factors when assessing depressed patients and may weigh up the information gained from such assessments to inform considerations of prognosis. However, ensuring these are assessed routinely, by all clinicians where appropriate, could be important, hence they might also be added to routine outcome measurement systems that operate in a number of primary care settings (211). As all of the factors found to be independently

associated with prognosis in this thesis were either able to be captured with single self-report questions or with the same computer administered self-report questionnaire, they might in some circumstances be able to be captured prior to consultations with clinicians, where infrastructure exists to link online surveys or other platforms to electronic patient records.

Future directions

The prognostic factors highlighted in this thesis could inform the development of a prognostic algorithm which could be used in a similar fashion to QRISK for cardiovascular prognoses (50,103). However, there are a number of further steps that would need to be taken before this could be utilised in clinical practice. Firstly, no predictive model was tested here; the models of prognosis built in Chapter 3 might be added to with the factors found in Chapters 4 and 5, but their predictive accuracy in a novel dataset and their utility would still need to be tested prospectively. One way of building a predictive prognostic model might be to use techniques which perform non-hypothesis driven 'variable selection', rather than by pre-specifying which variables should be included. A number of machine learning models have this capacity and maybe well suited to the data here, some of these models also have the benefit of being able to test multiple forms of interactions in the data (not just linear ones) (44,52,246). Testing of predictive models built in such a fashion is most robust if it involves cross-validation with data from one or more whole studies left out/withheld from the data used to train the predictive models (52).

There are many potential modelling techniques that could be applied to the data in Dep-GP and little evidence to suggest one type of technique is necessarily superior to the others (247–250). However, an emerging literature is developing on the benefits of running competitions on data used to train machine learning models in order to find the models that give the most accurate predictions when tested on withheld data (247–250). So, one avenue for further research might be to consider the predictive accuracy of machine learning models built using the Dep-GP data. Any such model would also need to be tested for utility and would therefore be best compared both to a simple model including just the baseline depressive symptom severity, and also another model including the variables found to be associated with prognosis here, but without the complex data manipulation and model building

processes that go into making the machine learning models. If it were found that such models were approximately equally accurate, then the most parsimonious model might be best retained in order to help explain how the model makes predictions in a way that is understandable to clinicians and patients. This has been found to be an important step in determining the likely clinical utility of such models (101) and the lack of the ability to do this has often been cited as a criticism of machine learning models (52).

In addition to the above, for most of this thesis I have either considered the association between baseline characteristics related to the severity of a depressed patient's experience of depression or considered the association between a range of baseline characteristics and prognosis independent of symptom severity. These considerations have relied on diagnostic classifications to consider what can be counted as part of depressive symptom severity, what can be counted as part of depressive 'disorder severity' (including symptoms of anxiety or anxiety disorders comorbid to depression), and what sits outside of these two constructs. In addition, considerations of severity have for the most part relied on total scores on symptom measures rather than techniques that seek to utilise the individual symptoms. A recent avenue of research involves methods to consider the role and relationship between individual symptoms which are not limited by diagnoses or the accuracy or inaccuracy of diagnostic classification systems; these techniques have grown out of network theory (251,252). This approach suggests that individual symptoms vary in their interactions and impacts upon each-other, and by considering the inter-relations of symptoms in a network which models their interactions, we can gain a better understanding of the problems experienced by patients, and in so doing, can work towards elucidating better personalised treatment plans (170,253).

A network theory informed approach might allow for a more nuanced and potentially more accurate understanding of prognosis. To illustrate, consider the theoretical example of two depressed patients both with the same score on a measure of symptoms, patient 1 has considerable insomnia symptoms, fatigue, poor concentration, and anhedonia, but patient 2 does not, instead patient 2 experiences low mood, suicidal ideation, feelings of guilt and worthlessness, and loss of libido, which patient 1 does not. Network theory offers an alternative explanation for

psychopathology compared to the “common-cause model” (which is the name given to the idea underlying diagnostic classification systems by proponents of network theory) (110). The proposal is that the symptoms experienced by each patient co-occur not because the patients share the same underlying disorder, but instead they are causally linked. So, e.g. sleep problems might give rise to fatigue and concentration problems which in turn gives rise to anhedonia; and loss of libido might lead to feelings of guilt and worthlessness, which might lead to suicidal ideation and low mood (254). In Chapter 3 the total score on the depressive symptom measure the patients were assessed with (depressive symptom severity) was strongly associated with their prognosis, but using symptom severity alone was less informative for prognosis than considering other related factors (i.e. depressive ‘disorder severity’ factors). So, it might also be the case that the prognostic effect of individual symptoms and their interactions could be used to gain a more accurate picture of prognosis (255). Such information could also be useful when considering treatment selection: for example, persistent sleep related problems are associated with considerably worse odds of remission post-treatment (256), we might therefore expect the pattern of symptoms experienced by patient 1 to respond better to treatments that target sleep problems than to those that do not, but we might reasonably expect that patient 2 would be unlikely to gain much from treatments targeting sleep over and above other symptoms (170). Symptom network models have recently begun to be used to inform predictive models but there is considerable debate about their utility, and to-date they have primarily been considered as explanatory and hypothesis generating models rather than those that can be applied to new data to inform predictions (257). This raises a potentially important question for future research on prognosis for adults with depression: whether symptom networks are able to be used to form useful predictive models and whether or not they are able to deliver more accurate predictions than either classical regression models or machine learning predictive models. This question might be best answered before predictive models using the variables identified in this thesis are tested prospectively with new patients, in order that such models are using the most informative variables with which to make prognostic predictions.

Finally, an important implication of the findings from the studies presented in this thesis relates to a key determination in both clinical practice and research settings

where there is a concern with balancing two competing aims: 1) to maximise information to determine a diagnosis or treatment plan, against 2) speed and efficiency to minimise the burden on the patient/participant and on the clinical/research team. Therefore, where a small number of factors have been identified as being potentially associated with clinically important differences in prognosis, the potential benefit of assessing these individual items (e.g. individual symptoms or markers of severity or individual aspects of social support in addition to depressive 'disorder severity') and their associations with treatment engagement or outcomes, could inform both clinical practice and future research. Among other benefits outlined above, such research might inform hypotheses regarding mechanisms of action.

In order not to overburden systems with additional assessment 'smart' assessments that seek only the essential information required to accurately determine the likely prognosis, are worth further investigation. Methods that use computerized adaptive testing (CAT) might achieve this (258): CAT systems are typically developed using item response theory (IRT) modelling procedures, such that loading of each item on a latent-trait factor assumed to underlie the construct being assessed is used to determine which variables are assessed and in what order (259).

Conclusions

In conclusion, the overall aim of this thesis was to investigate factors associated with prognosis for adults with depression independent of treatment. An individual patient dataset was formed from 12 randomised controlled trials of treatments for depression that recruited adults in primary care and randomised them to a number of different types of treatment. Differences in the severity of depressive symptoms pre-treatment had the largest effect on prognosis and accounting for symptom severity reduced the magnitude of effect of nearly all other factors. However, there were a number of other variables found to be independently associated with prognosis and with attrition. Differences in the baseline levels of some of those factors pre-treatment might be considered to be associated with clinically important differences in outcomes after accounting for treatment and for the severity of depression.

Knowledge of these associations might inform further research into the development of prognostic algorithms, parsimonious assessments including some or all of these factors, and particular patient groups to study in order to investigate mechanisms of change in prognosis independent of treatment. Such information could potentially be used to develop clinical assessment tools improving the efficiency and effectiveness of assessments, and might potentially be used to develop treatment decision support tools. More immediately, the factors identified as associated with prognosis or attrition could be assessed by primary care clinicians, and their potential impact on outcomes could be discussed with depressed patients, with transparency about the lack of certainty about the impact such factors might have on an individual's outcomes.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). American Psychiatric Pub; 2013.
2. World Health Organization. International classification of diseases for mortality and morbidity statistics (11th Revision) [Internet]. 11th ed. Vol. 11, World Health Organization. Geneva: World Health Organization; 2018. 2019 p. Available from: https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/719572464
3. American Psychiatric Association. UPDATE DSM-5 ® Supplement to DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS. American Psychiatric Pub; 2016.
4. Reed GM, First MB, Kogan CS, Hyman SE, Gureje O, Gaebel W, et al. Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. World Psychiatry. 2019;18(1):3–19.
5. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet. 1997 May;349:1436–42.
6. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The Epidemiology of Major Depressive Disorder. JAMA [Internet]. 2003 Jun 18;289(23):3095. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.289.23.3095
7. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry [Internet]. 2016;3(2):171–8. Available from: http://dx.doi.org/10.1016/S2215-0366(15)00505-2
8. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry. 2017;210(2):119–24.
9. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global , regional , and national incidence , prevalence , and years lived with disability for 310 diseases and injuries , 1990 – 2015 : a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388:1545–602.
10. Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne BJ, Polanczyk G, et al. How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. Psychol Med [Internet]. 2010 Jun 1;40(6):899–909. Available from: https://www.cambridge.org/core/product/identifier/S0033291709991036/type/journal_article
11. Burcusa SL, Iacono WG. Risk for recurrence in depression. Clin Psychol Rev [Internet]. 2007;27(8):959–85. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc5&NEWS=N&AN=2007-16751-006
12. Angst J, Kupfer DJ, Rosenbaum JF. Recovery from depression: risk or reality? Acta Psychiatr Scand. 1996;93:413–9.
13. Judd LL. The Clinical Course of Unipolar Major Depressive Disorders. Arch Gen Psychiatry [Internet]. 1997;54(11):989–91. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=9268780
14. Mueller TI, Leon AC, Keller MB, Solomon DA, Endicott J, Coryell W, et al. Recurrence

after recovery from major depressive disorder during 15 years of observational follow-up. Am J Psychiatry. 1999;156(7):1000–6.
15. Mauskopf JA, Simon GE, Kalsekar A, Nimsch C, Dunayevich E, Cameron A. Nonresponse, partial response, and failure to achieve remission: humanistic and cost burden in major depressive disorder. Depress Anxiety. 2009;26:83–97.
16. McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. Paying the price The cost of mental health care in England to 2026. Paying the price: the cost of mental health care in England to 2026. London; 2008.
17. Public Health England. Our priorities for 2013/14. London; 2013.
18. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet. 2007;370(9590):859–77.
19. World Health Organization. Improving health systems and services for mental health. Ment Heal Policy Serv Guid Packag World Heal Organ. 2009;112.
20. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: An analysis from the global burden of disease study 2010. PLoS One. 2015;10(2):1–14.
21. National Collaborating Centre for Mental Health. The NICE guidelines on the treatment and management of depression in adults Clinical Guideline CG90 (Updated ed.). National Institute for Health and Clinical Excellence. London; 2010 Oct.
22. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. JAMA Intern Med. 2016;176(10):1482–91.
23. McManus S, Bebbington P, Jenkins R, Brugha T. Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. McManus S, Bebbington P, Jenkins R, Brugha T, editors. NHS Digital. Leeds: NHS Digital; 2016. 1–405 p.
24. Siu AL, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, et al. Screening for Depression in Adults. JAMA [Internet]. 2016 Jan 26;315(4):380. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.18392
25. NHS BSA. Antidepressant prescribing patterns. Vol. 5, Medicine Today. NHS Business Services Authority; 2018. p. 8–9.
26. Clark DM. Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. Annu Rev Clin Psychol [Internet]. 2018 May 7;14(1):159–83. Available from: http://www.annualreviews.org/doi/10.1146/annurev-clinpsy-050817-084833
27. Commons H of. Cost of Depression in England. London; 2011.
28. Rush AJ, Trivedi MH, Wisniewski SR, Ph D, Nierenberg AA, Stewart JW, et al. Acute and Longer-Term Outcomes in Depressed Outpatients Requiring One or Several Treatment Steps : Am J Ps. 2006;163(11):1905–17.
29. Kessler RC. The potential of predictive analytics to provide clinical decision support in depression treatment planning. Curr Opin Psychiatry. 2018;31(1):32–9.
30. Buckman JEJ, Underwood A, Clarke K, Saunders R, Hollon SD, Fearon P, et al. Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. Clin Psychol Rev [Internet]. 2018 Aug;64(7):13–38. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0272735817302957
31. Cohen ZD, DeRubeis RJ. Treatment Selection in Depression. Annu Rev Clin Psychol [Internet]. 2018 May 7;14(1):209–36. Available from:

http://www.annualreviews.org/doi/10.1146/annurev-clinpsy-050817-084746
32. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet [Internet]. 2018 Apr;391(10128):1357–66. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673617328027
33. Saunders R, Buckman JEJ, Cape J, Fearon P, Leibowitz J, Pilling S. Trajectories of depression and anxiety symptom change during psychological therapy. J Affect Disord [Internet]. 2019 Apr;249:327–35. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0165032718324443
34. Buckman JEJ, Saunders R, Fearon P, Leibowitz J, Pilling S. Attentional control as a predictor of response to psychological treatment for depression and relapse up to 1 year after treatment: A pilot cohort study. Behav Cogn Psychother. 2019;47(3):318–31.
35. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: Rating confidence in estimates of event rates in broad categories of patients. Br Med J. 2015;350(h870):1–8.
36. Hayden JA, Windt DA Van Der, Cartwright JL, Côté P, Bombardier C. Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 2013;158:280–6.
37. Blasi Z Di, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet. 2001;357:757–62.
38. Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT, et al. Multiple recurrences of major depressive disorder. Am J Psychiatry [Internet]. 2000;157(2):229–33. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=10671391
39. Gopinath S, Katon WJ, Russo JE, Ludman EJ. Clinical factors associated with relapse in primary care patients with chronic or recurrent depression. J Affect Disord [Internet]. 2007;101(1–3):57–63. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed8&NEWS=N&AN=2007277742
40. Wang JL, Patten S, Sareen J, Bolton J, Schmitz N, Macqueen G. Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. Depress Anxiety. 2014;31:451–7.
41. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF. Recurrence of major depressive disorder and its predictors in the general population: Results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Psychol Med [Internet]. 2013;43(1):39–48. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2013-02413-004
42. Hardeveld F, Spijker J, De Graaf R, Hendriks SM, Licht CMM, Nolen WA, et al. Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study. Alonso Beshai, Bockting, Brugha, Cooper-Patrick, Costa, De Graaf, Fyer, Gilchrist, Gopinath, Hakkaart-van Roijen, Hansen, Hardeveld, Hovens, Judd, Kanai, Katon, Kaymaz, Kupfer, Lin, Lyketsos, Maj, Mueller, Murray, Penninx, Piek, Piet, Post, Solomon, Spij B, editor. J Affect Disord [Internet]. 2013;147:225–31. Available from:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2013-19360-016
43. Lorenzo-Luaces L, DeRubeis RJ, van Straten A, Tiemens B. A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. J Affect Disord. 2017;213(February):78–85.
44. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry [Internet]. 2016 Mar;3(3):243–50. Available from: https://linkinghub.elsevier.com/retrieve/pii/S221503661500471X
45. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. PLoS One. 2014;9(1):e83875.
46. Weitz E, Kleiboer A, Van Straten A, Hollon SD, Cuijpers P. Individual patient data meta-analysis of combined treatments versus psychotherapy (with or without pill placebo), pharmacotherapy or pill placebo for adult depression: A protocol. BMJ Open. 2017;7(2):1–8.
47. Menchetti M, Bortolotti B, Rucci P, Scocco P, Bombi A, Berardi D. Depression in primary care: Interpersonal counseling vs selective serotonin reuptake inhibitors. The DEPICS Study. A multicenter randomized controlled trial. Rationale and design. BMC Psychiatry [Internet]. 2010;10. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2011-00372-001
48. Cuijpers P, Hollon SD, van Straten A, Bockting C, Berking M, Andersson G. Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. BMJ Open [Internet]. 2013 Jan [cited 2014 Feb 20];3(4). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641456&tool=pmcentrez&rendertype=abstract
49. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine : strategic and economic implications of combining drugs and clinical biomarkers. Nat Rev Neurosci. 2007;6(April):287–93.
50. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ [Internet]. 2007 Jul 21;335(7611):136. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.39261.471806.55
51. Schoemaker JH, Kilian S, Emsley R, Vingerhoets AJJM. Factors associated with placebo response in depression trials: A systematic review of published meta-analyses (1990–2017). Neurol Psychiatry Brain Res. 2018;30(April):12–21.
52. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. Annu Rev Clin Psychol [Internet]. 2018 May 7;14(1):91–118. Available from: http://www.annualreviews.org/doi/10.1146/annurev-clinpsy-032816-045037
53. Harrell FE, Lee KL, Mark DB. Prognostic/Clinical Prediction Models: Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Tutorials Biostat Stat Methods Clin Stud. 2005;1:223–49.

54. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models. Epidemiology [Internet]. 2010 Jan;21(1):128–38. Available from: http://journals.lww.com/00001648-201001000-00022
55. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73.
56. Adibi A, Sadatsafavi M, Ioannidis JPA. Validation and Utility Testing of Clinical Prediction Models. JAMA [Internet]. 2020 Jul 21;324(3):235. Available from: https://jamanetwork.com/journals/jama/fullarticle/2762532
57. Rothwell PM. Subgroup analysis in randomised controlled trials : importance, indications, and interpretation. Lancet. 2005;365:176–86.
58. Ioannidis JPA, Lau J. Uncontrolled Pearls, Controlled Evidence, Meta-Analysis and the Individual Patient. J Clin Epidemiol. 1998;51(8):709–11.
59. Schmid CH, Landa M, Jafar TH, Giatras I, Karim T, Reddy M, et al. Constructing a database of individual clinical trials for longitudinal analysis. Control Clin Trials. 2003;24(3):324–40.
60. Fisher DJ, Copas AJ, Tierney JF, Parmar MKB. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol [Internet]. 2011;64(9):949–67. Available from: http://dx.doi.org/10.1016/j.jclinepi.2010.11.016
61. Lesko CR, Jacobson LP, Althoff KN, Abraham AG, Gange SJ, Moore RD, et al. Collaborative, pooled and harmonized study designs for epidemiologic research: Challenges and opportunities. Int J Epidemiol. 2018;47(2):654–68.
62. Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits From Antidepressants. Arch Gen Psychiatry. 2012;69(6):572–9.
63. Siddique J, Reiter JP, Brincks A, Gibbons RD, Crespi CM, Brown CH. Multiple imputation for harmonizing longitudinal non- commensurate measures in individual participant data meta- analysis. Stat Med. 2015;34(20):3399–414.
64. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments : daft , deluded , or deft approach ? most from particular treatments or other broad approaches used for testing such. BMJ Br Med J. 2017;356(j573):1–6.
65. Monroe SM, Harkness KL. Recurrence in Major Depression: A Conceptual Analysis. Psychol Rev. 2011;118(4):655–74.
66. Noma H, Furukawa TA, Maruo K, Imai H, Shinohara K, Tanaka S, et al. Exploratory analyses of effect modifiers in the antidepressant treatment of major depression: Individual-participant data meta-analysis of 2803 participants in seven placebo-controlled randomized trials. J Affect Disord [Internet]. 2019;250(December 2018):419–24. Available from: https://doi.org/10.1016/j.jad.2019.03.031
67. Marwood L, Wise T, Perkins AM, Cleare AJ. Meta-analyses of the neural mechanisms and predictors of response to psychotherapy in depression and anxiety. Neurosci Biobehav Rev [Internet]. 2018;95(September):61–72. Available from: https://doi.org/10.1016/j.neubiorev.2018.09.022
68. Wang J, Mann F, Lloyd-Evans B, Ma R, Johnson S. Associations between loneliness and perceived social support and outcomes of mental health problems : a systematic

review. BMC Psychiatry. 2018;18(156):1–16.
69. Haq A, Sitzmann A, Goldman M, Maixner D, Mickey B. Response of depression to electroconvulsive therapy: a meta-analysis of clinical predictors. J Clin Psychiatry. 2015;76(10):1374–84.
70. Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. Psychol Bull [Internet]. 2015;141(4):747–68. Available from: papers3://publication/doi/10.1037/bul0000015
71. Sugarman MA, Loree AM, Baltes BB, Grekin ER, Kirsch I. The efficacy of paroxetine and placebo in treating anxiety and depression: A meta-analysis of change on the Hamilton Rating Scales. PLoS One. 2014;9(8).
72. Dodd S, Berk M, Kelin K, Zhang Q, Eriksson E, Deberdt W, et al. Application of the Gradient Boosted method in randomised clinical trials: Participant variables that contribute to depression treatment efficacy of duloxetine, SSRIs or placebo. J Affect Disord [Internet]. 2014;168:284–93. Available from: http://dx.doi.org/10.1016/j.jad.2014.05.014
73. Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. J Affect Disord [Internet]. 2014 Jan;152–154(1):65–75. Available from: http://dx.doi.org/10.1016/j.jad.2013.10.017
74. Dodd S, Berk M, Kelin K, Mancini M, Schacht A. Treatment response for acute depression is not associated with number of previous episodes: Lack of evidence for a clinical staging model for major depressive disorder. J Affect Disord [Internet]. 2013;150(2):344–9. Available from: http://dx.doi.org/10.1016/j.jad.2013.04.016
75. Kampman O, Poutanen O. Can onset and recovery in depression be predicted by temperament? A systematic review and meta-analysis. J Affect Disord [Internet]. 2011;135(1–3):20–7. Available from: http://dx.doi.org/10.1016/j.jad.2010.12.021
76. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant Drug Effects and Depression Severity. JAMA [Internet]. 2010 Jan 6;303(1):47. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2009.1943
77. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ [Internet]. 2013 Feb 26;346(feb26 2):f540–f540. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.f540
78. Weitz ES, Hollon SD, Twisk J, Van Straten A, Huibers MJH, David D, et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. JAMA Psychiatry. 2015;72(11):1102–9.
79. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms. JAMA Psychiatry [Internet]. 2017 Apr 1;74(4):351. Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2017.0044
80. Driessen E, Cuijpers P, Hollon SD, Dekker JJM. Does Pretreatment Severity Moderate the Efficacy of Psychological Treatment of Adult Outpatient Depression ? A Meta-

Analysis. J Consult Clin Psychol. 2010;78(5):668–80.
81. Nakabayashi T, Hara A, Minami H. Impact of demographic factors on the antidepressant effect: A patient-level data analysis from depression trials submitted to the Pharmaceuticals and Medical Devices Agency in Japan. J Psychiatr Res [Internet]. 2018 Mar;98:116–23. Available from: https://doi.org/10.1016/j.jpsychires.2017.12.019
82. Gariepy G, Honkaniemi H, Quesnel-Vallee A. Social support and protection from depression : systematic review of current findings in Western countries. Br J Psychiatry. 2016;209:284–93.
83. Paykel ES. Life events, social support and depression. Acta Psychiatr Scand [Internet]. 1994;89(377, Suppl):50–8. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1994-37529-001
84. Carter GC, Cantrell RA, Zarotsky V, Haynes VS, Phillips G, Alatorre CI, et al. Comprehensive review of factors implicated in the heterogeneity of response in depression. Depress Anxiety. 2012;29(4):340–54.
85. Sockol LE. A systematic review and meta-analysis of interpersonal psychotherapy for perinatal women. J Affect Disord [Internet]. 2018 May;232(August 2017):316–28. Available from: https://doi.org/10.1016/j.jad.2018.01.018
86. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJH. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cogn Behav Ther [Internet]. 2018 Mar 4;47(2):91–106. Available from: https://www.tandfonline.com/doi/full/10.1080/16506073.2017.1420098
87. Cristea IA, Karyotaki E, Hollon SD, Cuijpers P, Gentili C. Biological markers evaluated in randomized trials of psychological treatments for depression: a systematic review and meta-analysis. Neurosci Biobehav Rev [Internet]. 2019 Jun;101(101):32–44. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0149763419300247
88. Ebrahim S, Montoya L, Truong W, Hsu S, Kamal el Din M, Carrasco-Labra A, et al. Effectiveness of Cognitive Behavioral Therapy for Depression in Patients Receiving Disability Benefits: A Systematic Review and Individual Patient Data Meta-Analysis. PLoS One. 2012;7(11).
89. Lin X-Q, Wang P, Cai W-K, Xu G-L, Yang M, Zhou M-D, et al. The Associations Between CYP2D6 Metabolizer Status and Pharmacokinetics and Clinical Outcomes of Venlafaxine: A Systematic Review and Meta-Analysis. Pharmacopsychiatry. 2019;52(05):222–31.
90. Stefan Kloiber, Ripke S, Kohli MA, Reppermund S, Salyakina D, Uher R, et al. Resistance to antidepressant treatment is associated with polymorphisms in the leptin gene, decreased leptin mRNA expression, and decreased leptin serum levels. Eur Neuropsychopharmacol. 2013;23(7):653–62.
91. Morris BH, Bylsma LM, Rottenberg J. Does emotion predict the course of major depressive disorder? A review of prospective studies. Br J Clin Psychol. 2009;48(3):255–73.
92. Iovieno N, Tedeschini E, Ameral VE, Rigatelli M, Papakostas GI. Antidepressants for major depressive disorder in patients with a co-morbid axis-III disorder: A meta-analysis of patient characteristics and placebo response rates in randomized controlled trials. Int Clin Psychopharmacol. 2011;26(2):69–74.
93. Fisher DJ. Two-stage individual participant data meta-analysis and generalized forest

plots. Stata J. 2015;2:369–96.
94. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev. 2018;63(June):80–92.
95. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol [Internet]. 2009 Dec 28;9(1):57. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-9-57
96. Dowrick C, Flach C, Leese M, Chatwin J, Morriss R, Peveler R, et al. Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study. Psychol Med [Internet]. 2011 Jan 29;41(1):141–50. Available from: https://www.cambridge.org/core/product/identifier/S0033291710000437/type/journal_article
97. Perroud N, Uher R, Ng MYM, Guipponi M, Hauser J, Henigsberg N, et al. Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the GENDEP project. Pharmacogenomics J [Internet]. 2012 Feb 28;12(1):68–77. Available from: http://www.nature.com/articles/tpj201070
98. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. Control Clin Trials [Internet]. 2004 Feb;25(1):119–42. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0197245603001120
99. Buckman JEJ, Naismith I, Saunders R, Morrison T, Linke S, Leibowitz J, et al. The Impact of Alcohol Use on Drop-out and Psychological Treatment Outcomes in Improving Access to Psychological Therapies Services: an Audit. Behav Cogn Psychother [Internet]. 2018 Sep 26;46(5):513–27. Available from: https://www.cambridge.org/core/product/identifier/S1352465817000819/type/journal_article
100. Boschloo L, Vogelzangs N, Van Den Brink W, Smit JH, Veltman DJ, Beekman ATF, et al. Alcohol use disorders and the course of depressive and anxiety disorders. Br J Psychiatry. 2012;200(6):476–84.
101. Delgadillo J, Huey D, Bennett H, McMillan D. Case complexity as a guide for psychological treatment selection. J Consult Clin Psychol [Internet]. 2017 Sep;85(9):835–53. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/ccp0000231
102. Hepgul N, King S, Amarasinghe M, Breen G, Grant N, Grey N, et al. Clinical characteristics of patients assessed within an Improving Access to Psychological Therapies (IAPT) service: results from a naturalistic cohort study (Predicting Outcome Following Psychological Therapy; PROMPT). BMC Psychiatry [Internet]. 2016 Dec 27;16(1):52. Available from: http://dx.doi.org/10.1186/s12888-016-0736-6
103. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart [Internet]. 2008 Jan 1;94(1):34–9. Available from: http://heart.bmj.com/cgi/doi/10.1136/hrt.2007.134890
104. NHMRC. Additional levels of evidence and grades for recommendations for developers of guidelines Introduction. Canberra: National Health and Medical Research Council; 2009. 1–23 p.

105. Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. Behav Res Methods. 2018;50(5):2016–34.
106. Debray TPA, Riley RD, Rovers MM, Reitsma B, Moons KGM, Cochrane IMMG. Individual Participant Data (IPD) Meta- analyses of Diagnostic and Prognostic Modeling Studies : Guidance on Their Use. PLoS Med. 2015;1–12.
107. Driessen E, Abbass AA, Barber JP, Connolly Gibbons MB, Dekker JJM, Fokkema M, et al. Which patients benefit specifically from short-term psychodynamic psychotherapy (STPP) for depression? Study protocol of a systematic review and meta-analysis of individual participant data. BMJ Open. 2018;8(2):1–8.
108. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data. Jama. 2015;313(16):1657.
109. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. Am J Psychiatry [Internet]. 2010 Jul;167(7):748–51. Available from: http://psychiatryonline.org/doi/abs/10.1176/appi.ajp.2010.09091379
110. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. J Affect Disord [Internet]. 2015 Feb;172:96–102. Available from: http://dx.doi.org/10.1016/j.jad.2014.10.010
111. Zimmerman M, Ellison W, Young D, Chelminski I, Dalrymple K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr Psychiatry [Internet]. 2015 Jan;56:29–34. Available from: http://dx.doi.org/10.1016/j.comppsych.2014.09.007
112. Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community : a standardized assessment for use by lay interviewers. Psychol Med. 1992;22:465–86.
113. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. International Classification. Geneva: World Health Organization; 1993.
114. Lewis G, Mulligan J, Wiles N, Cowen P, Craddock N, Ikeda M, et al. Polymorphism of the 5-HT transporter and response to antidepressants: randomised controlled trial. Br J Psychiatry [Internet]. 2011 Jun 2;198(6):464–71. Available from: https://www.cambridge.org/core/product/identifier/S000712500025544X/type/journal_article
115. Wing JK, Barbor T, Brugha TS, Burke J, Cooper JE, Giel R, et al. SCAN. Arch Gen Psychiatry [Internet]. 1990 Jun 1;47(6):589. Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archpsyc.1990.01810180089012
116. Williams J, Gibbon M, First M, Spitzer R, Davies M, Borus J, et al. The structured clinical interview for DSM-III-R (SCID): II. Multisite test-retest reliability. Arch Gen Psychiatry. 1992;49(8):630–6.
117. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? BMC Med. 2018;16(1):1–6.
118. Mynors-Wallis LM, Gath DH, Day A, Baker F. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. BMJ Br Med J. 2000;320:26–30.

119. Kendrick T, Peveler R, Longworth L, Baldwin D, Moore M, Win JC, et al. Cost-effectiveness and cost-utility of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine: Randomised controlled trial. Br J Psychiatry. 2006;188(4):337–45.
120. Richards DA, Hill JJ, Gask L, Lovell K, Chew-Graham C, Bower P, et al. Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. BMJ [Internet]. 2013 Aug 19;347(aug19 1):f4913–f4913. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.f4913
121. Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. Lancet [Internet]. 2013 Feb;381(9864):375–84. Available from: http://dx.doi.org/10.1016/S0140-6736(12)61552-9
122. Wiles NJ, Mulligan J, Peters TJ, Cowen PJ, Mason V, Nutt D, et al. Severity of depression and response to antidepressants: GENPOD randomised controlled trial. Br J Psychiatry [Internet]. 2012 Feb 2;200(2):130–6. Available from: https://www.cambridge.org/core/product/identifier/S0007125000257619/type/journal_article
123. Salisbury C, Cathain AO, Edwards L, Thomas C, Gaunt D, Hollinghurst S, et al. Effectiveness of an integrated telehealth service for patients with depression: a pragmatic randomised controlled trial of a complex intervention. The Lancet Psychiatry. 2016;3(6):515–25.
124. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet [Internet]. 2009 Aug;374(9690):628–34. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673609612575
125. Thomas H V, Lewis G, Watson M, Thomas H V, Lewis G, Watson M, et al. Computerised patient-specific guidelines for management of common mental disorders in primary care: a randomised controlled trial. Br J Gen Pract. 2004;54(508):832–7.
126. Kessler DS, MacNeill SJ, Tallon D, Lewis G, Peters TJ, Hollingworth W, et al. Mirtazapine added to SSRIs or SNRIs for treatment resistant depression in primary care: phase III randomised placebo controlled trial (MIR). BMJ [Internet]. 2018 Oct 31;363(k4218):k4218. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.k4218
127. Lewis G, Duffy L, Ades A, Amos R, Araya R, Brabyn S, et al. The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. The Lancet Psychiatry [Internet]. 2019 Nov;6(11):903–14. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2215036619303669
128. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): Large scale pragmatic randomised controlled trial. Br Med J. 2015;351:h5627.
129. Sharp DJ, Chew-graham CA, Tylee A, Lewis G, Howard L, Anderson I, et al. A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the

RESPOND trial. Health Technol Assess (Rockv). 2010;14(43).
130. Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH, et al. Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. BMJ [Internet]. 2012 Jun 6;344(jun06 1):e2758–e2758. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.e2758
131. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. Manual for the Beck Depression Inventory-II. San Antonio, Tx: Psychological Corporation; 1996.
132. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med. 2001 Sep;16(9):606–13.
133. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):361–70.
134. Goldberg D. General Health Questionnaire (GHQ-12) NFER. Nelson Publishing Co. Ltd, Windsor, UK; 1992.
135. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J psychiatry. 1987;150(6):782–6.
136. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder. Arch Intern Med [Internet]. 2006 May 22;166(10):1092. Available from: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.166.10.1092
137. Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, et al. The Health and Lifestyle Survey. Health Promotion Research Trust; 1987.
138. Holmes TH, Rahe RH. The social readjustment rating scale. J Psychosom Res. 1967;11(2):213–8.
139. Piccinelli M, Tessari E, Bortolomasi M, Piasere O, Semenzin M, Garzotto N, et al. Efficacy of the AUDIT as a screening tool for hazardous alcohol intake and related disorders in primary care: a validity study. Br Med J. 1997;314:314–20.
140. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;1727–36.
141. Hout B Van, Janssen MF, Feng Y, Kohlmann T, Busschbach J. Interim Scoring for the EQ-5D-5L : Mapping the EQ-5D-5L to EQ-5D-3L. JVAL [Internet]. 2012;15(5):708–15. Available from: http://dx.doi.org/10.1016/j.jval.2012.02.008
142. MRC. MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies. London: Medical Research Council; 2012.
143. Nelson JC, Delucchi KL, Ph D, Schneider LS. Moderators of Outcome in Late-Life Depression: A Patient-Level Meta-Analysis. Am J Psychiatry. 2013;170(6):651–9.
144. StataCorp LP. Stata Base Reference Manual - Release 15. College Station, Texas: Stata Press; 2017. 1–2987 p.
145. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.
146. Choi SW, Schalet BD, Cook KF, Cella D. Establishing a common metric for depressive symptoms: Linking the BDI-II , CES-D, and PHQ-9 to PROMIS Depression. Psychol Assess. 2014;26(2):513–27.
147. Thomas H V, Lewis G, Watson M, Thomas H V, Lewis G, Watson M, et al. Computerised patient-specific guidelines for management of common mental

disorders in primary care : a randomised controlled trial. 2004;(November):832–7.
148. Salaminios G, Duffy L, Ades A, Araya R, Button KS, Churchill R, et al. A randomised controlled trial assessing the severity and duration of depressive symptoms associated with a clinically significant response to sertraline versus placebo, in people presenting to primary care with depression (PANDA trial): study protocol for. Trials [Internet]. 2017 Dec 24;18(1):496. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/s13063-017-2253-4
149. A.J. R, H.C. K, H.A. S, M. F, M.H. T, E. F, et al. Report by the ACNP Task Force on response and remission in major depressive disorder [Internet]. Vol. 31, Neuropsychopharmacology. A.J. Rush, Department of Psychiatry, University of Texas, Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9086, United States. E-mail: john.rush@utsouthwestern.edu: Nature Publishing Group; 2006. p. 1841–53. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006396876
150. Fischer HF, Rose M. www.common-metrics.org: a web application to estimate scores from different patient-reported outcome measures on a common scale. BMC Med Res Methodol [Internet]. 2016;1–5. Available from: http://dx.doi.org/10.1186/s12874-016-0241-0
151. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Br Med J. 2015;349(g7647):1–25.
152. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol. 2008;8:1–15.
153. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Oxford: Blackwell; 2003.
154. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol [Internet]. 2014 Dec 5;14(1):75. Available from: http://www.biomedcentral.com/1471-2288/14/75%0ARESEARCH
155. RUBIN DB. Inference and missing data. Biometrika [Internet]. 1976;63(3):581–92. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/63.3.581
156. Royston P, White IR. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. J Stat Softw. 2011;45(4):1–20.
157. Horton NJ, Lipsitz SR. Multiple Imputation in Practice: Comparison of Software Packages for Regression Models With Missing Variables. Am Stat. 2001;55(3):244–54.
158. Resche-Rigon M, White IR, Bartlett JW, Peters SAE, Thompson SG. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. Stat Med. 2013;32(28):4890–905.
159. Royston P. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata J. 2009;9(3):466–77.
160. Team RC. A Language and Environment for Statistical Computing. Vol. 201. 2013.
161. Viechtbauer W. Conducting meta-analyses in R with the metafor. J Stat Softw. 2010;36(3):1–48.
162. Cuijpers P, Weitz E, Twisk J, Kuehner C, Cristea I, David D, et al. Gender as predictor and moderator of outcome in cognitive behavior therapy and pharamcotherapy for adult depression: An "individual patient data” meta-analysis. Depress Anxiety.

2014;31:941–51.
163. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J. 2003;327:557–60.
164. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Higgins JPT, Geen S, editors. Chichester: The Cochrane Collaboration; 2011.
165. Guyatt GH, Oxman A, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J. 2008;336(7650):924–6.
166. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med [Internet]. 2014 Oct 14;11(10):e1001744. Available from: https://dx.plos.org/10.1371/journal.pmed.1001744
167. Mynors-wallis LM, Gath DH, Day A, Baker F. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. Br J Psychiatry. 2000;320:26–30.
168. Buckman JEJ, Saunders R, Cohen ZD, Clarke K, Ambler G, DeRubeis RJ, et al. What factors indicate prognosis for adults with depression in primary care? A protocol for meta-analyses of individual patient data using the Dep-GP database. Wellcome Open Res. 2019;4(69):1–19.
169. Buckman JEJ, Saunders R, Cohen ZD, Clarke K, Ambler G, Derubeis RJ, et al. Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. BMJ Br Med J.
170. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One. 2014;9(2).
171. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, et al. Minimal clinically important difference on the Beck Depression Inventory - II according to the patient's perspective. Psychol Med [Internet]. 2015 Nov 13;45(15):3269–79. Available from: https://www.cambridge.org/core/product/identifier/S0033291715001270/type/journal_article
172. White IR. Multivariate Random-effects Meta-regression: Updates to Mvmeta. Stata J Promot Commun Stat Stata [Internet]. 2011 Jul 19;11(2):255–70. Available from: http://journals.sagepub.com/doi/10.1177/1536867X1101100206
173. Delgadillo J, Moreea O, Lutz W. Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. Behav Res Ther. 2016;79:15–22.
174. Saunders R, Cape J, Fearon P, Pilling S. Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. J Affect Disord. 2016;197:107–15.
175. Saunders R, Buckman JEJ, Pilling S. Latent variable mixture modelling and individual treatment prediction. Behav Res Ther [Internet]. 2020 Jan;124(October 2019):103505. Available from: https://doi.org/10.1016/j.brat.2019.103505
176. McCambridge J, Day M. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction. 2008;103(2):241–8.
177. Roiser JP, Elliott R, Sahakian BJ. Cognitive Mechanisms of Treatment in Depression.

Neuropsychopharmacology [Internet]. 2012;37(1):117–36. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3238070&tool=pmcentrez&rendertype=abstract
178. Lutz W, Schwartz B, Hofmann SG, Fisher AJ, Husen K, Rubel JA. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. Sci Rep [Internet]. 2018 Dec 18;8(1):7819. Available from: http://www.nature.com/articles/s41598-018-25953-0
179. Uchino BN. Social support and health: A review of physiological processes potentially underlying links to disease outcomes. J Behav Med. 2006;29(4):377–87.
180. Thoits PA. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav. 2011;52(2):145–61.
181. Krause N, Jay G. Stress, social support, and negative interaction in later life.pdf. Res Aging. 1991;13(3):333–63.
182. Berkman LF, Blumenthal J, Burg M, Carney RM, Catellier D, Cowan MJ, et al. Effects of treating depression and low perceived social support on clinical events after myocardial infarction: The Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) randomized trial. [Internet]. Arntz Beck, Beck, Berger, Berkman, Berkman, Blumenthal, Brummet, Bush, Carney, Carney, Case, Cox, Crow, D'Hoore, Dobson, Dusseldorp, Forrester, Frasure-Smith, Frasure-Smith, Frasure-Smith, Frasure-Smith, Freedland, Friedman, Glassman, Holohan, Ibrahim, J B, editor. Vol. 289, JAMA: Journal of the American Medical Association. Czajkowski, Susan M.: National Heart, Lung, & Blood Inst, Rockledge Ctr II, 6701 Rockledge Dr, Room 8114, Bethesda, MD, US, 20892, czajkows@nhlbi.nih.gov: American Medical Assn; 2003. p. 3106–16. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2003-05625-004
183. Brown GW, Andrews B, Harris T, Adler Z, Bridge L. Social support , self-esteem and depression. Psychol Med. 1986;16:813–31.
184. Nima A Al, Rosenberg P, Archer T, Garcia D. Anxiety, Affect, Self-Esteem, and Stress: Mediation and Moderation Effects on Depression. PLoS One. 2013;8(9).
185. Bandura A. Social cognitive theory : An agentic perspective. Asian J Soc Psychol. 1999;2:21–41.
186. Fancourt D, Tymoszuk U. Cultural engagement and incident depression in older adults : evidence from the English Longitudinal Study of Ageing. Br J Psychiatry. 2019;214:225–9.
187. Fancourt D, Perkins R, Ascenso S, Carvalho LA, Steptoe A. Effects of Group Drumming Interventions on Anxiety , Depression , Social Resilience and Inflammatory Immune Response among Mental Health Service Users. PLoS One. 2016;1–16.
188. Kinnier RT, Hofsess C, Pongratz R, Lambert C. Attributions and affirmations for overcoming anxiety and depression. Psychol Psychother Theory, Res Pract. 2009 Jun;82(2):153–69.
189. Smith L, Hill N, Kokanovic R. Experiences of depression, the role of social support and its impact on health outcomes. J Ment Heal. 2015;24(6):342–6.
190. Umberson D, Montez JK. Social Relationships and Health : A Flashpoint for Health Policy. J Health Soc Behav. 2010;51(S):54–66.
191. Haro JM, Ayuso-mateos JL, Bitter I, Demotes-Mainard J, Leboyer M, Lewis SW, et al. ROAMER : roadmap for mental health research in Europe. Int J Methods Psychiatr

Res. 2014 Jan;23(Suppl. 1):1–14.
192. Cruwys T, Dingle GA, Haslam C, Haslam SA, Jetten J, Morton TA. Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. Soc Sci Med. 2013;98:179–86.
193. Holvast F, Burger H, De Waal MMW, Van Marwijk HWJ, Comijs HC, Verhaak PFM. Loneliness is associated with poor prognosis in late-life depression: Longitudinal analysis of the Netherlands study of depression in older persons. J Affect Disord [Internet]. 2015;185:1–7. Available from: http://dx.doi.org/10.1016/j.jad.2015.06.036
194. Hallgren M, Lundin A, Tee FY, Burström B, Forsell Y. Somebody to lean on: Social relationships predict post-treatment depression severity in adults. Psychiatry Res [Internet]. 2017 Mar;249(November 2016):261–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0165178116314354
195. Sherboune CD, Stewart AL. The MOS Social Support Survey. Soc Sci Med. 1991;32(6):705–14.
196. Chalmers RP. mirt : A Multidimensional Item Response Theory Package for the R Environment. J Stat Softw [Internet]. 2012;48(6). Available from: http://www.jstatsoft.org/v48/i06/
197. Bock RD, Aitkin M. EM Solution of the Marginal Likelihood Equations. Psychometrika. 1981;46(4):443–59.
198. Swinscow TD V. “Statistics at square one.” 9th ed. British Medical Journal. London: BMJ Publishing Group; 1997.
199. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1(1):43–6.
200. Brugha TS, Weich S, Singleton N, Lewis G, Bebbington PE, Jenkins R, et al. Primary group size, social support, gender and future mental health status in a prospective study of people living in private households throughout Great Britain. Psychol Med. 2005;35(5):705–14.
201. Dimatteo MR. Social Support and Patient Adherence to Medical Treatment : A Meta-Analysis. Heal Psychol. 2004;23(2):207–18.
202. Akdağ EM, Kotan VO, Kose S, Tıkır B, Aydemir Ç, Okay İT, et al. The relationship between internalized stigma and treatment motivation , perceived social support , depression and anxiety levels in opioid use disorder. Psychiatry Clin Psychopharmacol. 2018;28(4):394–401.
203. Hammen C. Risk Factors for Depression: An Autobiographical Review. Ssrn. 2018;(December).
204. Monroe SM, Harkness KL. Life Stress, the “Kindling” Hypothesis, and the Recurrence of Depression: Considerations From a Life Stress Perspective. Psychol Rev [Internet]. 2005;112(2):417–45. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=15783292
205. Kendler KS, Gardner CO. Depressive vulnerability, stressful life events and episode onset of major depression: a longitudinal model. Psychol Med. 2016;46(9):1865–74.
206. Ormel J, Oldehinkel AJ, Brilman EI. The interplay and etiological continuity of neuroticism , difficulties , and li ... Am J Psychiatry. 2001;158(6):885–91.
207. Dowrick C, Shiels C, Page H, Ayuso-Mateos JL, Casey P, Dalgard OS, et al. Predicting long-term recovery from depression in community settings in Western Europe: Evidence from ODIN. Soc Psychiatry Psychiatr Epidemiol. 2011;46(2):119–26.

208. Bock C, Bukh JD, Vinberg M, Gether U, Kessing LV. Do stressful life events predict medical treatment outcome in first episode of depression? Soc Psychiatry Psychiatr Epidemiol. 2009;44(9):752–60.
209. Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, Gallop R. Prediction of Response to Medication and Cognitive Therapy in the Treatment of Moderate to Severe Depression. J Consult Clin Psychol. 2009;77(4):775–87.
210. Community and Mental Health team ND. Psychological Therapies, Annual Report on the use of IAPT services: England 2015-16. Leeds: NHS Digital; 2016.
211. Community and Mental Health team ND. Psychological Therapies, Annual report on the use of IAPT services 2018-19. Leeds; 2019.
212. Saunders R, Buckman JEJ, Stott J, Aguirre E, John A, Leibowitz J, et al. Older adults do better than younger adults in psychological therapy: evidence from the IAPT programme. Psychother Psychosom.
213. Stansfield S, Clark C, Bebbington P, King M, Jenkins R, Hinchliffe S. Chapter 2: Common mental disorders. In: Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. Leeds: NHS Digital; 2016.
214. Bailey RK, Mokonogho J, Kumar A. Racial and ethnic differences in depression: Current perspectives. Neuropsychiatr Dis Treat. 2019;15:603–9.
215. Pieterse AL, Todd NR, Neville HA, Carter RT. Perceived racism and mental health among black american adults: A meta-analytic review. J Couns Psychol. 2012;59(1):1–9.
216. Williams DR, Mohammed SA. Discrimination and racial disparities in health: Evidence and needed research. J Behav Med. 2009;32(1):20–47.
217. Marmot M, Bell R. Fair society, healthy lives (Full report). Public Health. 2012;126(SUPPL.1):S4–10.
218. Teo AR, Choi H, Valenstein M. Social Relationships and Depression: Ten-Year Follow-Up from a Nationally Representative Study. Coyne J, editor. PLoS One [Internet]. 2013 Apr 30;8(4):e62396. Available from: https://dx.plos.org/10.1371/journal.pone.0062396
219. Ivan Santini Z, Koyanagi A, Tyrovolas S, Haro JM. The association of relationship quality and social networks with depression, anxiety, and suicidal ideation among older married adults: Findings from a cross-sectional analysis of the Irish Longitudinal Study on Ageing (TILDA). J Affect Disord. 2015;179:134–41.
220. Holt-Lunstad J, Birmingham W, Jones BQ. Is there something unique about marriage? The relative impact of marital status, relationship quality, and network social support on ambulatory blood pressure and mental health. Ann Behav Med. 2008;35(2):239–44.
221. Akincigil A, Munch S, Niemczyk KC. Predictors of maternal depression in the first year postpartum: Marital status and mediating role of relationship quality. Soc Work Health Care. 2010;49(3):227–44.
222. Costello CG. Social factors associated with depression: A retrospective community study. Psychol Med. 1982;12(2):329–39.
223. Coyne JC, Thompson R, Palmer SC. Marital quality, coping with conflict, marital complaints, and affection in couples with a depressed wife. J Fam Psychol. 2002;16(1):26–37.
224. Trivedi MH, Morris DW, Pan J-Y, Grannemann BD, Rush AJ. What moderator characteristics are associated with better prognosis for depression? Neuropsychiatr

Dis Treat. 2005;1(1):51–7.
225. Sherbourne C, Schoenbaum M, Wells KB, Croghan TW. Characteristics, treatment patterns, and outcomes of persistent depression despite treatment in primary care. Gen Hosp Psychiatry. 2004;26(2):106–14.
226. Van Der Lem R, Stamsnieder PM, Van Der Wee NJA, Van Veen T, Zitman FG. Influence of sociodemographic and socioeconomic features on treatment outcome in RCTs versus daily psychiatric practice. Soc Psychiatry Psychiatr Epidemiol. 2013;48(6):975–84.
227. Arnow BA, Blasey C, Manber R, Constantino MJ, Markowitz JC, Klein DN, et al. Dropouts versus completers among chronically depressed outpatients. J Affect Disord. 2007;97(1–3):197–202.
228. Evans GW, Wells NM, Chan HYE, Saltzmand H. Housing quality and mental health. J Consult Clin Psychol. 2000;68(3):526–30.
229. Eaton WW, Shao H, Nestadt G, Lee BH, Bienvenu OJ, Zandi P. Population-based study of first onset and chronicity in major depressive disorder. [Internet]. Angst Berkson, Caspi, Chipman, Cohen, Dickel, Eaton, Eaton, Eaton, Eaton, Eaton, Eaton, Essau, Feinn, Frank, Furlong, Goldman, Grunblatt, Heils, Hu, Hu, Ishiguro, Itokawa, Judd, Keller, Keller, Kim, Lawless, Lee, Lesch, Levinson, Li, Liang, Lin, Lyketsos A, editor. Vol. 65, Archives of General Psychiatry. Eaton, William W.: Department of Mental Health, 624 N Broadway, Baltimore, MD, US, 21205, weaton@jhsph.edu: American Medical Assn; 2008. p. 513–20. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc5&NEWS=N&AN=2008-18338-004
230. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med. 2016;46(13):2679–93.
231. Paykel ES. Life events, Social support and depression. Acta Psychiatr Scand. 1994;89:50–8.
232. Ioannidis JPA. Why most published research findings are false. Jantsch W, Schaffler F, editors. PLoS Med [Internet]. 2005;2(8):e124. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16060722
233. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The Extent and Consequences of P-Hacking in Science. PLoS Biol. 2015;13(3):1–15.
234. Maslow AH. A Theory of Human Motivation. Psychol Rev. 1943;50(4):370–96.
235. Memon A, Taylor K, Mohebati LM, Sundin J, Cooper M, Scanlon T, et al. Perceived barriers to accessing mental health services among black and minority ethnic (BME) communities: A qualitative study in Southeast England. BMJ Open. 2016;6(11):1–9.
236. Dickens C, Coventry P, Khara A, Bower P, Mansell W, Bakerly ND. Perseverative negative cognitive processes are associated with depression in people with long-term conditions. Chronic Illn [Internet]. 2012 Jun 14;8(2):102–11. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2012-12771-003
237. Finegan M, Firth N, Delgadillo J. Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: the role of area-level income and crime. Psychother Res [Internet]. 2020 May 18;30(4):546–54. Available from: https://doi.org/10.1080/10503307.2019.1649500

238. Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, et al. The Healthy Activity Program (HAP), a lay counsellor- delivered brief psychological treatment for severe depression , in primary care in India : a randomised controlled trial. Lancet. 2016;6736(16):1–10.
239. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62:617–27.
240. Keefe JR, Wiltsey Stirman S, Cohen ZD, DeRubeis RJ, Smith BN, Resick PA. In rape trauma PTSD, patient characteristics indicate which trauma-focused treatment they are most likely to complete. Depress Anxiety. 2018;35(4):330–8.
241. Brugha TS, Bebbington PE, Jenkins R. A difference that matters : comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychol Med. 1999;29:1013–20.
242. Wahl I, Lowe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement : a common metric was developed for 11 self-report depression measures. J Clin Epidemiol. 2014;67:73–86.
243. Fischer HF, Rose M. www.common-metrics.org: a web application to estimate scores from different patient-reported outcome measures on a common scale. BMC Med Res Methodol. 2016;16(142):1–5.
244. Delgadillo J, de Jong K, Lucock M, Lutz W, Rubel J, Gilbody S, et al. Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. The Lancet Psychiatry. 2018;5(7):564–72.
245. Segal Z V, Shaw BF, Vella DD. Life stress and depression: A test of the congruency hypothesis for life event content and depressive subtype. Can J Behav Sci / Rev Can des Sci du Comport [Internet]. 1989 Oct;21(4):389–400. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1990-09899-001
246. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. AMIA Annu Symp Proc. 2013;2013:1109–15.
247. Webb CA, Cohen ZD, Beard C, Forgeard M, Peckham AD, Björgvinsson T. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. J Consult Clin Psychol [Internet]. 2020 Jan;88(1):25–38. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/ccp0000451
248. Rosellini AJ, Dussaillant F, Zubizarreta JR, Kessler RC, Rose S. Predicting posttraumatic stress disorder following a natural disaster. J Psychiatr Res. 2018;96:15–22.
249. Cohen ZD, Wiley J, Lutz W, Fisher AJ, Kim T, Saunders R, et al. SMART Mental Health Prediction Tournament - Protocol [Internet]. Open Science Framework. 2017. Available from: https://osf.io/wxgzu/%0D
250. Buckman JEJ, Cohen Z, Delgadillo J, Derubeis R, Consortium T. Results of the Stratified Medicine Approaches foR Treatment Selection (SMART) Mental Health Prediction Tournament. In: MQ Mental Health Science Meeting. London; 2019.
251. Borsboom D. A network theory of mental disorders. World Psychiatry [Internet]. 2017 Feb;16(1):5–13. Available from: http://doi.wiley.com/10.1002/wps.20375
252. Cramer AOJ, Van Borkulo CD, Giltay EJ, Van Der Maas HLJ, Kendler KS, Scheffer M, et al. Major depression as a complex dynamic system. PLoS One. 2016;11(12):1–20.

253. Bos FM, Fried EI, Hollon SD, Bringmann LF, Dimidjian S. Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. Soc Psychiatry Psychiatr Epidemiol [Internet]. 2018;0(0):0. Available from: http://dx.doi.org/10.1007/s00127-018-1506-1
254. Fried EI, Cramer AOJ. Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. Perspect Psychol Sci [Internet]. 2017 Nov 5;12(6):999–1020. Available from: http://journals.sagepub.com/doi/10.1177/1745691617705892
255. Boschloo L, van Borkulo CD, Borsboom D, Schoevers RA. A Prospective Study on How Symptoms in a Network Predict the Onset of Depression. Psychother Psychosom [Internet]. 2016;85(3):183–4. Available from: https://www.karger.com/Article/FullText/442001
256. Pigeon WR, Hegel M, Unützer J, Fan MY, Sateia MJ, Lyness JM, et al. Is insomnia a perpetuating factor for late-life depression in the IMPACT cohort? Sleep. 2008;31(4):481–8.
257. Jones PJ, Mair P, Riemann BC, Mugno BL, McNally RJ. A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. J Anxiety Disord. 2018;53(March 2017):1–8.
258. Gibbons RD, Hooker G, Finkelman MD, Weiss DJ, Pilkonis PA, Frank E, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): A screening tool for depression. J Clin Psychiatry. 2013;74(7):669–74.
259. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). Qual Life Res. 2005;14(10):2277–91.

Appendices

Appendix 1. Details of search terms and search results

Table 1. Bibliographic database searches and results for literature review on indicators of prognosis for adults with depression

Searches	Results
Cochrane Database of Systematic Reviews (last searched on 27th February 2020)	
1. (depression or MDD or Major Depression or depressive episode): ti.ab.kw (Word variations have been searched)	656
2. AND (Prognosis or Outcome): ti.ab.kw (Word variations have been searched)	574
3. AND (systematic review or meta-analysis or meta analysis): ti,ab,kw (Word variations have been searched)	350
4. NOT (psychosis or bipolar or bi-polar)	330
5. Limit to Topic "Mental Health"	**131**
Prospero (last searched on 27th February 2020)	
1. (depression or Depressive disorder or Major depression or Unipolar depression or MDD)	5785
2. Filter in Health area of review "Mental health and behavioural conditions, or Public health (including social determinants of health)	2302
3. Filter in Type and method of review "Epidemiologic, Prognostic, Systematic Review, Meta-analysis, Individual patient data (IPD) Meta-analysis, Network meta-analysis, Review of reviews, or Qualitative synthesis"	2254
4. Filter in Status of Review "Published"	**136**
Embase searched 1974 to 2020 February 28	
1. (Major depression or MDD or Major Depressive Disorder).m_titl.	18745
2. (minor depression or MinD).m_titl	13201
3. (depressive or depressive episode or depressive disorder).m.titl	39836
4. Depression.m_titl.	125022
5. 1 or 2 or 3 or 4	176324
6. treatment outcome/	839098
7. treatment response.mp. or treatment response/	271731
8. prognosis.mp. or prognostic.m[p. or prognostic assessment/	4939326
9. moderator	8831
10. systematic review.mp. or "systematic review"/ or meta analysis/	385120
11. 2 or 3 or 4 or 5	1477081
12. 1 and 6 and 7	1251
13. (children or adolescent or child).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	3167378
14. 12 NOT 13	1149
15. (old age or geriatric).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	142854
16. 14 NOT 15	1106
17. bipolar disorder/ or bipolar depression/ or bipolar.mp. or psychosis.mp. or psychotic.mp. or schizoaffective.mp. or schizophrenia.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word	355747
18. 16 NOT 17	962
19. (stroke or dementia or parkinson* disease or brain injury).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	893303
20. 18 NOT 19	900
21. Limit 20 to (human and English and journal)	**864**
Ovid MEDLINE 1946 to March Week 3 2019 (last searched on 27th February 2020)	
1. exp major depression/ or exp "depression (emotion)"/ or exp Depressive Disorder/	115136
2. prognosis.mp.	726849
3. exp Treatment outcome/	1028415
4. 2 or 3	1656920
5. systematic review.mp. or "Systematic Review"/	169051

6. meta-analysis.mp. or Meta-Analysis/	179131
7. 5 or 6	271962
8. 1 and 4 and 7	578
9. limit 8 to (English language and humans and "reviews (maximizes specificity)")	**513**

Table 2. Bibliographic database searches and results for randomised controlled trials to consider for IPD.

Searches	Results
Cochrane CENTRAL Trial Register (searched on 20th March 2019)	
1. ("Depression" or "MDD" or "Unipolar" or "Depressive"):ti,ab,kw (Word variations have been searched)	61130
2. ("RCT" or "controlled trial" or "randomized controlled trial" or "clinical trial"):ti,ab,kw (Word variations have been searched)	510513
3. ("CIS-R" or "Clinical Interview Schedule" or "Revised Clinical Interview Schedule" or "Clinical Interview Schedule Revised"):ti,ab,kw (Word variations have been searched)	35
4. #1 and #2 and #3	**28**
Embase 1947 to 2019 Week 12	
1. (depression or Depressive disorder or Major depression or Unipolar depression or MDD).mp.	615565
2. exp controlled clinical trial/ or exp "randomized controlled trial (topic)"/ or exp "clinical trial"/	1519513
3. ("Clinical Interview Schedule" or "CIS-R" or "CISR" or "Revised clinical interview schedule" or "clinical interview schedule revised").af.	732
4. 1 and 2 and 3	**27**
International Pharmaceutical Abstracts 1970 to March 2019	
1. (depression or Depressive disorder or Major depression or Unipolar depression or MDD).mp.	9677
2. (RCT or controlled trial or randomized controlled trial or clinical trial).mp.	14242
3. ("Clinical Interview Schedule" or "CIS-R" or "CISR" or "Revised clinical interview schedule" or "clinical interview schedule revised").af.	3
4. 1 and 2 and 3	**1**
Ovid MEDLINE 1946 to March Week 3 2019	
1. exp major depression/ or exp "depression (emotion)"/	107371
2. exp Depressive Disorder, Major/	27357
3. exp Depressive Disorder, Major/ or exp Depressive Disorder/ or exp Depression/	199532
4. 1 or 2 or 3	199532
5. exp controlled clinical trial/ or exp "randomized controlled trial (topic)"/	566718
6. ("Clinical Interview Schedule" or "CIS-R" or "CISR" or "Revised clinical interview schedule" or "clinical interview schedule revised").af.	584
7. 4 and 5 and 6	**20**
PsycINFO 1806 to March Week 3 2019	
1. exp major depression/ or exp "depression (emotion)"/	145063
2. (depression or Depressive disorder or Major depression or Unipolar depression or MDD).mp.	303299
3. 1 or 2	303598
4. exp "randomized controlled trial (topic)"/ or exp "clinical trial"/ or exp "controlled trial"/ or exp "randomized clinical trial"/	11270
5. (RCT or controlled trial or randomized controlled trial or clinical trial).mp.	37820
6. 4 or 5	43855
7. ("Clinical Interview Schedule" or "CIS-R" or "CISR" or "Revised clinical interview schedule" or "clinical interview schedule revised").af.	1110
8. 3 and 6 and 7	**49**

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist:

Section and topic	Item No	Checklist item	(Page No.#)
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	1
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	1
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	1
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	1
Support:			
Sources	5a	Indicate sources of financial or other support for the review	13
Sponsor	5b	Provide name for the review funder and/or sponsor	
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	3-4
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	4
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	5
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	5
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	5 and Appendix A
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	5
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	5
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	7
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	9-10
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	8
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	12
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	10-11

	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	12-13
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	12

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

Appendix 2. Distribution and degree of missing data for each variable in each study of Dep-GP

Appendix 2 Table 1.

		Study															
Prognostic Indicator	**Categories**	**AHEAD**				**CADET**				**COBALT**				**GENPOD**			
		N	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**
Sample size at baseline		**327**				**527**				**469**				**601**			
Baseline Characteristics																	
Depressive Symptom Severity																	
Z-score across measures		324	-0.2	0.97	3	523	0.4	0.91	4	469	0.13	1.01	0	601	0.31	0.92	0
PHQ-9 score					327	523	17.7	5.12	4	469	16.59	5.67	0				601
BDI-II Score					327				527	469	31.79	10.66	0	601	33.67	9.67	0
HADS Depression Score		324	10.5	3.87	3				527				469	601	12.22	3.77	0
GHQ Score					327				527				469				601
EPDS Score					327				527				469				601
PROMIS T-Score		324	66.1	7.69	3	527	73.1	6.43	0	469	71.71	7.19	0	601	74.16	6.63	0
Depressive Disorder Severity																	
CIS-R Total Score		327	23.7	10.61	0	527	29.3	9.15	0	469	30.08	8.91	0	601	30.82	8.00	0
CIS-R Depression Scales Sum Score		327	11.7	4.82	0	527	14.9	3.06	0	469	14.87	3.00	0	601	15.05	2.91	0
CIS-R Anxiety Scales Sum Score		324	12.0	7.01	3	527	14.4	7.02	0	469	15.20	6.83	0	601	15.77	6.17	0
Depression Duration					327	527	3.7	1.19	0	469	4.26	1.03	0	601	2.88	1.09	0
Average Anxiety Duration					327	527	2.3	1.01	0	469	2.72	1.04	0	601	1.92	0.83	0
Number of Comorbid Disorders					327	527	2.3	1.08	0	469	2.40	1.09	0	601	2.39	0.92	0
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Agoraphobia	No				327	446	84.6		0	408	87		0	526	87.5		0
	Yes					81	15.4			61	13			75	12.5		
CFS	No				327	142	26.9		0	126	26.9		0	125	20.8		0
	Yes					385	73.1			343	73.1			476	79.2		
GAD	No				327	209	39.7		0	157	33.5		0	191	31.8		0
	Yes					318	60.3			312	66.5			410	68.2		
MADD	No				327	388	73.6		0	362	77.2		0	479	79.7		0
	Yes					139	26.4			107	22.8			122	20.3		
OCD	No				327	449	85.2		0	390	83.2		0	487	81		0
	Yes					78	14.8			79	16.8			114	19		
Panic Disorder	No				327	453	86.0		0	402	85.7		0	550	91.5		0
	Yes					74	14.0			67	14.3			51	8.5		
Social Phobia	No				327	479	90.9		0	405	86.4		0	537	89.4		0
	Yes					48	9.1			64	13.6			64	10.6		
Specific Phobia	No				327	440	83.5		0	378	80.6		0	474	78.9		0
	Yes					87	16.5			91	19.4			127	21.1		
History of Depression	No	173	56.7		22	148	28.1		0	54	11.5		0	167	27.8		0

	Yes	132	43.3			379	71.9			415	88.5			434	72.2		
History of ADM treatment	No	173	56.7		22	195	37.3		4	0	0		0	276	45.9		0
	Yes	132	43.3			328	62.7			469	100			325	54.1		
History of Any Past Treatment	No	173	56.7		22	195	37.3		4	0	0		0	248	41.3		0
	Yes	132	43.3			328	62.7			469	100			353	58.7		
Social Support		**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**
Social Support Total Score		0			327	0			527	469	20.00	3.79	0	600	20.02	3.80	1
Accepted		0			327	0			527	469	2.53	0.60	0	600	2.55	0.60	1
Cared about		0			327	0			527	469	2.73	0.49	0	600	2.74	0.48	1
Can rely on others		0			327	0			527	469	2.56	0.61	0	600	2.57	0.61	1
Can talk to others		0			327	0			527	469	2.27	0.72	0	600	2.34	0.70	1
Encouraged		0			327	0			527	469	2.46	0.61	0	600	2.49	0.60	1
Made to feel Happy		0			327	0			527	469	2.39	0.66	0	600	2.36	0.65	1
Made to feel Important		0			327	0			527	469	2.48	0.65	0	600	2.42	0.65	1
Made to feel Loved		0			327	0			527	469	2.59	0.57	0	600	2.55	0.59	1
Life Events																	
Life Events Total Score		0			327	0			527	469	1.27	1.15	0	601	1.68	1.37	0
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Arguments/Disputed	No	0			327	0			527	362	77	0.42	0	413	69	0.46	1
	Yes									107	23			187	31		
Bereavement	No	0			327	0			527	363	77		0	489	82		1
	Yes									106	23			111	19		
Debt	No	0			327	0			527	327	70		0	343	57		1
	Yes									142	30			257	43		
Divorce	No	0			327	0			527	436	93		0	512	85		1
	Yes									33	7			88	15		
Illness or Injury	No	0			327	0			527	341	73		0	426	71		1
	Yes									128	27			174	29		
Legal Trouble/Court case	No	0			327	0			527	436	93		0	531	89		1
	Yes									33	7			69	12		
Lost job/Sacked	No	0			327	0			527	451	96		0	547	91		1
	Yes									18	4			53	9		
Victim of Crime/Assaulted	No	0			327	0			527	440	94		0	532	89		1
	Yes									29	6			68	11		
Socio-demographics		**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Age		327	43.1	15.43	0	527	44.4	13.17	0	469	49.59	11.70	0	601	38.82	12.35	0
Gender	Female	219	67.0		0	378	71.7		0	339	72.3		130	408	67.9		193
	Male	108	33.0			149	28.3			130	27.7		339	193	32.1		408
Ethnicity	white				327	453	86.0		0	459	97.9		10	575	95.7		26
	non-white					74	14.0			10	2.1		459	26	4.3		575
	Degree or higher				327	98	18.6		0	95	20.5		6				601

Highest Level of Educational Attainment	A-level or Diplomas					147	27.9			122	26.3					
	GCSE					129	24.5			130	28.1					
	None or Other					153	29.0			116	25.1					
Employment Status	Employed				327	236	45.0		2	206	43.9		0	357	59.4		0
	Not seeking employment					120	22.9			112	23.9			121	20.1		
	Unemployed					169	32.2			151	32.2			123	20.5		
Financial Wellbeing	Doing OK financially				327				527	167	35.6		0	210	35.1		2
	Just about getting by									174	37.1			176	29.4		
	Struggling financially									128	27.3			213	35.6		
Housing Tenancy Status	Home owner				327	216	41.0		0				469	289	48.1		0
	Tenant					256	48.6							232	38.6		
	Other					55	10.4							80	13.3		
Marital Status	Married/cohabiting				327	218	41.4		0	248	52.9		0	316	52.6		0
	Single					159	30.2			89	19			175	29.1		
	No longer married					150	28.5			132	28.1			110	18.3		
Others																	
Audit PC Total Score		0			327	0			527	469	1.56	2.27	0	596	1.91	2.11	5
EQ5D Cross-walked Index Score		320	0.8	0.09	7	526	0.7	0.09	1	0			469	0			601
LTC	No				327	321	60.9		0				469	454	75.5		0
	Yes					206	39.1							147	24.5		
Outcomes																	
Z-score across measures 3-4 months		224	0.0	0.91	103	490	0.2	1.07	37	440	0.36	0.93	29	486	-0.07	0.95	115
PROMIS T-Score 3-4 months		224	56.6	10.76	103	491	64.2	11.19	36	440	66.25	8.72	29	486	57.86	12.13	115
Z-score across measures 6-8 months		177	-0.4	0.82	150	0			527	419	0.48	1.16	50	0	.	.	601
PROMIS T-Score 6-8 months		152	51.7	10.82	175	0			527	0			469	486	57.86	12.13	115
Z-score across measures 9-12 months		169	-0.5	0.85	158	497	0.1	1.05	30	0			469	0			601
PROMIS T-Score 9-12 months		153	49.3	13.25	174	497	62.9	11.17	30	395	62.06	14.29	74	0			601
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Remission at 3-4 months	No	86	38.4		103	289	59		37	289	65.7		29	289	59.5		115
	Yes	138	61.6			201	41			151	34.3			197	40.5		
Attrition at 3-4 months	No	243	74.3		0	508	96.4		0	450	95.9		469	486	80.9		601
	Yes	84	25.7			19	3.6			19	4.1			115	19.1		

Appendix 2 Table 1. Continued (1).

		Study															
Prognostic Indicator	**Categories**	**HEALTHLINES**				**IPCRESS**				**ITAS**				**MIR**			
		N	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**

Sample size at baseline		**609**				**295**				798				480			
Baseline Characteristics																	
Depressive Symptom Severity																	
Z-score across measures		609	0.00	1.00	0	295	0.26	0.84	0	796	0.0	1.00	2	480	0.1	0.94	0
PHQ-9 score		609	16.87	4.60	0	0			295	0			798	480	16.4	5.48	0
BDI-II Score		0			609	295	33.16	8.80	0	0			798	480	31.1	9.91	0
HADS Depression Score		0			609	0			295	0			798	0			480
GHQ Score		0			609	0			295	796	7.7	3.22	2	0			480
EPDS Score		0			609	0			295	0			798	0			480
PROMIS T-Score		609	72.06	5.84	0	295	74.15	5.86	0	0			798	480	72.8	6.66	0
Depressive Disorder Severity																	
CIS-R Total Score		0			609	295	29.60	8.68	0	798	23.3	8.45	0	480	27.7	8.29	0
CIS-R Depression Scales Sum Score		598	14.66	2.88	11	295	14.79	3.03	0	798	11.1	4.03	0	480	14.2	3.06	0
CIS-R Anxiety Scales Sum Score		0			609	295	14.81	6.78	0	798	12.1	5.77	0	480	13.4	6.21	0
Depression Duration		598	4.15	1.16	11	295	3.05	1.19	0	798	2.9	1.33	0	480	4.2	1.01	0
Average Anxiety Duration		0			609	287	2.08	1.03	8	774	1.6	0.81	24	470	2.4	0.99	10
Number of Comorbid Disorders		0			609	295	2.32	1.00	0	798	1.0	0.82	0	480	2.1	0.97	0
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Agoraphobia	No				609	267	90.5		0	765	95.9		0	399	83.1		0
	Yes					28	9.5			33	4.1			81	16.9		
CFS	No				609	75	25.4		0	348	43.6		0	169	35.2		0
	Yes					220	74.6			450	56.4			311	64.8		
GAD	No				609	109	36.9		0	798	100		0	261	54.4		0
	Yes					186	63.1			0	0			219	45.6		
MADD	No				609	214	72.5		0	94	11.8		0	309	64.4		0
	Yes					81	27.5			704	88.2			171	35.6		
OCD	No				609	233	79		0	739	92.6		0	418	87.1		0
	Yes					62	21			59	7.4			62	12.9		
Panic Disorder	No				609	279	94.6		0	761	95.4		0	435	90.6		0
	Yes					16	5.4			37	4.6			45	9.4		
Social Phobia	No				609	251	85.1		0	744	93.2		0	422	87.9		0
	Yes					44	14.9			54	6.8			58	12.1		
Specific Phobia	No				609	249	84.4		0	776	97.2		0	418	87.1		0
	Yes					46	15.6			22	2.8			62	12.9		
History of Depression	No	44	7.7		38	68	23.1		0	292	36.6		0	84	17.5		0
	Yes	527	92.3			227	76.9			506	63.4			396	82.5		
History of ADM treatment	No	93	16.1		32	131	44.4		0	394	49.4		0	103	21.5		0
	Yes	484	83.9			164	55.6			404	50.6			377	78.5		
History of Any Past Treatment	No		.		609	131	44.4		0	360	45.1		0	94	19.6		0
	Yes		.			164	55.6			438	54.9			386	80.4		
Social Support		**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**
Social Support Total Score		0			609	295	20.05	3.77	0	0			798	480	20.5	4.08	0

Accepted		0			609	295	2.49	0.63	0	797	2.5	0.64	1	480	2.6	0.61	0
Cared about		0			609	295	2.77	0.47	0	797	2.7	0.52	1	480	2.7	0.49	0
Can rely on others		0			609	295	2.55	0.63	0	797	2.5	0.64	1	480	2.6	0.59	0
Can talk to others		0			609	295	2.27	0.71	0	797	2.3	0.74	1	480	2.4	0.71	0
Encouraged		0			609	295	2.55	0.61	0	797	2.4	0.66	1	480	2.5	0.64	0
Made to feel Happy		0			609	295	2.42	0.63	0	0			798	480	2.5	0.65	0
Made to feel Important		0			609	295	2.42	0.68	0	0			798	480	2.5	0.65	0
Made to feel Loved		0			609	295	2.58	0.58	0	794	2.5	0.64	4	480	2.6	0.58	0
Life Events																	
Life Events Total Score		0			609	295	1.44	1.25	0	798	2.45	1.75	0	480	1.04	1.04	0
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Arguments/Disputed	No	0			609	204	69	0.46	0				798	419	87		0
	Yes					91	31							61	13		
Bereavement	No	0			609	259	88		0	666	86		19	388	81		0
	Yes					36	12			113	15			92	19		
Debt	No	0			609	177	60		0	563	71		1	351	73		0
	Yes					118	40			234	29			129	27		
Divorce	No	0			609	245	83		0	721	95		36	457	95		0
	Yes					50	17			41	5			23	5		
Illness or Injury	No	0			609	224	76		0	600	84		86	365	76		0
	Yes					71	24			112	16			115	24		
Legal Trouble/Court case	No	0			609	275	93		0	718	90		0	443	92		0
	Yes					20	7			80	10			37	8		
Lost job/Sacked	No	0			609	273	93		0				798	458	95		0
	Yes					22	8							22	5		
Victim of Crime/Assaulted	No	0			609	278	94		0	742	95		19	462	96		0
	Yes					17	6			37	5			18	4		
Socio-demographics		**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Age		598	49.51	12.85	11	295	34.94	11.61	0	798	43.2	14.78	0	480	50.7	13.18	0
Gender	Female	417	68.5		192	200	67.8		95	543	68		0	332	69.2		0
	Male	192	31.5		417	95	32.2		200	255	32			148	30.8		
Ethnicity	white	592	97.5		17	281	95.3		14				798	468	97.7		1
	non-white	15	2.5		594	14	4.7		281					11	2.3		
Highest Level of Educational Attainment	Degree or higher	152	25.3		8	102	34.6		0				798	95	19.8		250
	A-level or Diplomas	117	19.5			88	29.8							135	28.1		
	GCSE	249	41.4			62	21							150	31.3		
	None or Other	83	13.8			43	14.6							100	20.8		
Employment Status	Employed	275	45.7		7	178	60.3		0	415	52.1		1	235	49.2		2
	Not seeking employment	149	24.8			82	27.8			256	32.1			141	29.5		
	Unemployed	178	29.6			35	11.9			126	15.8			102	21.3		
Financial Wellbeing	Doing OK financially				609	128	43.4		0	355	44.7		3	200	41.7		0
	Just about getting by					84	28.5			258	32.5			152	31.7		

	Struggling financially					83	28.1			182	22.9			128	26.7		
Housing Tenancy Status	Home owner	341	56.2		2	120	40.7		0	492	62		4	255	53.1		0
	Tenant	242	39.9			125	42.4			238	30			186	38.8		
	Other	24	4			50	16.9			64	8.1			39	8.1		
Marital Status	Married/cohabiting				609	108	36.6		0	467	58.7		3	285	59.4		0
	Single					141	47.8			178	22.4			94	19.6		
	No longer married					46	15.6			150	18.9			101	21		
Others																	
Audit PC Total Score		0			609	295	2.56	2.55	0	176	1.8	3.43	622	480	1.5	2.00	0
EQ5D Cross-walked Index Score		541	0.52	0.28	68	290	0.80	0.08	5	0			798	478	0.6	0.27	2
LTC	No				609	253	85.8		0	534	66.9		0	355	74		0
	Yes					42	14.2			264	33.1			125	26		
Outcome																	
Z-score across measures 3-4 months		523	0.00	1.00	86	206	0.16	1.07	89	0			798	424	0.2	1.03	56
PROMIS T-Score 3-4 months		525	67.42	8.57	84	222	58.89	13.93	73	0			798	431	61.7	11.98	49
Z-score across measures 6-8 months		515	-0.11	1.01	94	203	0.16	1.15	92	588	0.0	1.00	210	392	0.1	1.06	88
PROMIS T-Score 6-8 months		516	66.20	9.28	93	211	59.90	14.12	84	0			798	433	61.7	11.96	47
Z-score across measures 9-12 months		516	-0.30	1.02	93	0			295	0			798	378	0.0	1.02	102
PROMIS T-Score 9-12 months		516	64.52	9.76	93	0			295	0				389	59.6	13.79	91
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Remission at 3-4 months	No	364	69.6		86	136	66		89				798	297	70		56
	Yes	159	30.4			70	34							127	30		
Attrition at 3-4 months	No	523	85.9		609	206	69.8		295				798	424	88.3		480
	Yes	86	14.1			89	30.2							56	11.7		

Appendix 2 Table 1. Continued (2).

		Study															
Prognostic Indicator	**Categories**	**PANDA**				**REEACT**				**RESPOND**				**TREAD**			
		N	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**
Sample size at baseline		652				685				220				361			
Baseline Characteristics																	
Depressive Symptom Severity																	
Z-score across measures		650	-0.66	1.02	2	685	0.17	0.75	0	220	0.00	1.00	0	361	0.15	0.88	0
PHQ-9 score		650	11.98	5.79	2	685	16.65	4.25	0	0			220	0			361
BDI-II Score		652	23.90	10.25	0	0	.	.	685	0			220	361	32.06	9.24	0
HADS Depression Score		0			652	0	.	.	685	0			220	0			361
GHQ Score		0			652	0	.	.	685	0			220	0			361
EPDS Score		0			652	0	.	.	685	220	17.60	3.43	0	0			361
PROMIS T-Score		652	65.62	8.40	0	685	71.96	5.27	0	0			220	361	73.26	6.30	0
Depressive Disorder Severity																	361

CIS-R Total Score		652	21.25	10.14	0	685	26.87	9.54	0	220	26.30	7.62	0	361	28.08	7.83	0
CIS-R Depression Scales Sum Score		652	10.58	4.90	0	685	11.23	3.18	0	220	13.12	2.87	0	361	14.26	2.93	0
CIS-R Anxiety Scales Sum Score		652	10.66	6.37	0	685	13.27	6.66	0	220	13.19	5.80	0	361	13.74	5.95	0
Depression Duration		652	3.21	1.67	0	685	3.11	1.41	0	220	2.47	0.90	0	361	2.86	1.07	0
Average Anxiety Duration		643	1.94	0.94	9	685	2.04	0.97	0	220	1.56	0.70	0	361	1.85	0.80	0
Number of Comorbid Disorders		652	1.43	1.18	0	685	2.06	1.17	0	220	2.05	0.81	0	360	2.20	1.17	1
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Agoraphobia	No	610	93.6		0	601	87.7		0	197	89.5		0	325	90.3		1
	Yes	42	6.4			84	12.3			23	10.5			35	9.7		
CFS	No	364	55.8		0	253	36.9		0	75	34.1		0	104	28.8		0
	Yes	288	44.2			432	63.1			145	65.9			257	71.2		
GAD	No	353	54.1		0	268	39.1		0	104	47.3		0	122	33.9		1
	Yes	299	45.9			417	60.9			116	52.7			238	66.1		
MADD	No	400	61.3		0	484	70.7		0	135	61.4		0	275	76.2		0
	Yes	252	38.7			201	29.3			85	38.6			86	23.8		
OCD	No	600	92		0	597	87.2		0	199	90.5		0	311	86.1		0
	Yes	52	8			88	12.8			21	9.5			50	13.9		
Panic Disorder	No	610	93.6		0	606	88.5		0	209	95		0	347	96.1		0
	Yes	42	6.4			79	11.5			11	5			14	3.9		
Social Phobia	No	584	89.6		0	601	87.7		0	202	91.8		0	309	85.6		0
	Yes	68	10.4			84	12.3			18	8.2			52	14.4		
Specific Phobia	No	554	85		0	558	81.5		0	188	85.5		0	300	83.1		0
	Yes	98	15			127	18.5			32	14.5			61	16.9		
History of Depression	No	130	19.9		0	194	28.4		3	111	51.4		4	107	29.6		0
	Yes	522	80.1			488	71.6			105	48.6			254	70.4		
History of ADM treatment	No	261	40		0	254	37.2		3	115	52.3		0	137	38		0
	Yes	391	60			428	62.8			105	47.7			224	62		
History of Any Past Treatment	No	261	40		0	215	31.5		3	90	40.9		0	137	38		0
	Yes	391	60			467	68.5			130	59.1			224	62		
Social Support		**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**	**N**	**Mean**	**SD**	**N missing**
Social Support Total Score		652	20.64	3.82	0	0			685	0			220	361	20.13	3.77	0
Accepted		652	2.62	0.58	0	0			685	0			220	361	2.55	0.57	0
Cared about		652	2.76	0.49	0	0			685	0			220	361	2.75	0.47	0
Can rely on others		652	2.62	0.61	0	0			685	219	2.52	0.69	1	361	2.61	0.58	0
Can talk to others		652	2.42	0.70	0	0			685	219	2.38	0.73	1	361	2.30	0.72	0
Encouraged		652	2.58	0.60	0	0			685	219	2.56	0.64	1	361	2.54	0.58	0
Made to feel Happy		652	2.50	0.62	0	0			685	0			220	361	2.39	0.65	0
Made to feel Important		652	2.50	0.66	0	0			685	0			220	361	2.43	0.65	0
Made to feel Loved		652	2.65	0.55	0	0			685	0			220	361	2.57	0.59	0
Life Events																	
Life Events Total Score		652	1.22	1.19	0	0			685	0			220	361	1.49	1.28	0
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**

Arguments/Disputed	No				652				685				220				361
	Yes																
Bereavement	No	519	80		0				685				220	300	83		0
	Yes	133	20											61	17		
Debt	No	490	75		0				685				220	210	58		1
	Yes	162	25											150	42		
Divorce	No	589	90		0				685				220	300	83		0
	Yes	63	10			0				0				61	17		
Illness or Injury	No	272	42		0	0			685	0			220	265	73		0
	Yes	380	58			0				0				96	27		
Legal Trouble/Court case	No	604	93		0	0			685	0			220	335	93		0
	Yes	48	7			0				0				26	7		
Lost job/Sacked	No	613	94		0	0			685	0			220	337	93		0
	Yes	39	6			0				0				24	7		
Victim of Crime/Assaulted	No	619	95		0	0			685	0			220	329	91		0
	Yes	33	5			0				0				32	9		
Socio-demographics		**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Age		652	39.73	14.95	0	685	39.88	12.65	0	220	28.66	6.42	0	361	39.84	12.63	0
Gender	Female	384	58.9		0	461	67.3		0	220	100		0	237	66		2
	Male	268	41.1			224	32.7			0	0			122	34		
Ethnicity	white	579	88.8		0	668	97.5		0	175	79.9		1	336	93.1		0
	non-white	73	11.2			17	2.5			44	20.1			25	6.9		
Highest Level of Educational Attainment	Degree or higher	230	35.3		202	253	39.2		343	51	24.3		144	87	24.1		170
	A-level or Diplomas	220	33.7			89	13.8			25	11.9			104	28.8		
	GCSE	145	22.2			298	46.1			62	29.5			102	28.3		
	None or Other	57	8.7			6	0.9			72	34.3			68	18.8		
Employment Status	Employed	433	66.4		0	424	61.9		0	0	0		0	230	63.7		0
	Not seeking employment	146	22.4			138	20.1			220	100			83	23		
	Unemployed	73	11.2			123	18			0	0			48	13.3		
Financial Wellbeing	Doing OK financially	364	55.8		0				685				220	115	31.9		1
	Just about getting by	204	31.3											124	34.4		
	Struggling financially	84	12.9											121	33.6		
Housing Tenancy Status	Home owner	261	40		0	346	50.6		1				220	171	47.4		0
	Tenant	262	40.2			236	34.5							143	39.6		
	Other	129	19.8			102	14.9							47	13		
Marital Status	Married/cohabiting	255	39.1		0	348	50.9		1	155	71.4		3	167	46.3		0
	Single	296	45.4			229	33.5			42	19.4			116	32.1		
	No longer married	101	15.5			107	15.6			20	9.2			78	21.6		
Others																	
Audit PC Total Score		652	3.89	3.33	0	0	.	.	685	0			220	360	5.74	3.54	1
EQ5D Cross-walked Index Score		650	0.72	0.19	2	684	0.78	0.09	1	0			220	356	0.79	0.08	5
LTC	No	536	82.2		0	518	75.6		0	184	83.6		0	275	76.2		0

	Yes	116	17.8			167	24.4			36	16.4			86	23.8		
Outcome																	
Z-score across measures 3-4 months		524	-0.43	0.90	128	522	-0.09	0.95	163	185	0.00	1.00	35	288	0.04	1.00	73
PROMIS T-Score 3-4 months		527	56.94	11.05	125	522	61.20	10.61	163	0			220	290	59.59	12.03	71
Z-score across measures 6-8 months		0			652	0			685	0			220	222	-0.08	1.02	139
PROMIS T-Score 6-8 months		0			652	0			685	0			220	222	57.25	13.52	139
Z-score across measures 9-12 months		0			652	481	-0.35	0.90	204	140	-0.23	1.12	80	255	-0.25	0.96	106
PROMIS T-Score 9-12 months		0			652	481	58.22	10.79	204	0			220	256	54.90	14.58	105
		N	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**	**N**	**%**		**N missing**
Remission at 3-4 months	No	164	31.3		128	247	47.3		163	81	43.8		35	188	65.3		73
	Yes	360	68.7			275	52.7			104	56.2			100	34.7		
Attrition at 3-4 months	No	528	81		652	522	76.2		685	185	84.1		220	288	79.8		361
	Yes	124	19			163	23.8			35	15.9			73	20.2		

Appendix 3. Data Management Plan and Data Sharing Agreement agreed by all study teams

DATA MANAGEMENT PLAN

0. Proposal name

Understanding and predicting prognosis for adults with depression

1. Description of the data

1.1 Type of study: Secondary data analysis of individual patient data from randomised controlled trials.

1.2 Types of data: Patient data vary in each of the RCTs though most studies have data available on: diagnosis; the number and duration of previous episodes of depression; age at onset of first depressive episode; and social functioning. All have data on severity of depression and other affective disorder symptoms as measured on the CIS-R, socio-demographics, and treatment type

1.3 Format and scale of the data: The individual patient data will include data on approximately 7,000 RCT participants.

Data outputs from each of the software packages used for the analyses can be saved and/or transformed into '.csv' format and therefore easily shared. All data including outputs from analyses will be saved and stored on UCL computers and backed up on the UCL server so that there are always two electronic copies.

2. Data collection / generation

2.1 Methodologies for data collection / generation: Data have all been collected previously and will only need to be collated from all sources to form the dataset described above. No new data will be collected.

2.2 Data quality and standards: Consistency is ensured as all measures of symptom severity are extremely widely used and have very good validity and reliability. All of the RCTs used the Revised Clinical Interview Schedule (CIS-R), and other measures of depression including the Structured Clinical Interview for DSM Disorders (SCID), the Hamilton Rating Scale for Depression (HRSD or HAM-D), the Beck Depression Inventory Second Edition (BDI-II) or the Patient Health Questionnaire 9 item version (PHQ-9).

3. Data management, documentation and curation

3.1 Managing, storing and curating data: All data will be handled in accordance with the Data Protection Act 1998. All data will be pseudo-anonymised and electronically stored at UCL in password protected files on password protected computers at UCL. At UCL electronic files are backed up every night.

3.2 Metadata standards and data documentation: Statistical analyses will be completed with a 'log file' created of every command entered into the statistical software and a log of the outputs for all commands. 'Do files' will be kept with all the commands used to manage and analyse the data. 'Log files' and 'Do files' can be shared as text file documents and the same commands can be run by other researchers to replicate my analyses. In addition academic publications of this research will include sufficient details in the methods sections so other researchers could replicate the studies.

3.3 Data preservation strategy and standards: In accordance with the current Records Retention Schedule, research data are retained by UCL as sponsor for 20 years after the research has ended. The UCL Records Office maintains records in a safe and secure offsite location. Access to stored records is strictly controlled. The applicant and sponsors agree to archive study documents for 5 years from the study end.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards: This project will be conducted in adherence with UCL's Research Data Policy and Information Security Policy. To measure performance and improvement, UCL will use the international standard for records management, ISO 15489, and the *Lord Chancellor's Code of Practice on the management of records issued under section 46 of the Freedom of Information Act 2000* (2009). In conducting this study the Sponsor, UCL and IAPT services shall comply with all laws and statutes, as amended from time to time, applicable to the performance of research studies with human participants including, but not limited to: The Human Rights Act 1998; The Data Protection Act 1998; The Freedom of Information Act 2000; The Research Governance Framework for Health and Social Care, issued by the UK Department of Health (Second Edition 2005) or the Scottish Health Department Research Governance Framework for Health and Community Care (Second Edition 2006). The studies will also be conducted in accordance with the World Medical Association Declaration of Helsinki entitled 'Ethical Principles for Medical Research Involving Human Subjects' (1996 version).

4.2 Main risks to data security: No new data will be collected for this study therefore there is minimal risk of accidental breakdown of confidentiality. All data will be treated as confidential and will be stored electronically on password protected computers in locked rooms at UCL's Research Department of Clinical, Educational and Health Psychology

5. Data sharing and access

5.1 Suitability for sharing: Yes

5.2 Discovery by potential users of the research data:

Our local data policy is to always present data in a timely fashion at national and international conferences in line with the MRC's policy on data sharing and preservation. Data are discussed in detail at smaller meetings within the Divisions, presentations are made in academic seminars and at annual poster competitions, to allow early sharing of results. Studies are published as Open Access available on Europe PubMedCentral in line with UCL's publications policy) in high impact peer-reviewed journals and are often discussed on widely read academic blogs making them accessible to a wide audience.
Further, it is the policy of UCL that following primary use (e.g. publication) or when research data is archived for long term preservation, these data will be made available in the most open manner appropriate. Unless covered by third party contractual agreements, legislative obligations or provisions regarding ownership, UCL research data will be provided using a Creative Commons CC0 waiver similar to existing publishing conventions. This will ensure that re-used data are unambiguously identifiable and that appropriate credit and attribution is made; supported by data citation guidelines.

5.3 Governance of access: I will have primary responsibility and decision making powers over allowing access to the data. This will be subject to governance according to publication status, consent in place, and identification of protectable IP. Where required I will consult with UCLB and other UCL offices in order to make timely informed decisions.

5.4 The study team's exclusive use of the data: The individual patient data dataset will be made available on request subject to necessary permissions from each of the individual RCT's chief investigators or data custodians. Our local data policy is to share research data in a timely fashion in line with Wellcome Trust data sharing guidelines. Data will be presented at scientific meetings when considered robust and often well before publication; data will also be published as soon as appropriate.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions:
All study data will be pseudo-anonymised and so will not restrict data sharing.
5.6 Regulation of responsibilities of users: None of the collaborators will require access to the data gathered for these studies. Should this situation change all external users will be required to handle the data in a responsible manner, respecting consents involved and data management policy recommendations, including those of the Wellcome Trust.

6. Responsibilities

UCL's Provost and Council have overall responsibility for records management. Operational responsibility is delegated to the Records Manager, who is responsible for developing records management procedures, advising on good practice and promoting compliance with this Policy. UCL advocate a data lifecycle approach to data management. All members of staff are responsible for ensuring that their work is documented appropriately and that the records which they create or receive are managed correctly. They also have a responsibility to know what information they hold and where it is held. As the fellowship holder I will be responsible for data management, meta-data creation and both data security and data quality assurance, supported in these responsibilities by the sponsors and UCL Information Services Department who are responsible for ensuring all electronic files are backed-up on UCL's secure data servers.

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	URL or Reference
Data Management Policy & Procedures	http://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf http://www.ucl.ac.uk/jro/standingoperatingprocedures/documents/FINAL_Guide_on_Data_management_version_July_2009_revised_22.2.2010.pdf
Data Protection Policy	https://www.ucl.ac.uk/informationsecurity/policy/public-policy/Data_protection_policy_ISC_20110215

Research Data Policy	http://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf
Data Sharing Policy	http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/
Publication Policy	http://www.ucl.ac.uk/library/open-access/publications-policy
Institutional Information Security Policy	*https://www.ucl.ac.uk/informationsecurity/policy*
8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their **telephone & email contact details**	
Dr Joshua EJ Buckman	

Example of Signed Data Sharing Agreement

Dr Joshua EJ Buckman, Clinical Research Fellow
Research Department of Clinical, Educational and Health Psychology
Centre for Outcomes Research and Effectiveness (CORE)
University College London
1-19 Torrington Place
London WC1E 7HB
Tel: 0207 679 1785

12th March 2019

Professor Simon Gilbody (York)
Professor Tony Kendrick (Southampton)
Professor David Kessler (Bristol)
Professor Glyn Lewis (UCL)
Professor David Richards (Exeter)
Professor Deborah Sharp (Bristol)
Mrs Laura Thomas (Bristol)
Professor Nicola Wiles (Bristol)
Ms Sally Brabyn (York)
Dr Liz Littlewood (York)
Professor Chris Salisbury (Bristol)

Dear All,

Re: Agreement on use of data from RCTs for the 'Dep-GP' Individual Patient Data Dataset

I am writing to set out some conditions on the use of the data we have requested from you all to form an Individual Patient Data dataset of RCTs in which adults with depression recruited from UK GPs have been randomised to any treatment and have had depressive and anxious symptoms measured with the CIS-R. This document is intended to set out the key principles of the agreement for the use of your data, it is intended to work in conjunction with (not to supersede) any more detailed agreements with individual study teams already signed by our research team.

The studies derived from the IPD will be led by myself and the sponsors and collaborators listed below, and conducted as part of a Wellcome Trust Clinical Research Fellowship finishing in April 2020. We intend to maintain the IPD dataset beyond this until such as time as you tell us you would like us to remove your data or as we are required by data-handling laws or statues.

Do please let me know if there are any other details you would like to include in this agreement.

Principles of this data-sharing agreement

1. All data will be treated in accordance with the Data Protection Act 1998, data will be pseudonymised, stored securely and treated as confidential.

2. In conducting this study the Sponsor, UCL shall comply with all laws and statutes, as amended from time to time, applicable to the performance of research studies with human participants including, but not limited to: The Human Rights Act 1998, The Data Protection Act 1998, The Freedom of Information Act 2000, The Research Governance Framework for Health and Social Care, issued by the UK Department of Health (Second Edition 2005) or the Scottish Health Department Research Governance Framework for Health and Community Care (Second Edition 2006)

3. The studies using data from the IPD dataset will be conducted in accordance with the World Medical Association Declaration of Helsinki entitled 'Ethical Principles for Medical Research Involving Human Subjects' (1996 version) and in accordance with the terms and conditions of the ethical approval given to the trial.

4. All publications based on data included in the IPD will include each Chief Investigator from the included RCTs as co-authors and potentially nominated Principal Investigators where appropriate.

5. All articles to be submitted for publication will be sent to all co-authors for approval prior to submission.

6. Data from the IPD will not be used for any purposes outside of the approved investigations on understanding and predicting prognosis for adults with depression and closely related topics. Any additional analyses will be subject to agreement from the chief investigators of the individual studies.

7. Data from the IPD will not be shared with external parties, any requests for data sharing will be signposted to the chief investigators or custodians of the data for each individual RCT included in the IPD dataset and it is expected that these requests will then be reviewed as usual by those individuals or teams.

NAME: Dr Joshua E J Buckman
DATE: 12.03.2019

NAME: David Kessler
DATE: 04/07/2019

Sponsors of Wellcome Trust Fellowship: Professors Stephen Pilling and Glyn Lewis
Collaborators: Dr Gareth Ambler, Professors Rob DeRubeis, Simon Gilbody, Steve Hollon, Tony Kendrick, & Ed Watkins
Collaborators on individual analyses: Professor Ian White, Drs Rob Saunders, Ciaran O'Driscoll, Katherine Clarke, Zach Cohen, & Eiko Fried.

Appendix 4. Ethical approval and Trial Registration details

Table 1. Ethical approval and Trial Registration details of the studies included in the Dep-GP IPD database

Study	Ethical Approvals	Trial Registration details
AHEAD	South West Multicentre Ethics Committee and ethics committees covering Hampshire, East Dorset, Wiltshire, West Sussex and South West Surrey	ISRCTN14453847; https://doi.org/10.1186/ISRCTN14453847
CADET	Granted by NHS Health Research Authority & NRES Committee South West (NRES/07/H1208/60)	ISRCTN32829227; https://doi.org/10.1186/ISRCTN32829227
COBALT	Approvals were granted by West Midlands Research Ethics Committee (NRES/07/H1208/60) and research governance approval was obtained from the local Primary Care Trusts/Health Boards	ISRCTN38231611; https://doi.org/10.1186/ISRCTN38231611
GENPOD	Approvals granted by South West Research Ethics Committee (MREC 02/6/076) and research governance approval was granted by Bristol, Manchester and Newcastle Primary Care NHS Trusts.	ISRCTN31345163; https://doi.org/10.1186/ISRCTN31345163
HEALTHLINES	Approval was granted by the National Research Ethics Service Committee South West–Frenchay (Reference 12/SW/0009)	ISRCTN14172341; https://doi.org/10.1186/ISRCTN14172341
IPCRESS	Approval granted by Royal Free and Hampstead Research Ethics Committee, reference number 05/Q0501/18	ISRCTN45444578; https://doi.org/10.1186/ISRCTN45444578
ITAS	Bro Taf Health Authority and United Bristol Healthcare Trust Local Research Ethics Committee	ISRCTN57116180; https://doi.org/10.1186/ISRCTN57116180
MIR	Approvals were granted by South East Wales Research Ethics Committee Panel C (ref: 12/WA/0353); Bristol Clinical Commissioning Group (CCG), and other CCGs provided research governance assurance.	ISRCTN06653773; https://doi.org/10.1186/ISRCTN06653773
PANDA	Ethical approval was granted by Bristol Research Ethics Committee Centre (12/SW/0267).	ISRCTN84544741; https://doi.org/10.1186/ISRCTN84544741
REEACT	Approval was granted by Leeds (East) research ethics committee (08/H1306/77).	ISRCTN91947481; https://doi.org/10.1186/ISRCTN91947481
RESPOND	Approvals were granted by the Scotland A Multi-centre Research Ethics Committee (MREC; reference number MREC/03/0/127) and site-specific approval was obtained from 10 relevant local ethics committees and 10 primary care trusts (PCTs)	ISRCTN16479417; https://doi.org/10.1186/ISRCTN16479417
TREAD	Approvals were granted by West Midlands multicentre research ethics committee (MREC 05/MRE07/42), and research governance approval was given by the relevant local National Health Service primary care trusts	ISRCTN16900744; https://doi.org/10.1186/ISRCTN16900744

Appendix 5. Results from Sensitivity Analyses using the Multidimensional IRT conversion of BDI-II and PHQ-9 scores to the PROMIS T-score at 3-4 months post-baseline

Results from Chapter 4 for Social Support

Appendix 5 Table 1. Associations of Social Support with prognosis at 3-4 months post-baseline using the multidimensional IRT conversion to the PROMIS T-score as the outcome, adjusted for treatment, 'disorder severity' and covariates.

	Sensitivity Analysis	
Social Support Domain	**MIRT PROMIS T-Score at 3-4 months**	
	Mean difference (95%CI)	**I^2**
Total Score	-4.43(-7.18 to -1.68)	0
Accepted	-3.67(-5.85 to -1.50)	0
Cared about	-4.03(-6.72 to -1.33)	0
Made to feel happy	-2.18(-4.14 to -0.22)	0
Made to feel important	-0.96(-2.94 to 1.01)	0
Made to feel loved	-2.46(-4.66 to -0.26)	0
Can rely on others	-2.53(-4.62 to -0.45)	0
Supported or Encouraged	-3.53(-5.60 to -1.45)	0
Can talk to others	-2.32(-4.29 to -0.34)	18

Note: Total score is scaled 8-24 and unless otherwise stated, estimates are per one-point increase; individual items are scored 1-3 and estimates are per 1-point increase. All models adjusted for random allocation in each study, depressive symptom severity, average anxiety duration, depression duration, panic disorder, and a history of antidepressant treatment, gender, age, marital status, and employment status †Only available for 5 studies, excludes COBALT

Appendix 5 Table 2. Associations of life events, socio-demographics, and long-term health condition status with prognosis at 3-4 months post-baseline using the multidimensional IRT conversion to the PROMIS T-score as the outcome, adjusted for treatment, 'disorder severity' and covariates.

Prognostic Factors		Sensitivity Analysis	
Type	**Prognostic Indicator**	**MIRT PROMIS T-Score**	
		Mean Difference (95%CI)	**I^2**
Life events	Life events total score	0.36(-0.17 to 0.89)	46.91
	Any life events	0.92(-0.48 to 2.32)	54.00
	Two plus life events	0.67(-0.58 to 1.93)	45.00
	Arguments	0.45(-0.60 to 1.49)	0.00
	Bereavement	-0.19(-1.39 to 1.01)	23.28
	Debt	1.27(-0.32 to 2.86)	61.89
	Divorce	1.00(-0.92 to 2.92)	33.19
	Victim of violent crime/assault	1.26(-0.58 to 3.11)	17.01
	Illness or Injury	-0.16(-1.10 to 0.78)	0.00
	Legal troubles	-0.70(-2.35 to 0.94)	9.65
	Sacked/Lost job	-0.44(-2.18 to 1.30)	0.00
Socio-demographics	Age[†]	0.01(-0.06 to 0.07)	73.17
	Gender[†]	0.62(-0.47 to 1.70)	27.18
	Ethnicity	1.81(-0.18 to 3.80)	0.00
	Marital Status	0.82(0.27 to 1.37)	6.03
	Employment Status	1.44(0.62 to 2.25)	43.49
	Financial Wellbeing	0.64(0.07 to 1.21)	0.00
	Housing Status	1.33(0.52 to 2.14)	0.00
	Highest level of Educational Attainment	0.07(-0.38 to 0.53)	0.00
Long-term health condition status		1.29(0.07 to 2.52)	7.00

Note: Total score is scaled 0-8 and unless otherwise stated, estimates are per one-point increase; individual items are scored 0-1 and estimates are per 1-point increase.
All models are adjusted for allocated treatment, gender, age; depressive symptom severity, average anxiety duration, depression duration, and history of treatment with antidepressants;
[†] not adjusted for itself