Predicting and preventing relapse of depression in primary care: a mixed methods study

Andrew Stephen Moriarty MB BS MRes MRCGP FHEA

Submitted for the Degree of Doctor of Philosophy in Medical Sciences

The University of Hull and the University of York

Hull York Medical School

September 2023

Abstract

Background

Most people with depression are managed in primary care. Relapse (re-emergence of depression symptoms after improvement) is common and contributes to the burden and morbidity associated with depression. There is a lack of evidence-based approaches for risk-stratifying people according to risk of relapse and for preventing relapse in primary care.

Methods

In this mixed methods study, I initially reviewed studies looking to predict relapse of depression across all settings. I then attempted to derive and validate a prognostic model to predict relapse within 6-8 months in a primary care setting, using multilevel logistic regression analysis on individual participant data from seven studies (n=1244). Concurrently, a qualitative workstream, using thematic analysis, explored the perspectives of general practitioners (GPs) and people with lived experience of depression around relapse risk and prevention in practice.

Results

The systematic review identified eleven models; none could currently be implemented in a primary care setting. The prognostic model developed in this study had inadequate predictive performance on internal validation (C-statistic 0.60; calibration slope 0.81). I carried out twenty-two semi-structured interviews with GPs and twenty-three with people with lived experience of depression. People with lived experience of depression and GPs reflected that a discussion around relapse would be useful but was not routinely offered. Both participant groups felt there would be benefits to relapse prevention for depression being embedded within primary care.

Conclusions

We are currently unable to accurately predict an individual's risk of depression relapse. The longer-term care of people with depression in general practice could be improved by enabling continuity of care, increased

consistency and clarity around follow-up arrangements, and focussed discussions around relapse risk and prevention. Scalable, brief relapse prevention interventions are needed, which would require policy change and additional resource. We need to better understand existing interventions and barriers to implementation in practice.

Table of Contents

List of Tables

List of Figures

Acknowledgements

For Claire, Olive and Thom - my wonderful little family.

In memory of my Grandad, Barrie Hodgson (1936 – 2022).

Thanks also to:

Mum, Dad, Grandma, Aggie and Len.

My supervision team, Professors Dean McMillan, Carolyn Chew-Graham, and Simon Gilbody, for their unwavering support and generous mentorship. Also, for the supervision meetings that were always stimulating and good fun.

Professors Trevor Sheldon, Paul Tiffin and Joanne Reeve for invaluable insights as part of the Thesis Advisory Panel (TAP). Particular thanks to Trevor Sheldon for being so generous with his time and guidance while I was preparing my NIHR Fellowship application.

Professor Richard Riley, Dr Kym Snell and Dr Lucinda Archer, for helping me navigate the world of prognosis research and prognostic modelling. Dr Shehzad Ali for helping to support me through the Fellowship application and continuing to offer guidance throughout the PhD from Canada.

All other collaborators, patient advisory group colleagues, interview participants, and those who supported me through the NIHR Fellowship application and beyond. Special mention to Dr Lewis Paton, Dr Nick Meader, Dr Joshua Buckman, Jess Hendon, Sarah Dawson, Professor Rachel Churchill, Dr Bob Phillips and Professor Una Macleod. Thanks to Professors David Richards, Chris Salisbury and Jaime Delgadillo for sharing their data with me. Thanks to Kerry Cipriano, Jo Orchard, Wendy Cattle and Natalie Smith for operational support. Also, to my Partners and colleagues at MyHealth and the LMC.

Author's declaration

I confirm that this work is original and that if any passages or diagrams have been copied from academic papers, books, the internet or any other sources these are clearly identified by the use of quotation marks and the references are fully cited. I certify that, other than where indicated, this is my own work and does not breach the regulations of HYMS, the University of Hull or the University of York regarding plagiarism or academic conduct in examinations. I have read the HYMS Code of Practice on Academic Misconduct, and state that this piece of work is my own and does not contain any unacknowledged work from any other sources. I confirm that any patient information obtained to produce this piece of work has been appropriately anonymised.

Chapter One

Introduction

1.1. General introduction to thesis

This thesis presents research undertaken for a Doctoral Research Fellowship, funded by the National Institute for Health and Care Research (NIHR). I am a National Health Service (NHS) general practitioner (GP) and have sought to address the problem of relapse of depression in a primary care setting. The majority of people who present with depression are managed in primary care (Ferenchick, Ramanuj and Pincus, 2019). While there are robust guidelines for managing people with depression, evidence for longer-term treatment beyond the acute phase of depression is less robust (NICE, 2022). A significant proportion of people with depression experience relapse after initial improvement (Beshai et al., 2011), resulting in increased morbidity for the individual and impact on their family and carers. It also constitutes a source of significant economic burden to the health service (Gauthier et al., 2019).

My motivations for exploring this problem are driven in part by my own clinical experience. As a clinical academic, I think that an important part of my role is to look for research answers to real-world clinical problems. The ability to identify and offer appropriate longer-term care to the people most vulnerable to relapse is something that would improve the clinical care of patients and the day-to-day experience of GPs in a tangible and meaningful way.

I am also motivated to better understand depression as a condition. Depression is extremely common and I have seen it affect people in such profound ways. It is a heterogeneous, complex condition, which makes it

intellectually challenging to me. I am interested in unpicking the psychological and social factors involved and how these interact in different ways to impact on the patient experience.

Finally, my primary motivation is a desire to improve clinical outcomes for patients. People with depression deserve to know that they are being offered the best treatments that are available. As I have learned from talking to my own patients, as well as insights from the Patient Advisory Group involved in this project, the thought of becoming unwell again is a real concern for people who have had depression and the evidence around this area is not as good as we should want it to be. My hope is that this thesis is a contribution towards improving it.

1.2.   Rationale for the study

The Department of Health and Social Care has identified prevention as a priority for mental health research (Department of Health, 2017). Promoting health and avoiding the pressures on health services to focus solely on acute illness is one of the key messages of the Hewitt Review, which was published following the recent reorganisation of the NHS (Department of Health and Social Care, 2023). Prevention can be thought of as: primary prevention (preventing disease before it occurs); secondary prevention (intervening to prevent people who have experienced an illness from having recurrences or further problems) and tertiary prevention (reducing complications of and supporting people to manage long-term conditions) (Baumann and Ylinen, 2020). These are all important parts of a prevention strategy; this thesis is primarily focussed on secondary prevention.

The Five Year Forward View for Mental Health highlighted that prevention is also patients' top priority, particularly the ability to access help early to stop mental health problems escalating. It also noted that too few patients receive the full range of interventions recommended by the National

Institute for Health and Care Excellence (NICE) due to waiting times, underfunding and lack of resources (Mental Health Task Force, 2016). Despite scientific progress and an increased understanding of the biological and psychological underpinnings of mental health conditions, there is no evidence of a decrease in morbidity or mortality from depression, in contrast to physical health conditions such as cardiovascular disease, stroke and cancer (Insel, 2009).

As part of the funding application for this project, people with lived experience of depression were consulted on the direction of the research. A Patient Advisory Group (PAG) was formed, and they have remained involved throughout my Fellowship. I will detail throughout the thesis how their input has informed the research. Prior to undertaking any research, members of the PAG highlighted that relapse of depression is an area of priority and that they worried about being "forgotten" after acute-phase treatment, reporting that there is currently no on-going monitoring or support system in place. This contrasts unfavourably with other long-term conditions such as diabetes mellitus and chronic obstructive pulmonary disease (COPD), where patients receive regular review and proactive follow-up. This situation appears to contravene the government's aspiration to deliver "parity of esteem" between physical and mental health problems (Mental Health Task Force, 2016). One of central aims of this research study was to address these concerns, with a view to improving patient experience and outcomes.

I also sought feedback, at the funding application stage, from the NIHR Clinical Research Network (CRN) Yorkshire and Humber Primary Care Steering Group (PCSG), a group made up of research-interested GPs, advanced nurse practitioners and other primary care health professionals. The consensus was that there is a lack of long-term care, support and monitoring for people with depression. The PCSG agreed that they were unsure how to assess risk of relapse in patients with depression and did not know which patients should receive relapse prevention interventions. They unanimously felt that there was a need for a clinical tool to guide this in practice. The project aimed to benefit the NHS by helping to improve risk-

stratification of people with depression and better understanding the perspectives of people with lived experience of depression and GPs. The longer-term goal was for this to enable more targeted and efficient use of health resources and improvement in patient outcomes.

1.3. Study aims and objectives

This study aimed to better understand the problem of relapse of depression in primary care. I aimed to understand, first, whether we can predict who will relapse (and therefore target relapse prevention at higher-risk individuals), and, secondly, whether we can improve the longer-term care of people with depression, to prevent relapse more effectively.

This led to the following objectives:

1. To identify and critically appraise existing prognostic models to predict relapse or recurrence of depression.
2. To develop and validate a prognostic model for patients with remitted depression in primary care, to predict individualised risk of relapse.
3. To explore the perspectives of people with lived experience of depression and GPs on relapse of depression, relapse risk prediction and relapse prevention interventions.

1.4. Scope and structure of thesis

Here, I describe the scope and structure of the thesis. Several of the chapters (in particular, Chapters Two, Three and Five) are based on peer-reviewed publications, which I have published throughout the Fellowship[1]. I

---

[1] Chapter Two includes text from a commissioned Editorial (Appendix 1.2), written for the *British Journal of General Practice* (Moriarty et al., 2020), an *Evidently Cochrane* blog article (Moriarty et al., 2021a) and from a Cochrane Review protocol (Moriarty et al., 2021b). Chapter Three is adapted from a Cochrane Prognosis

have highlighted this again within the individual chapters and have obtained permission to reproduce the relevant content. The thesis presents a mixed methods study, incorporating a systematic review, a quantitative study and a qualitative study. To avoid confusion, I will refer to the individual studies as "studies" and the overall mixed methods programme of work as the "thesis".

Chapter Two presents a review of the literature outlining the background to the problem of depression relapse in primary care. I present a case for the need for improved prognosis research into depressive relapse in primary care and an explanation of why a multivariable prognostic model might be a desirable way of addressing this. I also review the pre-existing literature around relapse prevention in primary care and patient and GP perspectives on this.

Chapter Three presents a systematic review and critical appraisal of existing prognostic models for predicting relapse or recurrence of depression, across all settings. This review was undertaken with support from the Cochrane Common Mental Disorders (CCMD) group at York and the Cochrane Prognosis Methods Group. The review was one of the first published Cochrane Prognosis Reviews, and the first Prognosis Review for the CCMD group.

Chapter Four outlines the methodological approaches taken to the qualitative and quantitative studies and then outlines the mixed methods approach adopted to guide the integration of findings from the two studies. In Chapter Four, I also describe the patient and public involvement and engagement (PPIE) in this study.

Chapters Five and Six present the development and validation of a novel multivariable prognostic model to predict relapse in a primary care

---

Review (Moriarty et al., 2021c) and its co-publication in the *British Journal of Psychiatry* (Moriarty et al., 2022). Chapter Five is adapted from the protocol paper published in *Diagnostic and Prognostic Research* (Moriarty et al., 2021d).

setting. This study was carried out using individual participant data (IPD) from six randomised controlled trials (RCTs) and one longitudinal cohort study.

Chapters Seven to Nine present a qualitative study, which used thematic analysis of semi-structured interviews with people with lived experience of depression and GPs. The study explored participants' perspectives on depressive relapse, risk prediction and relapse prevention interventions.

Chapter Ten presents a discussion of the overall findings of the thesis, including the mixed methods integration of findings from the different studies. In Chapter Ten, I also discuss the strengths and limitations of the thesis; implications for clinical practice, policy and future research; my reflections on the thesis overall and final conclusions.

1.5. Primary care and general practice

Before discussing the literature around depression and relapse in primary care, I will spend some time here introducing primary care and its unique role in the health service. Primary care is a term, first used in the 1920s (Starfield, Shi and Macinko, 2005), now used to describe the part of the health service that is first encountered by patients seeking healthcare. It has been characterised in terms of its four central functions, or the "four C's": (first) contact, comprehensiveness, coordination and continuity (Jimenez et al., 2021). General practice is the main provider of primary care services; other primary care providers include community pharmacists, opticians and dentists.

It is well-established that investing in and strengthening the key facets of primary care has benefits for the whole health service (Haggerty et al., 2013). In particular, relational continuity of care (or relationship-based care) has been shown to have benefits to patients, GPs and the health service as

a whole, in terms of patient outcomes and patient and GP satisfaction (Gray et al., 2018; Jeffers and Baker, 2016; Royal College of General Practitioners, 2019). Relational continuity allows the development of an ongoing partnership between GPs and their patients, which is particularly important in the management of long-term conditions (Hudon et al., 2012; Brickley et al., 2020). For this reason, continuity is a key component of patient-centred care, which the World Health Organization (WHO) has defined as that which 'meets people's expectations and respects their wishes' (World Health Organization, 2015). While there is some evidence that other (non-GP) members of the primary care team, for example practices nurses, can be effectively supported to deliver depression care (Ekers et al., 2013; Morgan et al., 2009), the evidence shows that this is not routinely the case in practice (Webster, Ekers and Chew-Graham, 2016; Murphy, Ekers and Webster, 2014; Girard et al., 2017). This study will primarily focus on the role of the GP for this reason.

It is helpful to have a brief discussion here of how general practice is provided and funded. General practices in England are contracted by the NHS (NHS England), via local commissioners [now integrated care boards (ICBs), formerly clinical commissioning groups (CCGs)], to provide generalist medical services for a defined geographical or population area. Primary care services are commissioned through a nationally-negotiated GP contract, which defines the mandatory requirements and services all general practices must provide. This is usually a general medical services (GMS) contract (the national standard GP contract). Variations on this exist; personal medical services (PMS) and alternative provider medical services (APMS) contracts are similar to GMS but allow for local variation and/or flexibility in contract terms. In addition to the core services defined in these contracts, general practices can also opt to provide additional services through enhanced services arrangements. These are either nationally-negotiated and defined [known as Directed Enhanced Services (DES)], or locally-negotiated and therefore subject to local variation [Local Enhanced Services (LES)] (The King's Fund, 2020).

The NIHR Doctoral Research Fellowship funding application for this study was written between September 2017 and April 2018. The NHS in England has undergone significant change since then. First, the formation of Primary Care Networks (PCNs) through the introduction of the PCN DES in 2019 promoted increased collaboration between practices and at-scale working. Then, the Health and Care Act 2022 legislated for a wholesale reorganisation of the NHS, with the aim of increasing integration of services organised around a "place" (Figure 1.1.). CCGs were abolished and integrated care systems (ICSs) and ICBs assumed statutory commissioning responsibilities, changing the way in which primary care is funded, supported and organised.

Integrated care systems (ICSs)
Key planning and partnership bodies from July 2022

NHS England
Performance manages and supports the NHS bodies working with and through the ICS

Care Quality Commission
Independently reviews and rates the ICS

Statutory ICS

Integrated care board (ICB)

Membership: independent chair; non-executive directors; members selected from nominations made by NHS trusts/foundation trusts, local authorities, general practice; an individual with expertise and knowledge of mental illness

Role: allocates NHS budget and commissions services; produces five-year system plan for health services

Cross-body membership, influence and alignment

Integrated care partnership (ICP)

Membership: representatives from local authorities, ICB, Healthwatch and other partners

Role: planning to meet wider health, public health and social care needs; develops and leads integrated care strategy but does not commission services

Influence   Influence

Partnership and delivery structures		
Geographical footprint	**Name**	**Participating organisations**
System Usually covers a population of 1–2 million	**Provider collaboratives**	NHS trusts (including acute, specialist and mental health) and as appropriate voluntary, community and social enterprise (VCSE) organisations and the independent sector; can also operate at place level
Place Usually covers a population of 250-500,000	**Health and wellbeing boards**	ICS, Healthwatch, local authorities, and wider membership as appropriate; can also operate at system level
	Place-based partnerships	Can include ICB members, local authorities, VCSE organisations, NHS trusts (including acute, mental health and community services), Healthwatch and primary care
Neighbourhood Usually covers a population of 30-50,000	**Primary care networks**	General practice, community pharmacy, dentistry, opticians

The King's Fund>

Figure 1.1: Diagram showing structure of integrated care systems following the Health and Care Act 2022[2]

---

[2] Figure taken from The King's Fund (The King's Fund, 2022), published under a Creative Commons Attribution-NonCommercial-NonDerivs 4.0 licence, which can be distributed for non-commercial purposes free of charge as long as appropriately attributed.

The Coronavirus disease 2019 (COVID-19) pandemic then occurred after the first year of conducting this study. As well as increasing pressures generally across the health service, the COVID-19 pandemic had well-recognised and documented effects on primary care, including an increased use of remote consultation and "total triage" models (Rawaf et al., 2020; Wanat et al., 2021; Murphy et al., 2021). Many of these changes are likely to remain in place in some form into the future (Rawaf et al., 2020).

The final piece of context to consider here is how general practice interacts with other services when delivering care for people with mental health problems. The majority of people with mental health problems are managed in primary care (Ramanuj, Ferenchick and Pincus, 2019), where they might receive antidepressant medication, brief psychological therapies from mental health practitioners employed through the PCN DES, or, increasingly, non-medical interventions such as social prescribing (Drinkwater, Wildman and Moffatt, 2019). Beyond this support in primary care, patients may seek support from NHS Talking Therapies, for anxiety and depression [formerly called Improving Access to Psychological Therapies (IAPT), prior to 2023]. This service was established as IAPT in 2008 and allows GP- or self-referral for people with anxiety or depression to receive psychological therapies according to a stepped-care model, in line with NICE guidance (Clark, 2011). For patients for whom primary care and NHS Talking Therapies are unable to meet their needs, a referral is usually made to secondary care specialist mental health services.

While the studies presented in this thesis are concerned with the management of depression specifically within primary care, this section has been intended to provide useful context, explaining primary care's role within the wider health service. I also return to some of the subjects outlined in this section in the discussion in Chapter Ten. The next chapter will explore the background to the problem of relapse of depression in primary care in the form of a literature review.

Chapter Two

Background

2.1. Introduction

This chapter builds on Chapter One by exploring the problem of depression relapse in people managed in primary care in more detail[3]. I will describe the scope of the problem and focus further on the conceptual underpinnings of relapse and its associated constructs: recurrence, remission, recovery and response. I will move on to consider the role of prognosis research in helping to define the problem and what has been done in this area so far. I will then describe the role of interventions for preventing relapse of depression. Finally, I will describe and critically appraise the available qualitative evidence exploring the perspectives of patients and primary care health professionals.

2.2. Depression and relapse in primary care

Depression is now the leading cause of disability worldwide (World Health Organisation, 2017), with an estimated prevalence in excess of 264 million people globally (Global Burden of Disease 2017, 2018). It results in significant morbidity for patients and exerts a high societal and economic cost (Richards, 2011). In terms of a diagnosis of depression, two major classification systems exist and are commonly used: the World Health Organisation's International Statistical Classification of Diseases and Related Health Problems, 11th revision (ICD-11) (World Health Organization, 2019)

---

[3] This chapter includes some text reproduced from peer-reviewed articles published as part of my Fellowship (Moriarty et al., 2020, 2021a, 2021c, 2021b). All necessary permissions and licences have been obtained from the publishers prior to reproducing the content in this thesis (Appendices 2.1 to 2.3).

and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorder, 5th edition (DSM-5) (American Psychiatric Association, 2013). These are summarised in Table 2.1. Either can be used for diagnosing depression in a primary care setting (Ferenchick, Ramanuj and Pincus, 2019), and would usually be most available to GPs through the use of clinical guidelines (NICE, 2022). The most common type of depression is called major depressive disorder in the DSM-5 and referred to as either single episode or recurrent depressive disorder in the ICD-11. Depression, however, is a heterogeneous condition, with multiple types identified across the diagnostic classification systems and clinical presentation and trajectories vary between patients (Ferenchick, Ramanuj and Pincus, 2019).

Some people with depression experience a single, time-limited episode, and no further episodes beyond that (Monroe and Harkness, 2011). For others, depression is a recurrent condition, with patients experiencing the re-emergence of depressive symptoms (relapse or recurrence) after a period of relative wellness (Beshai et al., 2011). Indeed, there has been a shift in the understanding of depression as a discrete or episodic illness to being considered a long-term relapsing-remitting condition with possibly incomplete recovery between episodes for some patients (Bockting et al., 2015). There is evidence to suggest that relapse or recurrence of depression leads to an increased risk of subsequent relapse (Burcusa and Iacono, 2007), possibly increased treatment resistance (Post, 1992), and that the risk of relapse and recurrence decreases as the period of recovery gets longer (Beshai et al., 2011; Solomon et al., 2000). The economic burden of depression is also significantly higher in those who experience a relapse or recurrence compared to those who do not (Gauthier et al., 2019). Therefore, providing on-going care following remission and intervening to prevent relapse and recurrence of depression is likely to improve the overall course of illness for individual patients.

Table 2.1: Summary of DSM-5 and ICD-11 diagnostic criteria for depression

Diagnostic classification system	Condition	Definition	Duration and frequency	Main symptoms	Additional symptoms	Exclusions
DSM-5	Major depressive disorder	Single or recurrent major depressive episodes	Two-week period	Depressed mood or loss of interest or pleasure in almost all activities	Significant unintentional change in weight loss/gain or decrease/increase in appetite; Sleep disturbance (insomnia or hypersomnia); Psychomotor changes (agitation or retardation); Tiredness, fatigue, or low energy; A sense of worthlessness or excessive, inappropriate, or delusional guilt; Impaired ability to think, concentrate, or make decisions; Recurrent thoughts of death, suicidal ideation, or suicide attempts	Symptoms do not cause clinically significant distress or impairment in social, occupational, or other important areas of functioning; Symptoms are due to the direct physiological effects of a substance (e.g., drug abuse, a prescribed medication's side effects) or a medical condition (e.g., hypothyroidism); Presence of manic symptoms (mixed episode); Symptoms better explained by schizophrenia spectrum or other psychotic disorders
ICD-11	Single episode depressive	Presence or history of one depressive	Nearly every day during a period lasting	Depressed mood or diminished interest in	Difficulty concentrating; Feelings of worthlessness or excessive or inappropriate guilt;	Prior manic, hypomanic, or mixed episodes, which would indicate the

disorder (6A70)	episode with no history of prior depressive episodes	at least two weeks	activities occurring most of the day	Hopelessness; Recurrent thoughts of death or suicide; Changes in appetite or sleep; Psychomotor agitation or retardation; Reduced energy or fatigue	presence of a bipolar disorder
Recurrent depressive disorder (6A71)	A history or at least two depressive episodes separated by at least several months without significant mood disturbance	As for single episode depressive disorder			

2.2.1. Defining relapse: use of terminology

Relapse and recurrence of depression are usually defined with respect to three further terms: response, remission and recovery [collectively, these terms are often referred to as the 5 Rs (Figure 2.1)]. Relapse in the context of depression has been defined as the re-emergence of depressive symptoms following some level of remission but preceding recovery, and is distinguished in the literature from recurrence (the onset of a new episode of depression following an extended period of remission) (Beshai et al., 2011). Remission and recovery are similarly differentiated, with remission meaning asymptomatic but still 'in episode' and recovery being defined as resolution of the underlying episode (usually after 6 to 12 months) (Bockting et al., 2015).
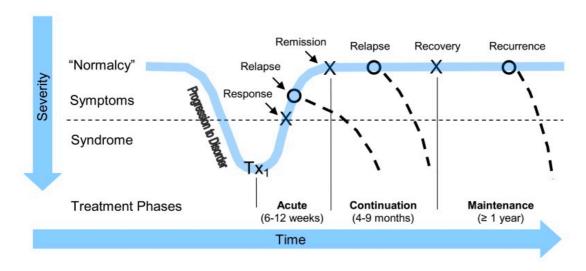
Figure 2.1: Diagram outlining depression change-points[4]

Tx1=treatment attempt 1; dashed lines indicate hypothetical worsening of depressive severity. Remission, the goal of for treatment, refers to the resolution of depressive symptoms and return to premorbid functioning; response refers to substantial clinical improvement which may or may not reach remission.

[4] Reproduced from an original report, which is in the public domain and free to reproduce, provided appropriate attribution (Gartlehner et al., 2015).

The original descriptions of relapse, recurrence and the other change-points by Frank et al. (1991) and Rush et al. (2006) defined the terms as follows: response is an initial improvement of symptoms (but not yet achieving remission), usually after treatment initiation and usually attributable to the treatment. After three weeks of minimal symptoms, a patient can be said to have entered remission. Any subsequent re-emergence of depressive symptoms after this point is described as a relapse (a return to the index episode of depression). If a relapse has not occurred by four months after remission, a patient is said to have entered recovery, after which point any re-emergence of depressive symptoms is termed a recurrence (a new episode of depression, separate from the index episode). More recent work has shown that recovery is most commonly operationalized as following an extended period of remission; between 6-12 months (Bockting et al., 2015). Relapse, then, occurs within 6-12 months, while recurrence occurs beyond 6-12 months (Frank et al., 1991; Rush et al., 2006).

The distinction between relapse and recurrence provides a useful theoretical framework and there may be some clinical relevance. The implication is that the re-emergence of symptoms in relapse is part of the unsuccessfully (or incompletely) treated index episode of depression; while in recurrence it is attributable to a new and separate episode of depression. When the MacArthur Foundation Research Network defined these terms in 1991 (Frank et al., 1991; Rush et al., 2006), their aim was to provide a framework that might be more consistently applied in the empirical literature, but also that the framework and definitions themselves be validated empirically by researchers. There have been only limited attempts to do this, though where this has been attempted researchers have found some evidence to support their validity (Beshai et al., 2011; Riso et al., 1997; De Zwart, Jeronimus and De Jonge, 2019). For example, recurrence rates following recovery are lower than relapse rates following remission (Beshai et al., 2011), and defining outcomes in line with Frank et al.'s criteria generally leads to accurate predictions about future outcomes (Riso et al., 1997).

Given the wide variability in the way in which the terms relapse and recurrence have been operationalized by researchers, however, Bockting et al. (2015) suggested using the terms interchangeably to describe the "re-emergence of symptoms following a period of relative wellness" (Bockting et al., 2015). I will use the term *relapse* throughout this thesis. The definition has been tightly operationalized for the purpose of the quantitative work. For the systematic review, I have been guided by the definitions used by the authors of the primary studies and have recorded what these are. For the qualitative work and for the purpose of drawing broader conclusions, I have been mindful of the issues with nomenclature discussed here, particularly where they may have implications for the findings reported in the thesis.

2.2.2. Relapse and recurrence of depression in primary care

A frequently-quoted statistic is that half of patients will experience a re-emergence of depressive symptoms at some point after their initial symptoms have improved, and that this increases to 70% and 90% after a second and third episode respectively (Kessler et al., 1996; Burcusa and Iacono, 2007; Solomon et al., 2000). The majority of work exploring the scope of the problem of depressive relapse has been done in secondary care settings and is likely to be of limited applicability to primary care (Buckman et al., 2018), which is where the vast majority of patients with depression are managed (Rait et al., 2009). Relapse rates, and longer-term outcomes generally, tend to be worse in speciality settings, compared to primary care settings (Ormel et al., 2020). One primary care cohort study that followed up people with remitted depression found that 37.1% of participants relapsed within one year (Lin et al., 1998). It is worth noting that this study was relatively small (n = 251) and followed up participants from randomised trials (all of whom had been prescribed antidepressant medication); therefore, this study's findings are potentially not generalisable to all primary care populations. A more recent UK-based cohort study followed up patients who had received low-intensity cognitive behavioural therapy (CBT) through IAPT. This study found that 53% of participants experienced a relapse within one

year, and that the majority (79%) of those participants did so within the first six months (Ali et al., 2017). This study is also not necessarily representative of a typical primary care population, as not all patients with depression in primary care would be referred to the IAPT (now NHS Talking Therapies) service.

Potentially more useful information can be gained from non-clinical, naturalistic cohort studies (i.e., a real-world setting, free from experimental intervention). Two systematic reviews have examined studies within naturalistic cohorts (Steinert et al., 2014; van Weel-Baumgarten et al., 2000). One review estimated primary care relapse or recurrence rates (over five years) to be between 30% and 40% (van Weel-Baumgarten et al., 2000). This review was limited by small number of studies (only two from a primary care setting), methodological weaknesses in the included studies and loosely-defined end-points. The other review found that 35% to 60% of participants experienced stable recovery (Steinert et al., 2014). Both of these reviews reported significant heterogeneity in the included studies and therefore difficulty synthesising findings to draw firm conclusions.

Finally, a study of trajectories of depression in primary care, over one year, suggested that the majority of patients with depression in primary care follow a mild trajectory of illness (Gunn et al., 2013). Other patients followed either moderate or severe static trajectories and, less commonly, dynamic trajectories (increasing or decreasing severity). This study concluded that depression in primary care is most commonly mild and not episodic. However, we should bear in mind that depression measures in this single cohort were gathered only three-monthly (by self-report postal survey, with structured interviews only at baseline and 12 months), and so the data may not have been sufficiently granular to identify fluctuations in symptoms over time.

In addition to people who experience single or recurrent episodes of depression, there is a further group of patients who have a more chronic and persistent form of depression. An estimate from the review of naturalistic cohorts suggested that between 10% and 17% of depressed patients in primary care follow a chronic rather than episodic course of illness (Steinert et al., 2014). The concepts of relapse and recurrence are less easily applied to these patients, although this group of patients can include those with recurrent depression, with incomplete remission between episodes. Indeed, researchers have cautioned against excluding patients with chronic or persistent depression from studies of relapse and recurrence; the majority of these patients do in fact still remit or recover at times and could therefore be said to experience a relapse and recurrence, despite the longer duration of index episodes (Monroe and Harkness, 2011). In summary, the cohort of patients seen in primary care is different from those in secondary care. The studies discussed here have limitations that prevent us from making definitive statements about relapse rates in primary care settings, although it is evident that a significant number of people with depression in primary care do experience relapse.

2.2.3. Why do some people experience relapse?

Here, I discuss the underlying aetiology and mechanisms of relapse. It is unclear whether the same mechanisms are implicated in the different phenotypic groups we considered in the previous section (single episode, recurrent and chronic/persistent depression). There are several theories to explain why some people experience a relapse while others do not, and why a relapse may in turn increase the risk of a further episode: these include the diathesis-stress model, the kindling hypothesis and the scar theories.

The diathesis-stress hypothesis posits that causal factors can include both a biological vulnerability (diathesis) to depression, which is then precipitated in the individual by, for example, stressful life events (stress) (Monroe and Harkness, 2005, 2011). The kindling hypothesis suggests that

psychosocial and environmental stressors are more strongly implicated in the initial episode and that subsequent episodes are increasingly driven by endogenous factors, for example underlying genetic susceptibility (Post, 1992; Kendler, Thornton and Gardner, 2000; Bockting et al., 2015). This may be explained by "scarring" wherein the initial episode of depression changes something at a mind/brain level, making subsequent episodes more likely (Burcusa and Iacono, 2007). Possibilities include cognitive scarring (whereby depression activates negatively biased interpretations of experience, making a recurrence of depression more likely after a first episode); sensitisation to stressful life events (less stress is required in order to exert the same depressogenic effect in recurrent episodes); and psychosocial and personality scars (Burcusa and Iacono, 2007). Burcusa and Iacono (2007) noted there is only limited empirical evidence supporting these theories of recurrence. Furthermore, it seems likely that these models would be of only limited value to GPs, and there is no evidence exploring primary care perspectives or applications of the explanatory models discussed here.

2.3. Predicting relapse: prognosis research and prognostic models

So far, we have established that a significant number of people with depression in primary care will experience a relapse (although it is difficult to be very certain about depression relapse rates in primary care) and it is unclear precisely what causes some patients to experience relapse where others do not. This makes it challenging for GPs and other primary care professionals to identify higher-risk individuals and intervene to prevent relapse. If relapse and remission of depression could be reliably predicted at the individual patient-level, then resources could be better targeted towards relapse prevention of depression and support precision medicine, i.e., tailoring of intervention decisions conditional on an individual predicted risk and response to treatment (Riley et al., 2019b). This process requires prognosis research; specifically, the identification of prognostic factors and

the development, validation and impact evaluation of prognostic models for outcome risk prediction.

Prognosis refers to future outcomes given a particular baseline condition or disease and has been defined as "the forecast of future outcomes for people with a particular disease or health condition" (Riley et al., 2019b). The PROGnosis RESearch Strategy (PROGRESS) framework was developed in 2013 (Hemingway et al., 2013), and describes four main categories of prognosis research: overall prognosis; prognostic factor research; prognostic model research; and predictors of treatment effect. This thesis focuses on prognostic model research, and to a lesser extent prognostic factor research.

2.3.1. Can we predict relapse of depression?

2.3.1.1. Prognostic factors for relapse of depression

A prognostic factor is a variable that is associated with an increased risk of a future outcome. In contrast to prognostic models, which provide individualised risk prediction of particular outcomes conditional on multiple factors, prognostic factor studies generally focus on the factors themselves and whether they add (causal or prognostic) value over existing factors. In the UK, NICE guidance highlights a number of these to guide the identification of people who are, on average, at higher risk of depression relapse. These are people who: have had two or more episodes of depression; have a history of incomplete response to previous treatment or residual symptoms; have a history of severe depression; have other chronic physical or mental health problems; or have personal, social or economic factors that contributed to their depression and are still present (NICE, 2022).

The consensus view has long been that the two factors that most affect risk of relapse and recurrence of depression are residual depressive

symptoms (subthreshold symptoms of depression that persist once acute treatment has ended) and a prior history of recurrence (Campbell, 2009). Two systematic reviews and meta-analyses explored prognostic factors associated with relapse and recurrence of depression (Buckman et al., 2018; Wojnarowski et al., 2019). Buckman et al. (2018) performed a four-stage meta-synthesis which consisted of: an umbrella review (or meta-review) of 10 systematic reviews, a meta-analysis of 12 cohort studies, a meta-review of 27 non-systematic reviews and a systematic review of 20 experimental and neuroimaging studies. Wojnarowski et al. (2019) performed a systematic review of predictors of relapse and recurrence of depression after cognitive behavioural therapy, with a meta-analysis of five studies (n = 369).

This pre-existing evidence, drawn from evidence synthesis, has reported "strong evidence" that residual depressive symptoms are prognostic for relapse and recurrence, and "good" evidence that the number of previous episodes is associated with increased risk of relapse and recurrence (Buckman et al., 2018). In addition, the following factors are associated with relapse and recurrence: childhood maltreatment, comorbid anxiety (anxiety which is present at the same time as depression), neuroticism, younger age of first onset, rumination (the tendency towards excessive, repetitive thoughts which interferes with other mental processing) (Buckman et al., 2018), experiencing a higher number of dependent chronic stressors, or a severe independent life event post-treatment (Wojnarowski et al., 2019).

Individual participant data meta-analyses (IPDMA) have also been used to explore prognostic factors (Kuyken et al., 2016; Breedvelt et al., 2021a) and have been broadly in agreement, finding that younger age of onset, residual symptoms and a shorter duration of remission are associated with an increased risk of relapse. Previous research has also found a higher-odds of recurrence associated with both psychosocial impairment and poor coping skills, and that avoidant coping style and "daily hassles/life events" were predictive of recurrence (Hardeveld et al., 2010; Beshai et al., 2011).

Some of the clinical factors that the pre-existing literature has concluded do not appear to be predictive of relapse or recurrence include: insidiousness of onset; presence of precipitant (cause or trigger for current episode); previous treatment with tricyclics; history of hospitalisation; history of suicidal ideation or attempts; history of alcoholism or substance misuse; history of substance abuse; family history of depression; general level of functioning; biological functions and abnormal sleep patterns. Demographic factors lacking evidence supporting their role as prognostic factors for relapse or recurrence are: age, socioeconomic status, employment status, gender, marital status and intelligence (Buckman et al., 2018; Burcusa and Iacono, 2007; Thase et al., 1992; Wojnarowski et al., 2019).

2.3.1.2. Prognostic models for predicting relapse of depression

While a single prognostic factor can help refine the estimate of overall prognosis to particular subgroups, they are seldom sufficient to effectively aid risk-stratification at the individual level. Rather, individualised outcome prediction is better shaped by using multiple prognostic factors in combination, in the form of a multivariable prognostic model (Riley et al., 2019b). A multivariable prognostic model is a way (usually a mathematical equation) of combining information about multiple prognostic factors (hence multivariable) to produce an estimate of an individual's risk of developing a particular outcome in the future. Such risk prediction tools are increasingly recommended by policymakers and, in general practice, can be successfully built into IT systems (Riley et al., 2019b).

Reliable prediction of individuals' risk of relapse and recurrence might enable a precision medicine approach to relapse prevention, personalising the allocation and potentially type of relapse prevention interventions offered to ensure maximum benefit. A robust clinical tool to risk-stratify patients and then target relapse prevention interventions to those at increased risk would potentially be of significant benefit to patients, healthcare professionals and the NHS as a whole. A systematic review of existing prognostic models for

the intended population, outcome and setting is a recommended first step before considering the development of a novel prognostic model. If an existing model performs satisfactorily, adjusting this for the intended population (recalibration) and externally validating the model is likely to be a better use of resources than developing a model from the beginning (Riley et al., 2019b).

The predictive performance of a prognostic model can be measured in several ways which include: overall measures of model fit (for example R^2, which measures explained variation for models with continuous outcomes, or generalisations of R^2 for models with binary or time-to-event outcomes); calibration (which measures the extent to which risk predictions and observed outcomes are in agreement); and discrimination [the model's ability to separate patients who develop the outcome of interest and those who do not, usually measured using the Concordance (C-) statistic or area under the curve (AUC)]. Clinical utility is also important to consider when a model's predicted risks are to be used to inform decision-making. This can be measured by the net benefit at a particular risk threshold, and by plotting decision curves of the net-benefit across a range of relevant thresholds (Vickers, Van Calster and Steyerberg, 2016).

There have been some attempts to derive and validate prognostic models to predict depression-related outcomes (Angstman et al., 2017; King et al., 2010; Rubenstein et al., 2007; van Bronswijk et al., 2019). In an initial scoping of the literature, I identified only one model developed to predict risk of recurrence of depression over three years (C-statistic of 0.72 on external validation; confidence interval not reported) (Wang et al., 2014). There has been no previous systematic review to identify all prognostic models designed to predict relapse or recurrence of depression.

2.3.2. Should we predict relapse of depression?

In addition to whether a prognostic model can accurately predict an outcome in a generalizable way, other considerations include whether the

results of risk predictions can be used and shared in a clear and helpful manner and result in improved outcomes or lower costs when applied. To be useful in practice, prognostic models must include unambiguous prognostic factors, address a common and important problem and have face validity (doctors must trust a model to guide their practice rather than their own experience) (Riley et al., 2019b).

It is possible that a statistical prediction tool for relapse of depression might align too closely with a biomedical model of depression that does not fully describe the course of depression in many patients. However, due to limitations imposed by the healthcare system, such as workforce and workload challenges (Gopal and Mulla, 2020), decreased continuity (Royal College of General Practitioners, 2016; Murphy et al., 2021), and remote working (Murphy et al., 2021), GPs must gather and synthesise information to aid clinical decision-making in a relatively short amount of time. A prognostic model could facilitate the identification and stratification of these different risk groups and, longer-term, might enable more effective and efficient use of resources. The views and preferences of patients, healthcare professionals, commissioners and policymakers need to be more robustly explored. Provided that stakeholder perspectives are considered and used to aid implementation, there are grounds for thinking that risk prediction has a role to play in addressing this clinical problem.

2.4. Preventing relapse of depression in primary care

2.4.1. Overview

This section considers the evidence for relapse prevention of depression. The vast majority of studies of relapse prevention for depression are in secondary care settings. There has been only one previous systematic review of (non-pharmacological) relapse prevention strategies in a primary care setting (Gili et al., 2015); this review identified only three studies and the

authors were unable to draw firm conclusions. There are three recognised treatment phases for depression: treatments implemented before any symptomatic improvement, with a view to achieving remission (acute phase); those employed after symptomatic improvement but before recovery (continuation phase); and those that extend past the point of recovery (maintenance phase) (Bockting et al., 2015). Interventions to prevent relapse might be targeted at patients who are in the continuation or maintenance phases, having had symptomatic improvement, or might be implemented during the acute phase, with the intention of exerting a protective effect against relapse in the future (Bockting et al., 2015).

Interventions for depression can be broadly considered as either pharmacological (antidepressant medication) and psychological. Pharmacological and psychological interventions are comparable in terms of efficacy for acute phase depression (Cuijpers et al., 2013b; Kamenov et al., 2016; Cuijpers et al., 2020). Combination therapy is superior to either type of intervention given singly for people with moderate or severe depression (Cuijpers et al., 2020). For these patients, combination therapy has been shown to lead to improvements in quality of life and functioning as well as initial treatment response (Kamenov et al., 2016).

Interventions for preventing relapse of depression, as for acute phase interventions, are also principally pharmacological or psychological. Aside from psychological therapies, other kinds of non-pharmacological therapies also exist. These include social interventions [for example, social prescribing (Drinkwater, Wildman and Moffatt, 2019)], service-level interventions [for example, collaborative care (Gunn et al., 2006)], and combination interventions. Physical interventions such as electroconvulsive treatment (ECT) can be used to prevent relapse in some patients, although ECT is almost exclusively a secondary care intervention, evidence for its efficacy is of low quality and there are concerns around harm caused by ECT (NICE, 2022).

2.4.2. Antidepressant medication

There is reasonable evidence that, compared with placebo, antidepressant medications prevent depression relapse; there do not appear to be major differences in this effect across different antidepressants, or even different classes of antidepressants (Geddes et al., 2003; Glue et al., 2010; Hansen et al., 2008; Kaymaz et al., 2008). A systematic review found that, compared to placebo, continuation of antidepressants for at least six months after remission significantly improved relapse rates, and that this was true for selective serotonin reuptake inhibitors (SSRIs), tricyclic antidepressants (TCAs) and other newer agents (Kato et al., 2020). Absolute benefits from antidepressants are greater for those at higher initial risk of relapse (Geddes et al., 2003), and the effect may decrease after a number of depressive episodes, as patients appear to develop resistance to their prophylactic properties (Kaymaz et al., 2008). There are some limitations in the available literature around the effect of antidepressants on relapse, which prevent us from drawing firm conclusions about their effects in primary care populations. The studies included in existing reviews have all been undertaken in secondary care populations, where outcomes are generally poorer. The included studies, on the whole, have also not adjusted for potential withdrawal (discontinuation) effects upon stopping antidepressants which are likely to be a significant confounding factor when estimating rates of relapse (Recalt and Cohen, 2019; Cohen and Recalt, 2019; Hengartner, 2020).

A more recent primary care based RCT examined relapse rates in people stopping antidepressants (Lewis et al., 2021). In the study, participants with a history of relapse, who had been on common antidepressants for two years or more and who were willing to stop their medication, were assigned to either continue or discontinue their antidepressants. People who discontinued antidepressants experienced higher rates of relapse (56%) than those who continued medication (39%), most commonly 12 to 26 weeks after the study started. The lower rate of relapse in the continuation group could not be explained by the placebo effect as this was a double-blinded study

(dummy pills, with tapering doses of the participants' original antidepressants, were used for the discontinuation group). The findings do not necessarily translate to people who have had only one episode of depression, and there was a lack of ethnic diversity in the study sample (participants were mainly White). The findings do, however, suggest that antidepressants continue to confer a prophylactic effect against relapse for some people with remitted depression. They also suggest that approximately 40% of people could discontinue their longer-term antidepressants without experiencing relapse, but it is not yet clear how we identify who will remain well and who will relapse after discontinuing.

Antidepressants are not without their drawbacks:

- adverse effects are common (including sleep disturbance, gastrointestinal disturbance and sexual dysfunction, among others) (van Leeuwen et al., 2020), and can be worse in older adults (Coupland et al., 2011);

- concerns about 'dependence' and discontinuation symptoms (which can often be confused with relapse, although these generally occur much sooner after antidepressant discontinuation than relapse, and can be severe and long-lasting) (Davies and Read, 2019; van Leeuwen et al., 2020);

- reluctance by patients to take medication (Goldman et al., 1999);

- poor concordance (i.e., not taking the medication as prescribed or recommended) (Badger and Nolan, 2013);

- and high financial cost to the health service (Maund et al., 2019), although this is explained by the volume of prescriptions rather than the cost of the medications themselves, which are relatively inexpensive (Moore et al., 2009).

The other problem with antidepressants is that any prophylactic or protective effects disappear when the medication is discontinued (Bockting et al., 2015). The effect from some types of psychological therapy, like cognitive

behavioural therapy (CBT) and problem-solving therapy, can endure beyond the point at which the therapy is actually being delivered, provided people use the skills and strategies learned as a result (Cuijpers et al., 2013a; Vittengl et al., 2007).

Antidepressants primarily exert their pharmacological effects on the serotonergic, dopaminergic and noradrenergic pathways (Healy, 1997). While these neurotransmitters are likely to be implicated in some patients with depression, and modification of these pathways is a sound theoretical basis for explaining some of the effectiveness of antidepressants at preventing relapse, it is widely accepted that depression is multifactorial. There has been research challenging the "serotonin hypothesis" of depression (the theory that depression is caused by abnormally low levels of the neurotransmitter serotonin, and that increasing those levels by inhibiting reuptake is the mechanism by which antidepressant medications work) (Moncrieff et al., 2022; Ang, Horowitz and Moncrieff, 2022). While GPs still use this explanation as part of their discussions with patients (Read et al., 2020), the literature suggests that GPs increasingly avoid giving simple biological explanations to patients when explaining the effect of antidepressant medications (Tickell et al., 2020).

A comprehensive overview of the neurobiological processes underlying the mechanism of action of antidepressants is beyond the scope of this thesis. The evidence demonstrates that antidepressants are effective, even though the mechanism of action remains uncertain and is probably not fully explained by reference to specific neurochemical pathways (Kendrick and Collinson, 2022). One mechanism by which antidepressant medications are thought to prevent relapse is by reducing the presence of residual depressive symptoms (those that are sub-threshold and persist after remission), which are strongly associated with increased risk of relapse (Buckman et al., 2018; Lin et al., 1998).

2.4.3. Psychological therapies

Psychological therapies for preventing relapse include CBT, mindfulness-based cognitive therapy (MBCT), and interpersonal therapy (IPT) (Clarke et al., 2015). CBT aims to modify thoughts and behaviours, such as reducing avoidance, increasing the time spent in pleasurable or rewarding activities and challenging negative thoughts. CBT also involves the teaching of cognitive skills which focus on challenging underlying dysfunctional beliefs (cognitive content) that can persist after remission or recovery despite a non-depressed state, presenting a vulnerability that might be more easily triggered by, for example, life events or stress (Beshai et al., 2011; Bockting et al., 2015; Clarke et al., 2015). MBCT was developed specifically as an intervention to prevent relapse and places greater emphasis on cognitive processes that cognitive content (Kuyken et al., 2015). It includes meditation techniques to help people become aware of their experiences in the present moment, teaches patients to experience thoughts without judgement and to recognise that negative thoughts are transient and do not have to guide feelings or behaviours (Williams and Kuyken, 2012). IPT focuses on interpersonal and societal role problems, which can be implicated in the onset and recurrence of depressive symptoms (Clarke et al., 2015).

Psychological therapies exert their effect by modifying a broader range of therapeutic targets than antidepressants. Most psychological therapies for preventing relapse are designed to occur during the continuation or maintenance treatment phases; although, as discussed, these can occur during the acute phase and exert a longer-term benefit. As well as aiming to reduce residual depressive symptoms (like antidepressants do), psychological therapies additionally target cognitive and information processing mechanisms (specifically those involved in integrating affective and cognitive information; processing negatively valanced stimuli; social skills and the ability to use social support; problem solving skills; and degree of negative self-concept) and interpersonal stress pathways. Relapse of

depression is associated with negative thinking styles, such as rumination (Buckman et al., 2018). CBT specifically targets these thoughts and aims to educate patients on how to modify and transform such thoughts into more adaptive thoughts (Clarke et al., 2015). MBCT focusses on teaching patients to improve their awareness of and relationship to such thoughts, rather than on modifying the thoughts (Kuyken et al., 2015). Psychological therapies also focus on being aware of and planning for early warning signs of relapse, and also focus on healthy lifestyle behaviours. In summary, there is a range of mechanisms by which psychological interventions might work to improve relapse-related outcomes.

2.4.4. Relapse prevention in primary care

Studies of relapse prevention specifically in a primary care have included MBCT (Kuyken et al., 2008, 2015; Ma & Teasdale, 2004), collaborative care approaches (Howell et al., 2008; Katon et al., 2001), counsellor-led supportive self-help (Biesheuvel-Leliefeld et al., 2017), and preventative cognitive behavioural therapy (Bockting et al., 2018; De Graaf et al., 2011). MBCT has been found to be effective for depression relapse prevention in a primary care setting (Kuyken et al., 2015), and is particularly effective for those with residual symptoms (Kuyken et al., 2016). Given the focus on meditation techniques, it is not likely to be an acceptable intervention for all patients and, as it requires a greater degree of training for the person delivering the intervention, it may not be feasible for all primary care settings (Williams and Kuyken, 2012).

A collaborative care-based relapse prevention intervention including patient education and proactive telephone monitoring increased medication adherence and decreased depressive symptoms overall, although did not reduce relapse rates (Katon et al., 2001). A counsellor-led supportive self-help relapse prevention intervention, taking place over eight weeks, did reduce relapse rates (Biesheuvel-Leliefeld et al., 2017), but was not considered cost-effective (Biesheuvel-Leliefeld et al., 2018). Preventative

cognitive therapy was also found to be effective for reducing relapse in a primary care setting (Bockting et al., 2018). In summary, the evidence is limited for relapse prevention depression in a primary care setting but there have been some promising studies.

2.4.5. What do the guidelines say?

The NICE guideline for depression, updated in 2022, recommends that patients starting antidepressant medication for depression should continue treatment for a minimum of six months after remission to reduce the risk of relapse. People "with a higher likelihood of relapse" (i.e., people who have one of the prognostic factors listed by NICE and outlined earlier in this chapter) are advised to continue antidepressant medication to prevent relapse, until there is good reason to reduce it. Recommendations for relapse prevention psychological therapies are group CBT or MBCT, with an explicit focus on the development of relapse prevention skills, for people who do not wish to continue antidepressant medication (NICE, 2022). In more severe or refractory cases, patients are usually referred for specialist mental health assessment.

The Canadian Network for Mood and Anxiety Treatments (CANMAT) depression guideline concurs with NICE that evidence to support maintenance therapy for longer than two years is less well-established, but that certain risk factors (early depression onset, ongoing psychosocial adversity, older age, and comorbid medical or psychiatric conditions) might justify extended maintenance antidepressant treatment (Lam et al., 2009). The American Psychiatric Association (APA) guideline recommends continuation treatment with antidepressant medication and depression-focussed psychological treatments. For those with three or more previous depressive episodes, chronic depression or other risk factors (including family history of mood disorder), the APA advises maintenance treatment with the medication that produced and maintained remission during the acute

and continuation phases, and for ECT to be considered beyond that. They recommend regular monitoring for "signs of relapse" throughout (Van Kempen et al., 2010).

2.5. Relapse prevention: a review of the qualitative literature

Improving risk-stratification and the targeted allocation of relapse prevention interventions in primary care will involve discussion with patients about the risk of relapse. For some patients, it is more appropriate to frame depression as a potentially chronic, on-going illness to be managed, rather than something that can be "cured". Do patients want to have these discussions and is relapse something that concerns people with a lived experience of depression? Are such discussions required for all patients following a first episode of depression? How do clinicians decide when to adopt a chronic disease model of depression management and for which people aiming towards a more definitive treatment might be appropriate? Patient expectations and understanding may affect outcomes and so these are important questions to consider.

There is some pre-existing evidence on this subject, and previous qualitative research has explored patients' understanding of relapse, the extent to which this is a concern for them, and their experiences within a primary care context. A qualitative study, which was embedded in the PREVENT RCT of MBCT for relapse prevention (Kuyken et al., 2015), explored the views of patients with recurrent depression (Tickell et al., 2020). This qualitative study specifically addressed the role of GPs and found that participants felt supported when they felt able to make an appointment with their GP when needed, whereas they felt unsupported when they found it "very difficult" to access the GP. In addition, the authors reported that people with depression often feel disheartened when their GPs advise antidepressant medication in response to relapse or do not show sufficient

interest in psychological approaches such as MBCT (Tickell et al., 2020). This latter point should be interpreted in the context that the qualitative work was undertaken in a cohort of people who had chosen to take part in a trial of MBCT.

Other pre-existing research addressing patient preferences has been in the context of discussions around antidepressants, with fear of relapse recognised as a barrier to patients discontinuing antidepressant medication (Maund et al., 2019; Bowers et al., 2020; van Leeuwen et al., 2022) and some patients confusing relapse with discontinuation symptoms (Leydon, Rodgers and Kendrick, 2007). Research has also shown that patients may not have full confidence in the GPs' ability to discuss discontinuation of antidepressants due to a perceived lack of knowledge and time (Bosman et al., 2016). Interestingly, GPs felt that they did have sufficient knowledge to manage continuation therapy and would be more inclined to continue antidepressant medication in patients with a history of relapse (Bosman et al., 2016). They did agree, however, that time-limited consultations and a lack of evidence-based guidance on long-term depression management resulted in some patients being sub-optimally managed (Bosman et al., 2016).

A discrete choice experiment (DCE) among 109 people with partially or fully remitted depression or anxiety explored patient preferences around components to be included in a relapse prevention intervention (Muntingh et al., 2019). The participants in this study were recruited from outpatient mental health clinics in the Netherlands, and therefore are not directly applicable to a UK primary care population. They did, however, have anxiety or unipolar depression, with severe mental illness more typical of NHS secondary care excluded, and therefore the findings are relevant. Participants reported that high effectiveness (defined as reduced relapse risk), regular contact with a health professional, low time investment and the inclusion of a personal prevention plan were priorities. Interestingly, people of younger age generally valued effectiveness more than older participants and

those with previous episodes had a greater preference for more regular contact with a health care professional. Treatment modality appeared to be less of a concern for the participants surveyed. This DCE study involved a relatively small number of participants choosing between two or more hypothetical interventions, using a multiple-choice format, and did not explore underlying reasons and meanings in detail using more interpretative qualitative methods, leaving a gap in the literature. The qualitative part of this mixed methods study will explore these topics in more detail, with a focus on discussion of risk, preferences around interventions and patient and GP perspectives on relapse.

2.6. Summary

This chapter has defined and explored the scope of the problem of depression relapse in primary care; the role of prognosis research and how a multivariable prognostic model could help to risk stratify patients; the pre-existing evidence around relapse prevention interventions in primary care; and discussed the need for an increased understanding of patient and GP perspectives in this area. The thesis will now follow a format of presenting work undertaken to address these needs. As explained earlier in this chapter, a review of existing prognostic models is advised before attempting to develop a new model. For this reason, the next chapter presents a systematic review and critical appraisal of prognostic models for predicting relapse or recurrence of depression, to guide the subsequent research.

Chapter Three

Systematic review of prognostic models for predicting relapse or recurrence of depression

3.1. Introduction

This chapter presents a systematic review, undertaken with support from the Cochrane Common Mental Disorders Group at the University of York, and the Cochrane Prognosis Methods Group[5]. I attended the funded course, "Systematic reviews and meta-analysis of prognosis studies", which was run by the Cochrane Prognosis Methods Group in Utrecht, in July 2019 as part of my training plan to undertake this review.

The chapter incorporates content from the two peer-reviewed papers resulting from this work, published as part of my NIHR Doctoral Research Fellowship. The review was published as a Cochrane Prognosis Review (Moriarty et al., 2021c) in May 2021 and, following an updated literature search, as a co-publication in the British Journal of Psychiatry "Precision Medicine and Personalised Healthcare in Psychiatry" special issue in January 2022 (Moriarty et al., 2022). All necessary permissions and licences have been obtained from the publishers prior to reproducing the content in this thesis (Appendices 2.2 and 2.3).

Protocol preregistration: Cochrane Database of Systematic Reviews and PROSPERO; CD013491; doi:10.1002/14651858.CD013491.pub2.

---

[5] I conceived and led the review; wrote the protocol and final article; undertook the database search, study selection, risk of bias and applicability assessment, and data extraction; led the analysis and wrote the discussion. A second reviewer duplicated the study selection, risk of bias and applicability assessment and data extraction to satisfy Cochrane review standards.

3.2. Objectives of review

To identify and critically appraise prognostic model development and validation studies aimed at predicting relapse, recurrence, sustained remission or recovery in adults with major depressive disorder who meet the criteria for remission or recovery. In addition, I planned to summarise and meta-analyse their predictive performance, to describe the characteristics of the models identified, and to review the clinical utility (net benefit) of the identified models, where possible.

3.3. Methods

The protocol was preregistered in the Cochrane Database of Systematic Reviews (CD013491) (Moriarty et al., 2019) and is reported in line with the Preferred Reported Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline (Page et al., 2021). See Appendix 3.1 for a completed PRISMA Checklist.

3.3.1. Eligibility criteria

I specified the following eligibility criteria [see Table 3.1 for PICOTS criteria (Debray et al., 2017)]:

3.3.1.1. Population

Adult population (18 years and over) with major depressive disorder (defined using validated diagnostic criteria) who met criteria for remission or recovery (i.e., no longer meeting diagnostic criteria for major depressive episode) at the point of prediction. I excluded models developed in populations with comorbid severe mental illness (for example, schizophrenia and bipolar affective disorder), as these patients typically receive more intensive psychiatric input and the results would be less generalisable. I

excluded people below 18 years old, as children with depressive disorders are treated in very different settings with different practitioners and follow-up schedules, and are likely to have meaningfully different predictors from adults. I planned to include older adults, being mindful that multimorbidity may be more common in the older population and may impact on depression outcomes in this population, more so than in a general adult population.

3.3.1.2. Prognostic models (index and comparator)

All multivariable prognostic models developed to predict individual risk of relapse, recurrence, sustained remission, or recovery of depression over any time period. The included models had to have been developed with the intention of providing individualised risk predictions (binary or time-to-event outcomes) and I excluded papers reporting multivariable models not intended for this purpose. I also planned to include models predicting outcomes on a continuous scale if these had been identified, provided they met the other inclusion criteria (i.e., remitted major depressive disorder at start-point). I did not specify a comparator prognostic model.

There are three types of prognostic model study (Wolff et al., 2019):

- Prediction model development without external validation: these studies aim to identify important predictors of the outcome of interest, assign weights (usually in the form of regression coefficients) to each predictor during multivariable analysis, develop a prediction model for individualised risk predictions and quantify the model's predictive performance in the development set. They should use internal validation techniques to adjust for optimism and reduce overfitting;
- Prediction model development with external validation: these studies undertake the development steps as described previously and then attempt to quantify the model's performance in data external to the development data;
- Prediction model external validation studies: attempt to externally validate an existing prediction model.

I included all model development and validation (internal and external) studies, including those that updated existing models (i.e., extended or modified existing models with new predictor information). While external validation is described as the "evaluation of performance in data that were not used to develop the model" (Collins et al., 2014), this does not generally mean a random split of the development dataset to produce two separate datasets. This approach is best considered an inefficient form of internal validation (Riley et al., 2019b). External validation can, however, be performed in a dataset produced by a non-random split, for example participants from the same institution but at different time points (temporal validation) or by location (geographical validation) (Collins et al., 2014; Moons et al., 2012). I included these as examples of external validation studies for the purpose of this review. If a sufficient number of external validation studies were identified for a particular model, I planned to perform a meta-analysis to provide a quantitative summary of that model's predictive performance. I planned to treat updated models as separate models for the purposes of meta-analysis.

Eligible studies included those that developed prognostic models using data from cohort studies (prospective and retrospective, including registries and cohorts from randomised controlled trial data) and any other sources of data if they meet the other inclusion criteria. Reports of impact assessments of prognostic models (studies that assess the impacts of the models when translated and implemented into practice, for example in randomised trials) were not included in this review, as these studies require different methodology. I did not include prognostic factor studies, which generally examine the adjusted association of prognostic factors on risk of relapse or recurrence (generally in the form of relative risk ratios or odds ratios) but do not derive a multivariable prognostic model to calculate individualised risk of outcome (Riley et al., 2019b).

3.3.1.3. Outcomes

As outlined in Chapter Two, remission and recovery are terms used to describe an improvement in depressive symptoms; remission meaning improved but still 'in episode' and recovery being the resolution of the underlying episode (usually after 6 to 12 months of remission) (Bockting et al., 2015). Relapse occurs following some level of remission but precedes recovery, while recurrence is the onset of a new episode of depression following recovery (Frank et al., 1991; Rush et al., 2006). Sustained remission can be thought of as the inverse, or opposite of relapse; and recovery as the inverse of recurrence. Both hold potentially valuable prognostic information pertinent to relapse risk prediction models in depression, and are therefore included as outcomes in this review. The precise temporal cut-offs are inconsistently operationalized in the literature (Buckman et al., 2018). For this reason, I accepted all definitions of these terms, as operationalized by the authors of the primary studies.

I did not include models that predict sustained depressive symptoms, as these models require a different population (i.e., those who have been diagnosed as depressed and continue to experience symptoms rather than those with depression who have subsequently entered remission).

3.3.1.4. Timing

The starting point of prediction is when a person with depression has responded to treatment and meets criteria for remission and I accepted models that predicted outcomes over any time period.

3.3.1.5. Setting

I included models developed in any setting (primary, secondary, or community care) for this review.

Table 3.1: Summary of PICOTS Criteria

P	Population	Adult patients (18 years and over) diagnosed with depression and meeting criteria for remission
I	Index prognostic model	All prognostic models predicting relapse, recurrence, sustained remission or recovery in patients with remitted depression
C	Comparator prognostic model	None
O	Outcomes	Relapse, recurrence, sustained remission or recovery in depression
T	Timing	Start-point: the point at which a patient has responded to treatment and is identified as meeting criteria for remission
S	Setting	Any setting (primary, secondary or community care)

3.3.2. Information sources and search strategy

I consulted with an Information Specialist to develop the search strategy. The following bibliographic databases were searched, using relevant subject headings (controlled vocabularies) and search syntax, appropriate to each resource: the Cochrane Library (current issue); Ovid MEDLINE (1946 onwards); Ovid Embase (1980 onwards); and Ovid PsycINFO (1806 onwards), up to May 2021. I also searched several grey literature resources primarily for dissertations and theses (Open Grey (www.opengrey.eu); ProQuest Dissertations & Theses Global (www.proquest.com/products-services/pqdtglobal.html); DART-Europe E-theses Portal (www.dart-europe.eu); EThOS - the British Libraries e-theses online service (ethos.bl.uk); Open Access Theses and Dissertations (oatd.org)), also up to May 2021. I applied no restrictions by date, language, or publication status. I checked the reference lists of all included articles and conducted forward citation searches on the Web of Science (12 March 2021 and 19 May 2021), to identify additional studies missed from the original

electronic searches (e.g., unpublished or in-press citations). I contacted corresponding authors for information on unpublished or ongoing studies.

3.3.3. Selection process

Two review authors (ASM and NM) independently reviewed the titles and abstracts of studies identified by the search strategy. We excluded prognostic model studies that clearly did not meet the inclusion criteria at the title and abstract screening stage. For any studies where there was uncertainty, we undertook a full-text review. We resolved disagreement in judgements through discussion or, if necessary, by referral to a third review author (KIES or DM).

3.3.4. Data collection process and items

Two independent review authors (ASM and NM) conducted the data extraction. The Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) has been developed to guide data extraction in systematic reviews of prognostic models, and was used for this review. We extracted the following data for all included studies:

- method of depression diagnosis;
- year of participant recruitment and follow-up;
- setting;
- source of data;
- participants' characteristics;
- study design;
- definition of relapse and recurrence;
- information on number and type of candidate predictors;
- sample size;
- number of events;
- missing data;

- type of model used for development (e.g., logistic regression, Cox regression, machine learning, neural network) and any adjustment for model overfitting (e.g., using penalisation or shrinkage techniques);
- model performance: we extracted information about the models' predictive performance, in terms of discrimination (C-statistic) and calibration (calibration slope, ratio of observed (O) to expected (E) events (O:E ratio), calibration plots), including optimism-adjusted estimates in the development data. Calibration (preferably a calibration plot) and discrimination (C-statistic) should be reported, at a minimum. A C-statistic of 1 indicates that a model has perfect discrimination while a C-statistic of 0.5 means that the model performs no better than chance (Riley et al., 2019b);
- model evaluation: whether internal and external validation were done, whether optimism-adjusted measures were reported from internal validation, model updating in case of poor validation;
- results: interpretation and discussion of generalisability, strengths and weaknesses;
- clinical utility: usually assessed through net benefit analysis (Vickers, Van Calster and Steyerberg, 2016), a means of progressing beyond the predictive performance of the developed model and considering its implementation and impact in a healthcare setting, usually using decision analytic techniques. We describe this for included studies where it has been reported.

We also collected information on how the model was presented (risk chart, nomogram, full regression formula) and whether it is possible to use a model based on the information presented in the article. Where measures of predictive performance were not available directly, I planned to calculate these from other information available with reference to recent guidance (Debray et al., 2019).

3.3.5. Data synthesis and planned meta-analysis approach

If there were enough studies reporting external validation performance, I planned to conduct random-effects meta-analyses to summarise performance of prognostic models, as data were likely to be highly heterogeneous. I aimed to pool information about each model's discrimination (using C-statistic or equivalent), calibration (using calibration slope, calibration-in-the-large; and O:E ratio) and equivalents from time-to-event models (e.g., Harrell's C-statistic, calibration slope, D statistic, O:E at each time point). I planned to summarise performance measures separately, first transforming them to an appropriate scale where necessary (logit C-statistic and log O:E ratio) to produce summary results (with 95% confidence intervals (CIs)) that quantified the average performance across studies (Snell et al., 2018). To better account for the uncertainty in the estimated between-study heterogeneity, I planned to use the restricted maximum likelihood (REML) estimation, with 95% CIs for the summary (average) performance of a model, derived using the Hartung-Knapp-Sidik-Jonkmann method (Debray et al., 2017; Langan et al., 2019). If there were insufficient data for a meta-analysis, I planned to use a narrative synthesis instead.

3.3.5.1. Subgroup analysis and investigation of heterogeneity

I planned that, if there were sufficient data (a minimum of 10 studies), I would investigate potential sources of heterogeneity using meta-regression with the summary estimate of model performance (e.g., logit C-statistic or log O:E ratio) as a dependent variable and study-level covariates (population/case-mix (age of participants and multimorbidity), study setting of models (primary and secondary care settings) and study design (follow-up time, source of data, outcome definition and sample size)) as explanatory variables.

3.3.5.2. Sensitivity analysis

If I had sufficient studies for meta-analysis, I planned to evaluate the impact of risks of bias by conducting analyses only including studies assessed at low risk of bias.

3.3.5.3. Dealing with missing data

When performance measures (such as C-statistic, O:E ratio) were not reported in the paper, I contacted authors. Where possible, I planned to use standard methods and formulae described by Debray and colleagues to estimate the O:E ratio and C-statistic and associated standard errors (Debray et al., 2017).

3.3.5.4. Assessment of heterogeneity

Reviews of prognostic studies often have to deal with a substantial amount of heterogeneity. I planned to assess the impact of heterogeneity in predictive performance across validation studies, where there were enough data to do so, by calculating prediction intervals that provide a range for the potential performance of a model in a new validation study (Debray et al., 2017). I also planned to calculate I^2 and Tau^2 statistics. If reported, I would have extracted performance in subgroups.

3.3.6. Risk of bias assessment in included studies

Risk of bias (assessed as low, high or unclear) relates to the ability of the primary study to answer its own question and whether shortcomings in the methods used mean that the authors' conclusions lack internal validity, with the predictive accuracy of the model likely to be distorted (Wolff et al.,

2019). Applicability (also assessed as low, high, or unclear concern about applicability) refers to the extent to which the primary study is relevant to the systematic review criteria (how well the study meets the inclusion criteria *of the review*). Two independent review authors (ASM and NM) assessed risk of bias using the Prediction model risk of bias assessment tool (PROBAST), which assesses risk of bias over four domains, as well as applicability (Wolff et al., 2019; Moons et al., 2019; Riley et al., 2019b):

- Participants: this domain assesses whether appropriate data sources and inclusion/exclusion criteria were used;
- Predictors: assesses whether predictors were defined and assessed in a similar way for all participants; assessed without knowledge of outcomes; and available at the time at which the model is intended for use;
- Outcomes: assesses whether outcomes were determined appropriately; whether they were prespecified; whether predictors were excluded from outcome definition; whether they were defined and determined in a similar way for all participants; whether they were determined without knowledge of predictors; and whether there was an appropriate time interval between predictor assessment and outcome determination;
- Analysis: assesses whether there was a reasonable number of participants with the outcome; whether there was appropriate handling of continuous and categorical predictors; whether all enrolled participants were included in the analysis; whether missing data were handled appropriately; whether relevant model performance measures were presented; whether overfitting and optimism in performance were accounted for; and whether predictors and assigned weights in the final model correspond to results from multivariable analysis.

I will report how the included studies performed later in this chapter. Here, I will expand on some aspects of the 'Analysis' domain and how I applied this when making judgements in this review. Predictor selection is an important part of prognostic model development and occurs in two stages: selecting predictors for consideration in the model (candidate predictors) and selecting predictors during model development (predictors in final model).

When using regression analysis, selection of candidate predictors is best done on robust *a priori* grounds and usually following a literature search or clinical consensus, or both (Riley et al., 2019b). When selecting predictors for inclusion in the final model, it is recommended that statistical significance on univariable analysis between a candidate predictor and the outcome of interest is avoided as a method of selection. Forward selection is also best avoided. These approaches risk overfitting the model to the development dataset and excluding important predictors from the final model. Recommended approaches include fitting the full model (including all predictors felt to be important either clinically or based on the literature, regardless of statistical significance), using variable selection using backward selection (all predictors included and those found not to be statistically significant are excluded in a stepwise manner, with internal validation to then apply shrinkage to deal with overfitting) (Riley et al., 2019b), or penalised regression such as the LASSO or elastic net.

When determining whether an appropriate sample size was used, I adhered to PROBAST recommendations, which use the rule of thumb using events per predictor parameter (EPP). The PROBAST guidance suggests an EPP of 20 and over for development studies (although those between 10 and 20 EPP can be rated 'probably yes' or 'probably no', depending on outcome frequency, overall model performance and distribution of predictors in the model), and that validation studies must have at least 100 participants with the outcome and 100 without the outcome. EPP refers to the number of candidate predictors rather than just those included in the final model. Specifying the number of parameters rather than the number of predictors takes into account whether there have been any transformations of continuous variables (e.g., when checking for correct functional form) and indicator variables for categorical predictors with multiple categories and interactions.

Because prognostic models are often developed on data collected for a different purpose, missing data are common. A complete-case analysis to compensate for missing data is not generally recommended (unless there is very little missing), due to waste of valuable data. There are several more acceptable ways of accounting for missing data. Multiple imputation is considered more appropriate when data are missing at random (Riley et al., 2019b) and is recommended by PROBAST (Moons et al., 2019).

The PROBAST tool has been developed primarily for studies that used a more traditional regression method and guidance on best practice for machine learning (ML) models is less widely available. There is debate over the minimum number of EPP required, with guidance stating between 10 and 50 required for model development using classical modelling techniques, such as logistic regression. The guidance and literature that does exist would suggest that we should demand, if anything, significantly larger sample sizes when using a ML approach to prognostic model development, with one paper estimating that one would need more than 10 times the EPP required for regression models to achieve a stable area under the curve (AUC) and small optimism (van Der Ploeg, Austin and Steyerberg, 2014). Another suggestion is that prediction models developed using ML techniques require EPP of more than 200 to avoid overfitting (Wolff et al., 2019). In the case of any ML models identified, I applied the PROBAST guidance as described for traditional regression techniques, but judgements should be interpreted with these limitations in mind.

3.3.7. Certainty assessment
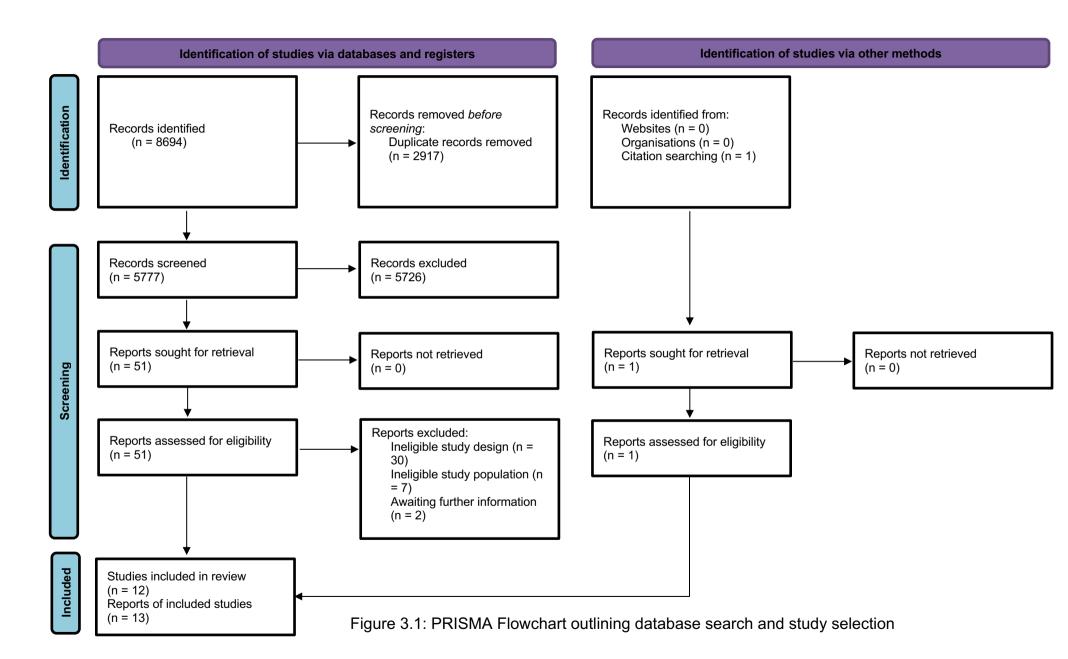
The GRADE system was developed to guide the interpretation of certainty (or confidence) in the results of intervention reviews. GRADE assesses the overall certainty of evidence for the estimate of effect by addressing the domains of: risk of bias, inconsistency, imprecision,

indirectness and publication bias. GRADE can be applied to some prognosis reviews, with proposed extensions available for reviews of overall prognosis (Iorio et al., 2015) and prognostic factors (Foroutan et al., 2020; Huguet et al., 2013). As discussed, heterogeneity is more likely and might be more acceptable in reviews of prognostic model and factor studies due to the inevitable study differences in methods of measurement, adjustment factors and statistical analysis methods, amongst others. Publication bias is also likely to be more severe in prognosis reviews than in those of intervention studies. Due to incomplete guidance on application of GRADE to prognostic model reviews, I did not conduct GRADE assessments for this review. I have focused on risk of bias (using PROBAST) to guide our assessment of the certainty of the evidence.

3.4. Results

3.4.1. Results of the search

I identified a total of 8694 studies initially, with one study located through a forward citation search performed on 12 March 2021 (van Loo et al., 2020). Deduplicated records (n=5777) records underwent title and abstract screening by two independent review authors (ASM and NM), 51 underwent full-text screening, and 12 studies were included in the final review. These included 11 unique prognostic models; one of the studies (van Loo et al., 2020) externally validated a model developed elsewhere (van Loo et al., 2018). Studies excluded after full-text screening (n=37) fell into two categories: not meeting study design criteria (i.e., model not intended for prediction) or not meeting participant population criteria. Two studies (awaiting further information) were conference proceedings; I was unable to obtain further information on these studies and so did not include them in the review (Trivedi et al., 2016; Cohen et al., 2021) (Figure 3.1).

Figure 3.1: PRISMA Flowchart outlining database search and study selection

Identification of studies via databases and registers

Identification of studies via other methods

Identification

Records identified
(n = 8694)

Records removed *before screening*:
Duplicate records removed
(n = 2917)

Records identified from:
Websites (n = 0)
Organisations (n = 0)
Citation searching (n = 1)

Screening

Records screened
(n = 5777)

Records excluded
(n = 5726)

Reports sought for retrieval
(n = 51)

Reports not retrieved
(n = 0)

Reports sought for retrieval
(n = 1)

Reports not retrieved
(n = 0)

Reports assessed for eligibility
(n = 51)

Reports excluded:
Ineligible study design (n = 30)
Ineligible study population (n = 7)
Awaiting further information
(n = 2)

Reports assessed for eligibility
(n = 1)

Included

Studies included in review
(n = 12)
Reports of included studies
(n = 13)

3.4.2. Description of studies

Of the included studies (Table 3.2), three were development and external validation studies (Klein et al., 2018; van Loo et al., 2015; Wang et al., 2014), eight were development only studies (Backs-Dermott, Dobson and Jones, 2010; Berlanga et al., 1999; Johansson, Lundh and Bjärehed, 2015; Judd, Schettler and Rush, 2016; Mocking et al., 2021; Ruhe et al., 2019; van Loo et al., 2018) and one (van Loo et al., 2020) was an external validation study. Only three (Mocking et al., 2021; Ruhe et al., 2019; van Loo et al., 2018) of the development only studies reported internal validation. No prognostic model was externally validated in more than one included study and, therefore, a meta-analysis was not necessary.

3.4.2.1. Source of data and setting

The ideal sources of data for a prognostic model development or validation study are prospective cohort (including RCTs), nested case-control or case-cohort studies. All included studies used prospectively-gathered data for developing the prognostic models. Four of the models were developed in secondary care (Berlanga et al., 1999; Johansson, Lundh and Bjärehed, 2015; Judd, Schettler and Rush, 2016; Pintor et al., 2009), while the other seven were developed in primary care (Klein et al., 2018; Ruhe et al., 2019) or community settings (Backs-Dermott, Dobson and Jones, 2010; van Loo et al., 2015, 2018; Wang et al., 2014; Mocking et al., 2021). Two different development studies (van Loo et al., 2015, 2018) used data drawn from the same source: the Virginia Adult Twin Study of Psychiatric and Substance Use Disorder (VATSPSUD), a population-based longitudinal study of male–male and male–female white twin pairs. Van Loo et al. (2015) used data from female-female twin pairs and Van Loo et al. (2018) used data from male–male and male–female twin pairs from VATSPSUD. Van Loo et al. (2020) used a data-set drawn from primary care, secondary care and community settings (the Netherlands Study of Depression and Anxiety (NESDA)) for

external validation. Table 3.2 summarises the specific outcome definitions used.

3.4.2.2. Participants

All studies identified were developed in a population matching the review inclusion criteria: adults with a diagnosis of depression that met criteria for remission at the point of prediction. Two studies included only women (Backs-Dermott, Dobson and Jones, 2010; van Loo et al., 2015). The authors of van Loo et al. (2015) explained that studying men and women separately might lead to more accurate prediction models because risk factors for relapse can be sex-dependent.

3.4.2.3. Outcome (end-point)

All of the studies included in this review developed prognostic models to predict either relapse or recurrence in participants with remitted depression at the start-point. None were identified predicting sustained remission or recovery. The included studies varied in their outcome definition. Most referenced Frank's relapse criteria (Frank et al., 1991; Rush et al., 2006) or used similar criteria using a mixture of diagnostic instruments and clinical interview. All primary studies identified gave a clear definition of relapse or recurrence and used this consistently across all participants in their studies.

'Recurrence' was defined in a number of ways, ranging from a re-emergence of depressive symptoms at any point but not before two months (Johansson, Lundh and Bjärehed, 2015) to within a median follow-up time of 6.1 years (van Loo et al., 2015). 'Relapse' was defined as a re-emergence of depressive symptoms occurring either within two months of achieving remission (Johansson, Lundh and Bjärehed, 2015), within six months but after at least eight weeks of remission (Judd, Schettler and Rush, 2016) or

within 12 months (Backs-Dermott, Dobson and Jones, 2010). See Table 3.2 for further information on specific definitions used.

3.4.2.4. Predictors

The included studies covered a wide range of predictors (Table 3.3 outlines the different predictors included in the final models and how they were measured for the individual studies). Most commonly, these were disease-related characteristics and demographic factors. Some studies explored some less common predictors such as: neuropsychological predictors (emotional categorisation, emotional memory, and facial expression recognition) (Ruhe et al., 2019); personality characteristics such as neuroticism (Berlanga et al., 1999); psychosocial predictors such as life stress and interpersonal difficulties (Backs-Dermott, Dobson and Jones, 2010); biochemical predictors such as results from the corticotrophin-releasing factor test (Pintor et al., 2009); peripheral blood metabolomic markers (Mocking et al., 2021); and combinations of items from the Symptom Checklist (SCL-90) (Judd, Schettler and Rush, 2016).

3.4.2.5. Statistical analysis methods for model development

Of the 11 development studies, nine used regression analysis [five used logistic regression (Berlanga et al., 1999; Johansson, Lundh and Bjärehed, 2015; Judd, Schettler and Rush, 2016; Pintor et al., 2009; Wang et al., 2014) and four used Cox proportional hazards regression to study time to recurrence (Klein et al., 2018; Mocking et al., 2021; van Loo et al., 2015, 2018)]. Of the remaining two included studies, one used an ML support vector machine model to predict recurrence over a median period of 233 days (Ruhe et al., 2019) and the other used discriminant function analysis (DFA), a statistical method to identify which continuous variables (predictors) best discriminate between two or more groups (in this case, relapse or stable remission) (Backs-Dermott, Dobson and Jones, 2010).

Table 3.2: Characteristics of studies included in systematic review

Study	Type of study	Setting (country)	Source of data (Year of recruitment)	Participants		End-point (follow-up)	Number of participants (Number with event)		Number of candidate predictor parameters (Number of predictors in final model)	Method of model development
				Mean age (SD)	Gender (% Female)		Development	External validation		
Backs-Dermott 2010	Model development	Community (Canada)	Prospective longitudinal cohort study (Not reported)	Relapse group: 43.1 (10.87); Stable remitted group: 43.65 (11.72)	100	Relapse: meeting current criteria for MDE according to SCID-I (12 months)	49 (29)	NA	11 (5)	Differential Function Analysis
Berlanga 1999	Model development	Secondary care (Mexico)	Post-RCT prospective follow-up study (1994-1996)	Recurrence group: 34.8 (11.1); No-recurrence group: 37.2 (11.2)	Recurrence group: 83; No-recurrence group: 71	Recurrence: Fulfilling criteria for MDD on clinical interview (12 months)	42 (18)	NA	Not reported (3)	Logistic regression (multivariable analysis with a stepwise backward method in which variables that were significant in the univariable analysis were introduced into the model)
Johansson 2015	Model development	Secondary care (Sweden)	Prospective cohort study	47 (17)	71	Relapse: depressive episode within	51 (31)	NA	4 (2)	Logistic regression (the 2 predictor

			(Not reported)			2 months of discharge; Recurrence: depressive episode at least 2 months after discharge (12-14 months)				variables were chosen which showed the strongest independent correlations with relapse/recurrence)
Judd 2016	Model development	Secondary care (US)	Prospective cohort study: the National Institute of Mental Health Collaborative Depression Study (1978-1981)	37.8 (14.4)	58.5	Relapse: 2 consecutive weeks of psychiatric status ratings at threshold for defining episode of major or minor/dysthymic depression (6 months)	188 (58); 514 SCL-90 assessments (73 with relapse)	NA	17 (12)	Forward and backward selection of pre-selected predictors using stepwise mixed-model logistic regression
Klein 2018	Model development with external validation	Primary care (The Netherlands)	Prospective data from 2 pragmatic RCTs (Development data: 2010 - 2013; Validation data: 2009 - 2015)	Development dataset: 46.8 (10.6); Validation dataset: 48.3 (9.9)	Development dataset: 74.5; Validation dataset: 66.5	Recurrence (time to): assessed using SCID-I (2 years)	235 (104)	205 (116)	8 (4)	Cox proportional hazards regression (backward selection at P < 0.05)

Mocking 2021	Model development	Community setting (US)	Cross-sectional study comparing people with remitted recurrent MDD (rrMDD) with never depressed controls (2011-2014)	Males: 54 (SEM: 1.4); Females: 53 (SEM: 1.2)	66.1	Recurrence: ≥5 depressive symptoms lasting at least 2 weeks according to the DSM-IV criteria (2.5 years)	62 (35)	NA	399 (Unclear)	Cox proportional hazards regression
Pintor 2009	Model development	Secondary care (Spain)	Prospective cohort study (2001-2005)	Relapsed group: 50.67 (8.04); Non-relapsed group: 51.88 (8.54)	Relapsed group: 50; Non-relapsed group: 56	Relapse: identified using Hamilton Depression Rating Scale (HDRS-21); "Frank et al. (1991) criteria were applied" (does not describe exactly how) (2 years)	43 (18)	NA	Not reported (3)	Logistic regression
Ruhe 2019	Model development	Primary care (The Netherlands)	Prospective cohort study (Not reported)	53.4 (7.7)	65.8	Recurrence: MDD according to SCID-I (Median follow up: 233 days (IQR 92 - 461))	64 (35)	NA	Not reported (4)	Machine learning support vector machine (SVM); data-driven model (classification-based algorithm)

Van Loo 2015	Model development with external validation	Community setting (US)	Prospective longitudinal data (1988-1997)	Development dataset: 30.7 (7.1); Validation dataset: 32.4 (7.1)	100	Recurrence: first episode meeting DSM-III-R criteria after a period of not meeting the criteria (remission or recovery) for at least 4 months (Development dataset: median follow-up 5.5 years; Validation dataset: median follow-up 6.1 years)	194 (45)	133 (57)	81 candidate predictors; number of parameters unclear (26)	Elastic net penalised Cox proportional hazards regression
Van Loo 2018	Model development	Community setting (US)	Longitudinal cohort study (1988 – 1997)	35 (8.8)	34.6	Time to recurrence: First reported episode meeting DSM-III-R criteria in the year prior to follow-up interview (5 years)	653 (Not reported)	NA	70 Predictors, number of parameters unclear (24)	Cox proportional hazards model with elastic net penalised regression analysis

Van Loo 2020	External validation	Primary care, secondary care and community setting (The Netherlands)	Longitudinal cohort study (2004 – 2007)	42 (12.4)	68.6	Recurrence: Any episode of MD during follow-up (9 years)	NA	1925 (Not reported)	NA (24)	Logistic regression used for external validation
Wang 2014	Model development with external validation	Community setting (US)	Prospective longitudinal dataset (2001 – 2005)	Development dataset: 45.38 (0.37); Validation dataset: 45.37 (0.41)	Development dataset: 77.4; Validation dataset: 74.9	Recurrence: Meeting DSM-IV diagnostic criteria for MDE (3 years)	1518 (362)	1195 (307)	Not reported (24)	Logistic regression with combined forward and backward selection

3.4.3. Predictive performance of prognostic models

The predictive performance of all included models is summarised in Table 3. Six of the model development studies identified (Klein et al., 2018; Wang et al., 2014; Ruhe et al., 2019; van Loo et al., 2015, 2018; Mocking et al., 2021) reported internal validation to account for overfitting and optimism within the developed model. Three also reported external validation, using a dataset separate from the training dataset to give a truer reflection of model performance and generalisability (Klein et al., 2018; van Loo et al., 2015; Wang et al., 2014). Van Loo et al. (2020) presented the external validation of the model developed in Van Loo et al, (2018).

Klein et al. (2018) used an RCT dataset separate from that used for development for external validation and presented a calibration slope of 0.56 (0.81 on internal validation) and a Harrell's C-statistic of 0.59 (0.56 on internal validation). Van Loo et al. (2015) used a temporal cut-off to define their development and validation samples (temporal validation). They presented "comparable" Kaplan-Meier curves as evidence that their prognostic model was well calibrated for people at lower risk of relapse but less so for higher-risk participants, and an AUC of 0.61 on external validation (0.79 on internal validation). Wang et al. (2014) used data from the same source but from a different geographical region (geographical validation) to define development and external validation datasets. The authors presented a C-statistic of 0.72, indicating good discrimination, and presented the result of the Hosmer-Lemeshow goodness-of-fit test (3.51, P = 0.9) as evidence of "excellent calibration".

Van Loo et al. (2020) presented the results of external validation in two "test" sets. One of these, the Virginia Adult Twin Study of Psychiatric and Substance Use Disorder (VATSPSUD), was data from the same sample used in Van Loo et al. (2018) for model development and I have therefore classified this as an internal validation. The second test sample (NESDA) is separate from the development dataset and I have focused on this as the

external validation. Discrimination was reported as good [AUC = 0.68 (95% confidence interval (CI) 0.66 to 0.71) predicting recurrence over 0 to 2 years; AUC = 0.72 (95% CI 0.69 to 0.75) predicting recurrence over 0 to 9 years]; calibration was not reported. Of the external validations included in this review, only Van Loo et al. (2020) included 95% confidence intervals for measures of predictive performance.

Klein et al. (2018) was the only included study to present all of the regression coefficients for the predictors included in the final model as well as the intercept and associated 95% confidence intervals. This model could therefore be used based on the information provided in the primary source. None of the included studies explored net benefit analysis (clinical utility) of the developed models.

Table 3.3: Summary of final predictors and predictive performance of prognostic models

| Study | Predictors included in final model | Predictive performance | | | | Other performance statistics presented |
| | | Internal validation | | External validation | | |
		Calibration	Discrimination	Calibration	Discrimination	
Backs-Dermott 2010	'Psychosocial' predictors: Life stress; Cognitive-Personality Vulnerability Factors; Social support; and Coping style: • Interpersonal marked difficulties (Short Life Events and Difficulties Scale, SLEDS); • Perceived social support from a significant other (Multidimensional Scale of Perceived Social Support, MSPSS) • Perceived social support from friends (MSPSS) • Emotion-oriented coping (Coping Inventory for Stressful Situations, CISS); • Avoidance-oriented coping (CISS)	Not reported	Not reported	Not applicable	Not applicable	The DFA was significant: Wilk's Lambda = 0.69, x^2 (5) = 16.35, P = 0.006 Standardised discriminant function coefficients: • MSPSS (Significant Other): 0.48; • MSPSS (Friends): 0.35; • CISS (Emotion-Oriented Coping): 0.67; • CISS (Avoidance-Oriented Coping): −0.58; • Presence of interpersonal severe difficulties: −0.63

Berlanga 1999	'Personality and clinical predictors': • Elevated EPQ (Eysenck Personality Questionnaire) score on the neuroticism subscale; • Short duration of treatment of the index episode; • A slow onset of response to treatment of the index episode	Not reported	Not reported	Not applicable	Not applicable	Combination of 3 variables predicted recurrence of depression in 90% of cases. Threshold not specified Sensitivity: 89% Specificity: 92% Positive Predictive Value: 89% Negative Predictive Value: 92%
Johansson 2015	• Number of previous episodes (0/1/2/3 or more) • Having a partner (yes/no)	Not reported	Not reported	Not applicable	Not applicable	Sensitivity: 90% Specificity: 60% Overall accuracy: 78% (Threshold not defined) Measure of overall model fit: Nagelkerke's R^2 = 0.45 R^2 = 2.97 (Hosmer and Lemeshow), 0.33 (Cox and Snell) Model X^2 = 20.66 (df = 2, P < 0.001) (compared with constant-only model) Final model presented with regression coefficients and intercept: Intercept = −0.68 Previous episodes Beta coefficient = 1.19 (1.55 to 7.06) P = 0.00

						Partner Beta coefficient = −2.14 (0.02 to 0.64) P = 0.01
Judd 2016	12 SCL-90 items in final model: Feeling blocked in getting things done Feeling pushed to get things done Feeling tense or keyed up Having ideas/beliefs others do not share Feeling inferior to others Feeling low in energy or slowed down Feeling very self-conscious with others Headaches Crying easily Feelings being easily hurt Worrying too much about things Trouble concentrating	Not reported	Not reported	Not applicable	Not applicable	Predictive statistics for "experiencing any one or more of the 12 symptoms most predictive of relapse at a moderate or worse level of severity for the past week": Sensitivity: 80.8% Specificity: 51.2% Positive Predictive Value: 21.5%; Negative Predictive Value: 94.2%
Klein 2018	Number of previous MDEs (life-chart of SCID-I), categorised as less than 3, 3 or 4, and 5 or more;	Calibration slope = 0.81	Harrell's C-statistic = 0.56	Calibration slope = 0.56	Harrell's C-statistic = 0.59	Total risk score calculated from final model "scores": low (< 35), moderate (35 - 50), high (> 50)

	Number of residual depressive symptoms (Inventory of Depressive Symptomatology, continuous) Severity of the last MDE (SCID-I), mild or moderate vs severe Treatment in RCT also included as a non-significant predictor					Cut-off score 35 or more (37% risk of recurrence): Sensitivity: 52% Specificity: 69% PPV: 59% NPV: 63% Cut-off score 50 or more (71% risk of recurrence): Sensitivity: 16% Specificity: 95% PPV: 72% NPV: 57%
Mocking 2021	Predictors were all metabolites (peripheral blood metabolomics) known to be core features of the cell danger and integrated stress response (CDR and ISR) pathways. 80% of the metabolic predictors of recurrence in both males and females belonged to 6 pathways: (1) phospholipids, (2) sphingomyelins, (3) glycosphingolipids, (4) eicosanoids, (5) microbiome, and (6) purines.	Not reported	Females: C-statistic=0.90 (95% CI 0.69 to 1.0) Males: C-statistic=0.99 (95% CI 0.90 to 1.0)	Not applicable	Not applicable	
Pintor 2009	Corticotrophin-releasing factor test (net area under cortisol curve (NAUCC), cut-off point of 251.24 µg/ml/min) Previous suicide attempt	Not reported	Not reported	Not applicable	Not applicable	Nagelkerke's R^2 = 0.797 Sensitivity: 89% Specificity: 92% Hosmer-Lemeshow Goodness-of-fit test: χ^2 = 2.23, df = 8, P = 0.97

	Stress during follow-up					
Ruhe 2019	Best classifier included 4 predictors: Number of previous episodes in last 10 years Age of onset CTQ-physical abuse subscale-score CTQ-physical abuse of 8 or more	Not reported	Not reported	Not applicable	Not applicable	Results for "best classifier": Sensitivity: 71.4 Specificity: 79.3
Van Loo 2015	Recent depressive episode: Loss of interest (HR 1.10) Appetite loss (HR 1.02) Weight loss (HR 1.05) Weight gain (HR 0.99) Insomnia (HR 1.07) Concentration difficulties (HR 1.07) Feeling anxious, nervous, worried (HR 1.03) Feeling tense, jumpy, shaky (HR=1.06); Sum of 9MD criteria (HR 1.02) Current state: SCL past 30 days (HR 1.03)	Not reported.	AUC = 0.79	Not reported.	AUC = 0.61	Comparable KM-curves for the 2 lowest risk groups was used as evidence that the model is well-calibrated for those at lower risk but less so for higher-risk patients

	Psychiatric history (lifetime):					
	Age at first depression (HR 1.06) Number of MD episodes ≥ 6 (HR 1.05) Duration of most severe MD episode 1 - 3 months (HR 0.98) Duration of most severe MD episode ≥ 3 months (HR 1.03) Early anxiety (HR 1.06) Family history: GAD co-twin (HR 1.06) Personality: Extraversion (HR 1.02) Adverse life events (early): Parental loss childhood/adolescence (HR 1.03) Disturbed family environment (HR 1.02) Sum of lifetime traumas 3 - 4 (HR 1.06)					

	Childhood sexual abuse (severe) (HR 1.04) Adverse life events (recent): Number of stressful life events in past year (HR 1.03) Social and economic environment: Marital status (HR 1.03) Low marital satisfaction (HR 1.04) Problems with relatives (HR 1.02) Financial problems (HR 1.15)					
Van Loo 2018	Recent depressive episode: Loss of interest (HR 1.11) Appetite gain (HR 1.01) Weight loss (HR 1.02) Feeling restless (HR 1.02) Fatigue (HR 1.04) Hypersomnia (HR 1.04) Feeling irritable/angry (HR 1.06) Feeling tense (HR 1.04) Cardio-respiratory panic symptoms (HR 1.11) Sum of 9 MD criteria (HR 1.05)	Not reported	AUC in the male training sample = 0.785 AUC in male test sample = 0.710	Not applicable	Not applicable	KM-curves for the low-risk group in both training and test data were very similar, indicating good discrimination and calibration for participants with lower risk for depression. The KM-curves for the intermediate and high-risk groups were more similar in the test data than in the training data, which indicated that the model was less well- calibrated for higher risk patients

	Current state:
	
	  SCL last 30 days (HR 1.06)
	
	Psychiatric history (lifetime):
	
	  Early anxiety (HR 1.15)
	  History of GAD (HR 1.76)
	  2 – 3 MD episodes lifetime (HR 1.02)
	  $\geqslant$ 6 MD episodes lifetime (HR 1.14)
	  History of alcohol dependence (HR 1.03)
	
	Family history:
	
	  MD mother (HR 1.09)
	
	Early adverse life events:
	
	  Childhood sexual abuse (HR 1.19)
	  Traumas $\geqslant$ 5 (HR 1.13)
	
	Recent adverse life events:
	
	  Number of stressful life events in past year (HR 1.01)
	
	Social and economic environment:					

	No partner (HR 1.03) Low marital satisfaction (HR 1.13) Support from relatives (HR 0.99) Problems with relatives (HR 1.03)					
Van Loo 2020	As for Van Loo 2018	Not reported	Predicting MD over 0 - 1 year: AUC = 0.73 (95% CI 0.69 to 0.76)*	Not reported	Predicting MD over 0 - 2 years: AUC = 0.68 (95% CI: 0.66 to 0.71) Predicting MD over 0 -9 years: AUC = 0.72 (95% C: 0.69 to 0.75)	None.
Wang 2014	Female sex Age (continuous); Married/common-law Divorced/separated/single White Had MDD last year 2 depressive episodes 3+ depressive episodes Lifetime GAD or specific phobia Avoidant personality disorder Depressive symptoms in MDE:	Not reported	C statistic = 0.75	Not reported	C statistic = 0.7195	Model development: Hosmer-Lemeshow χ^2 (8) = 10.48, P = 0.23 "Excellent calibration" External validation: Hosmer-Lemeshow χ^2 (8) = 3.51, P = 0.90 "Excellent calibration" In the combined development and validation data: C statistic of 0.7365 and "excellent calibration" (H–L χ^2 (8) = 6.22, P = 0.62)

	Difficulties in concentration Wanted to eat more Felt guilty Took medication for low mood SF-12 physical disability scores (53.9 to 57.8; 43.3 to 53.8; 0 to 43.2) SF-12 mental disability scores (48.4 to 54.5; 37.7 to 48.3; 0 to 37.6) Experience of racial discrimination Ever physically attacked/beaten/injured); by spouse, partner, or anyone else (abuse) (Experience of sexual assault) Before 18, parents/caregiver swear, insult, or say hurtful things to you (Almost never/sometimes; fairly often/very often) Before 10 being left alone/unsupervised by parents/care givers (Almost never/sometimes; fairly often/very often) Interaction terms: Sex × SF-physical Marital × Abuse					Observed risk of recurrence over 3 years = 25.40% (95% CI 23.76% to 27.04%) Mean predicted risk of recurrence based on the model = 25.34% (95% CI 24.73% to 25.95%). "We visually compared the predicted versus the observed risk of recurrence by decile risk groups"

	Race × Avoid SF-physical × Guilty					

*This internal validation used the same data as development data [Van Loo et al. (2018)]

3.4.4. Risk of bias and applicability assessment of included studies

I rated 11 of the 12 included studies as being at high overall ROB (see Figure 3.2 for summary and Appendix 3.4 for full details). Only one study, Klein et al (2018), was assessed to be at low risk of bias in all four domains. ROB was generally assessed as being low for most studies in the domains of participants and predictors. Predictors were generally measured appropriately and in the absence of knowledge about outcomes. An exception was Van Loo et al. (2020), where predictor information was not available until after the point of prediction for some predictors. There were some infrequent examples of lack of clarity around the measurement of some of the predictors and outcomes; for example, Pintor et al. (2009) described the assessment of relapse according to Frank et al. (1991)'s criteria applied to the Hamilton Depression Rating Scale-21 but did not report cut-offs or the evidence for them. ROB was unclear for nine out of 12 of the studies in the domain of outcomes, because the studies did not state that outcomes were determined blind to the predictor information. For the fourth domain (analysis), there was variable quality of the reported methods and some weaknesses and potential sources of bias were identified in this domain for 11 of the 12 included studies.

The most common limitation was related to sample size or number of events, or both, a lack of which adversely impairs the predictive ability of a statistical model in the real world due to a significant risk of overfitting. Most studies did not describe how the sample size was determined. Only one study (Klein et al., 2018) reported sufficient EPP for model development (104 recurrences for eight candidate predictor parameters). While this study did not meet the cut-off of 20 EPP, I rated it as 'probably yes' for Item 4.1 (reasonable number of participants with the outcome) because the authors had used internal validation techniques to account for optimism in the model.

All other regression models (Berlanga et al., 1999; Johansson, Lundh and Bjärehed, 2015; Judd, Schettler and Rush, 2016; Pintor et al., 2009;

Mocking et al., 2021; van Loo et al., 2015, 2018; Wang et al., 2014) had inadequate sample size, according to PROBAST (see Section 3.3.6). The sample size determination used by Backs-Dermott et al. (2010), which used DFA, appeared to be appropriate according to their reported methods. Ruhe 2019 used a ML approach for model development (Ruhe et al., 2019). Formal guidance is lacking to aid sample size determinations for prognostic model studies using ML techniques. The guidance and literature that does exist suggests that we should demand, if anything, significantly larger sample sizes when using a ML approach to prognostic model development, with one paper estimating that one would need more than 10 times the EPP required for regression models to achieve a stable area under the curve (AUC) and small optimism (van Der Ploeg, Austin and Steyerberg, 2014). This study did not have an adequate sample size according to any of the existing guidance and recommendations. For Van Loo et al. (2020), while it was not explicitly stated, I made the assessment that the sample size probably met PROBAST requirements for external validation (at least 100 events).

Another common limitation of the included studies (n = 8) was their handling of missing data. Multiple imputation was used to handle missing data in only four of the identified studies (Klein et al., 2018; Judd, Schettler and Rush, 2016; van Loo et al., 2018, 2020). The remaining studies either did not report their approach (Backs-Dermott, Dobson and Jones, 2010; Berlanga et al., 1999; Johansson, Lundh and Bjärehed, 2015; Pintor et al., 2009) or used non-PROBAST recommended approaches for handling missing data, such as imputing the mean (Ruhe et al., 2019) or single imputation (van Loo et al., 2015; Wang et al., 2014). Finally, most studies (n = 11) did not present appropriate performance statistics. The PROBAST guidance recommends that, as a minimum, a calibration plot and discrimination statistics (C-statistic for binary and time-to-event outcome models) are presented as relevant performance measures for a prognostic model study (Wolff et al., 2019). Classification measures, such as sensitivity and specificity, can be presented in addition to calibration and discrimination statistics, but they have the drawback of loss of information and of requiring risk thresholds to be specified, often based on the data rather than on

meaningful, clinical grounds. One study (Klein et al., 2018) presented both a calibration plot and C-statistic in line with minimum best practice.

I had low concern about applicability for all included studies except for Berlanga et al. (1999), which was rated at an unclear level of concern (Figure 3.3). It was unclear whether all of the participants had reached remission and it appears that a proportion of participants would have met criteria for depression according to the Hamilton Depression Rating Scale. The inclusion criteria were purposefully broad, as I was interested in exploring a range of models and settings, which might explain the overall low concern about applicability. A point of note is that five of the studies here had significant time periods between data collection and publication of the data analysis. This period was nine years in the case of Wang et al. (2014) and for four of the studies, this gap was more than a decade (13 years for Van Loo et al. (2020); 18 years for Van Loo et al. (2015); 21 years for Van Loo et al. (2018); and 35 years for Judd et al. (2016). While not explicitly addressed in the 'Risk of bias' assessment, it is worth noting that this could have implications for reliability and applicability of results (for example, the time from data collection to analysis means that data queries can be less easily addressed and that findings are potentially less applicable to the health service in its current form).

Risk of bias

Study	D1	D2	D3	D4	Overall
Backs-Dermott 2010	+	+	–	✗	✗
Berlanga 1999	+	+	–	✗	✗
Johansson 2015	+	+	–	✗	✗
Judd 2016	+	+	+	✗	✗
Klein 2018	+	+	+	+	+
Mocking 2021	+	+	+	✗	✗
Pintor 2009	+	–	–	✗	✗
Ruhe 2019	+	+	–	✗	✗
Van Loo 2015	+	+	–	✗	✗
Van Loo 2018	+	–	–	✗	✗
Van Loo 2020	+	✗	–	✗	✗
Wang 2014	+	+	–	✗	✗

D1: Participants
D2: Predictors
D3: Outcome
D4: Analysis

Judgement
✗ High
– Unclear
+ Low
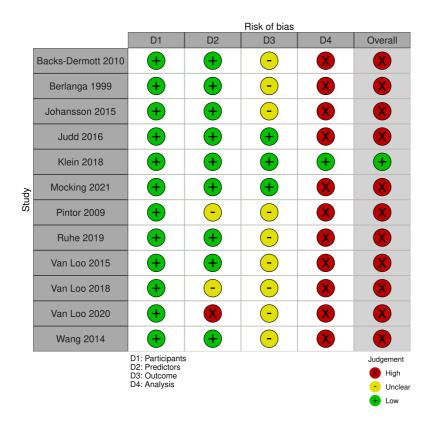
Figure 3.2: Risk of bias assessment (PROBAST) of included studies

Concerns about applicability

Study	D1	D2	D3	Overall
Backs-Dermott 2010	+	+	+	+
Berlanga 1999	–	+	+	–
Johansson 2015	+	+	+	+
Judd 2016	+	+	+	+
Klein 2018	+	+	+	+
Mocking 2021	+	+	+	+
Pintor 2009	+	+	+	+
Ruhe 2019	+	+	+	+
Van Loo 2015	+	+	+	+
Van Loo 2018	+	+	+	+
Van Loo 2020	+	+	+	+
Wang 2014	+	+	+	+

D1: Participants
D2: Predictors
D3: Outcome

Judgement
– Unclear
+ Low
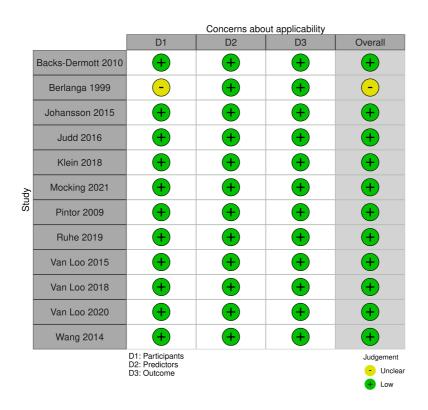
Figure 3.3: Applicability assessment (PROBAST) of included studies

88

3.5. Discussion and implications of review for this thesis

This is the first systematic review of prognostic models predicting relapse and recurrence of depression. I identified 11 unique models, across 12 included studies. The planned meta-analysis was not indicated, due to an insufficient number of studies reporting performance statistics for the same model. I therefore presented a narrative synthesis and critical appraisal of the existing literature reporting efforts to develop relapse prediction models for people with remitted depression.

None of the models identified underwent independent external validation (i.e., by researchers not involved in the original model development) or net benefit analysis to assess clinical utility. Only one of the included models was found to be at overall low risk of bias (Klein et al., 2018). This prognostic model, developed using Cox proportional hazards regression, predicted time to recurrence within two years and included the following predictors: number of previous episodes of depression (less than 3; 3 or 4; 5 or more), number of residual symptoms, severity of last depressive episode according to SCID-I (mild or moderate; severe) and treatment (this was to control for the treatment received in the RCT and was a non-significant predictor). The discrimination and calibration of this model were both poor on external validation. The other eleven studies had weaknesses in their analysis, particularly for sample size, handling of missing data and not presenting appropriate performance statistics. The discussion of the systematic review presented here is intended to inform the quantitative workstream.

3.5.1. Strengths and limitations of the review

This was a wide-ranging review in an innovative and developing area for Cochrane as a whole, and for the Cochrane Common Mental Disorders group. I was guided by the recent prognosis literature and guidance in developing the review methods, searches and in critically appraising the

included studies. I identified models incorporating a range of predictors and using a variety of statistical methods. One limitation is the absence of a quantitative synthesis due to a lack of eligible studies presenting the predictive performance of the same model. I undertook the 'Risk of bias' assessment using the PROBAST tool. It is important to note that PROBAST was primarily designed for the assessment of primary prognostic model studies using regression-based techniques. One study identified in this review used ML techniques (Ruhe et al., 2019). The PROBAST guidance is less directly applicable to ML techniques, although the guidance does recommend tailoring the tool for different methodological approaches, and this can include ML (Wolff et al., 2019). Longer-term, formal guidance developed by experts is expected to ensure a more robust and consistent assessment of risk of bias for prognostic model studies using ML techniques.

3.5.2. Comparison with the previous literature

The findings from this review are broadly in agreement with previous findings from prognosis studies (mainly prognostic factor studies or reviews of prognostic factor studies) for relapse and recurrence of depression (discussed in Section 2.3). The number of previous episodes was the most commonly-included predictor across the models identified in this review (n = 6) (Johansson, Lundh and Bjärehed, 2015; Klein et al., 2018; Ruhe et al., 2019; van Loo et al., 2015, 2018; Wang et al., 2014). The presence of residual symptoms was used as a predictor only in one developed model (Klein et al., 2018). Adverse childhood experiences were included as predictors in four of the included studies (Ruhe et al., 2019; van Loo et al., 2015, 2018; Wang et al., 2014), comorbid anxiety in three (Wang et al., 2014; van Loo et al., 2015, 2018), neuroticism in one (Berlanga et al., 1999), and age of onset in two models (Ruhe et al., 2019; van Loo et al., 2018). Notably, rumination was not explored as a predictor in any of the included prognostic models, despite good evidence that this is associated with increased risk of relapse (Buckman et al., 2018; Hardeveld et al., 2010).

Wang et al. (2014) found that marital status "contributed to" the prediction of recurrence, while another study (Johansson, Lundh and Bjärehed, 2015) included having a partner or not as one of the two predictors in their final model [odds ratio of 0.12 (95% CI 0.02 to 0.64), p=0.01]. The extant literature does not support marital status as a predictor of recurrence (Burcusa and Iacono, 2007; Buckman et al., 2018) and weaknesses in the methodology of the prognostic model studies mean that we cannot make conclusive statements about this. However, given the strength of the association presented (Johansson, Lundh and Bjärehed, 2015), the prognostic significance of "having a partner or not" warrants further investigation. The model development study by Van Loo et al. (2018) supports the findings of earlier research suggesting that gender is unlikely to be predictive of relapse. It is worth stating here that where there is a lack of evidence for an association between a variable and an outcome, this variable should not necessarily be excluded from a prognostic model study. In summary, with respect to relapse, this review is broadly in agreement with, and has not found strong evidence to challenge, the findings from the pre-existing literature.

This review focussed on prognostic models which predicted relapse, recurrence, sustained remission or recovery in people with remitted depression. In addition to the studies included in this review, it may be helpful to consider other attempts to develop prognostic models for depression-related outcomes. There have been some previous attempts to derive and validate multivariable prognostic models to predict depression-related outcomes other than relapse and recurrence. Existing prognostic models for depression outcomes include a model (the Depression Outcomes Calculator-Six Items, (DOC-6©)) to predict remission (C-statistic (AUC) of 0.62 (95% CI 0.57 to 0.66)) or persistent depressive symptoms (C-statistic (AUC) of 0.67 (95% CI 0.61 to 0.72)) at six months' post-diagnosis (Angstman et al., 2017); a model to predict persistent symptoms at six months (C-statistic not reported; R^2 of 0.40 in development sample and 0.27 in validation sample) (Rubenstein et al., 2007); and a model to predict onset of depression in non-depressed general practice attendees (C-statistic of 0.79 (95% CI 0.77 to

0.81)) (King et al., 2010). The studies in this review present predictive performance statistics broadly in line with these, suggesting that successful individualised prediction might be possible for depression outcomes, but better-quality studies and potentially different combinations of predictors are needed to explore this further.

3.5.3. Implications of review for thesis

This review has implications for depression research, for prognosis research more generally, and for this thesis. First, the findings of this review confirm that the terms 'relapse' and 'recurrence' were inconsistently used in the primary prognostic model studies. Given the range of definitions used, this review is unlikely to inform future deployment of the terms; future research should look to empirically test and update these definitions. The important point for this thesis is that a prognostic model will aim to predict the outcome according to the definition used at the time of development. It is therefore important that the outcome definition used is valid and that clinicians are informed about the underlying theoretical basis for models used in practice and the context in which they were developed.

Secondly, the range of models presented in this review suggests that this is a subject that researchers recognise as important. However, while many of the studies identified have reported promising predictive performance, the high risk of bias in the analysis and lack of external or independent validation mean that the results must be interpreted with caution. Similarly, the clinical utility (net benefit) of using the models, which quantifies the overall utility of using a model to inform clinical decisions at thresholds of predicted risk (Vickers, Van Calster and Steyerberg, 2016), has not been examined. In summary, none of the prognostic models identified in this review had sufficiently-high performance metrics to enable clinicians in primary care or other settings to accurately predict an individual's risk of relapse at present.

Finally, I reported some key methodological weaknesses in the studies identified in this review, particularly with respect to sample size. Unless the sample size is adequate, there will be limitations to how far we can trust the predictive performance statistics presented by the model development study as overfitting is likely. I discussed the limitations of the sample sizes within the primary studies in this review, and considered events per predictor parameter (EPP) as a 'rule of thumb' for determining minimum sample size. EPP has recently been criticised as being too simplistic and not evidence-based (Van Smeden et al., 2016). More sophisticated guidance has been developed and reported (Riley et al., 2019a, 2018) in which adequate sample size is dependent on the number of outcome events, number of predictors and desired accuracy of the model. EPP is still the method of sample size determination suggested in the most recent iterations of the TRIPOD (Moons et al., 2015) and PROBAST (Wolff et al., 2019) guidance. A recurring concern, noted by the authors of the PROBAST guidance, is that prognostic models are often developed using the data available, and sample size justifications are often unreported or are *post hoc* and descriptive (Moons et al., 2015). The study reported in this thesis aimed to address the methodological limitations identified in this review.

3.6. Summary

This review identified 11 prognostic models developed to predict the risk of relapse or recurrence in people with remitted depression. The models were developed in a variety of clinical settings and patient populations and with a range of included predictors. In summary, it is not possible to reliably predict outcomes for a given person with remitted depression based on their demographic, clinical, and disease-level characteristics. There is, however, good evidence about some of the predictors of relapse and recurrence of depression. There is less strong evidence that these predictors can be incorporated into multivariable prognostic models to provide accurate individualised risk predictions. This review suggests that this might be

possible, although the studies identified here were limited by their high risk of bias due to methodological weaknesses.

In Chapter Four, I will discuss the research methodology and philosophy underpinning the thesis, the mixed methods approach taken and how the findings from this review informed the quantitative study. Chapters Five and Six will then present the quantitative study, a novel prognostic model development and validation study that aimed to build on the findings of this review for a primary care setting. Specifically, this study aimed to ensure that the developed model was applicable to a primary care population, and that the predictors included were relevant to and available in primary care setting. I also aimed to address the methodological limitations identified in the majority of studies in this systematic review, by ensuring an adequate sample size, robust statistical analysis in line with the PROGRESS framework, and an assessment of model generalisability.

Chapter Four

Research Methodology

4.1. Overview of chapter

Research methodology refers to the theory underpinning research design and is distinguished from methods, which are the specific tools and techniques used for data collection and analysis (Kelly, 2009). In this chapter, I will describe the methodology applied first for the quantitative and qualitative studies in turn and then for the mixed methods approach taken throughout the thesis (an overview of the mixed methods approach is provided in Figure 4.1). The methods for the specific studies are detailed later in the thesis (Chapters Five and Seven). This chapter also outlines and explains the role of patient and public involvement and engagement (PPIE) throughout the study.
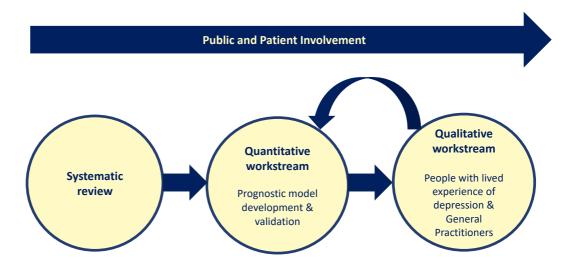
Figure 4.1: An overview of convergent mixed methods approach

4.1.1. Research philosophy

When conducting research, it is essential to consider one's theoretical position and the ontological and epistemic assumptions underlying the methodology (Braun and Clarke, 2013). Ontology refers to the reality that we are attempting to understand, and whether the researcher believes there is a reality that exists independent of the researcher (a realist ontological position) or whether we believe that reality only exists as far as we can experience and understand it subjectively (a relativist ontological position). These two more extreme standpoints may not be helpful when engaging in health services research. There are certainly some sociological phenomena where a relativist and purely experiential approach would be appropriate and, similarly, some scientific disciplines where I think a realist position would be a reasonable position to adopt. Health services research, however, involves a complex intersection between biological, psychological and social processes. In this context, I believe there are some fixed facets of the external reality that are knowable, but that individuals' prior experiences and ideas are bound to affect their understanding and interpretation of reality, and that this must be borne in mind when attempting to make sense of data. A critical realist ontological position is a middle ground of sorts and holds that there is a reality that we are trying to understand but that, as researchers, we can only ever partially know it (Cruickshank, 2012; Bergin, Wells and Owen, 2008). Critical realism is the ontological position I have adopted in this thesis.

Epistemology is related to ontology, but refers to the knowledge we are producing (i.e., how we come to understand reality) rather than how we think of reality itself. The epistemic position adopted usually arises from one's ontological assumptions. Epistemology can be thought of on a spectrum from a positivist standpoint (which generally maps to a realist ontological position) through to a constructivist standpoint (which is associated, although not exclusively, with a relativist ontology). Critical realism is distinct from some of the other ontological positions in that it also encapsulates some epistemic concerns (Bergin, Wells and Owen, 2008). Critical realism is also aligned

with a contextualist epistemology, which, while aiming to understand "the truth", allows one to acknowledge that knowledge arises from the context in which is studied and that truth is situational and dependent on factors relating to the research methodology used to understand it. A contextualist epistemological position does not assume that the data gathered directly represent reality but posits that knowledge is local, situated and provisional and that reality is made sense of by the researcher's interpretative practices. This is the epistemological position I have adopted throughout the thesis.

4.2. Systematic review and quantitative study

4.2.1. Philosophical considerations

The research methodology adopted for each of the studies is aligned with the research objectives, outlined in Chapter One (Section 1.3). As a reminder, objectives one and two were:

Objective 1: To identify and critically appraise existing prognostic models to predict relapse or recurrence of depression.

Objective 2: To develop and validate a prognostic model for patients with remitted depression in primary care, to predict individualised risk of relapse.

These were addressed through a systematic review (reported in Chapter Three) and a quantitative study, respectively. For narrative purposes, this chapter follows the systematic review, because the findings from the review directly informed the methodology used in the quantitative study. However, a research philosophy underlies the systematic review, just as for the subsequent quantitative and qualitative studies.

The systematic review and quantitative study are unavoidably both underpinned by a more realist ontology and positivist epistemology than the qualitative work (described later). For both studies, I made the assumption

that reality can be objectively measured for the purpose of analysis and drawing conclusions (Tariq and Woodman, 2013). I assumed that depression is a real illness that exists objectively in the world and that people with depression can experience remission and subsequently relapse (or not). Epistemically, I assumed that we can know this by identifying the existence of depression, remission and relapse using validated tools. My approach to these studies was aligned with Popper's hypothetico-deductive approach to scientific enquiry (Popper, 1959; Musgrave, 2011). I hypothesised that, following a review of the literature and consensus of my multidisciplinary supervisory team and PAG, the variables I selected for having strong evidence as relapse predictors would combine to provide accurate and precise individualised risk estimates. I used statistical reasoning and mathematical calculation of sample size to ensure I could infer from my sample to the wider population. I then set out to test this, to measure the uncertainty in the analysis (in the form of 95% confidence intervals) and to assess the generalisability of the results. Bias is perhaps easier to control for when working within a quantitative rather than qualitative paradigm (discussed later). I was able to apply methodological and statistical rigor in the form of adhering to expert guidance on statistical methods, reporting measures of uncertainty and quantitatively validating the findings of the primary analysis.

Despite these assumptions, taken for pragmatic reasons for the purpose of quantitative analysis, my overall approach remained consistent with that taken throughout the thesis, which is aligned with critical realism and contextualism. Throughout the thesis, I have retained a commitment to questioning and challenging the validity of the underlying constructs and the integration of the mixed methods enabled me to do this more robustly in the concluding chapter. Critical realism allowed me to approach the research questions under the assumption that there is a real world to be investigated, but that that my underlying assumptions and perspective as a researcher will impact the analysis and findings from the research

4.2.2. Methodological approach to prognostic model development

The quantitative methodology was guided by the systematic review presented in Chapter Three. As discussed in that chapter, there are a number of ways of approaching the development of a prognostic model. The systematic review highlighted a range of statistical approaches to prognostic model development for relapse of depression and, as I have discussed, in most studies the risk of bias was high for predominantly methodological reasons. There was no dominant algorithm or approach that outperformed others. One of the key distinctions between approaches to prognostic modelling is between those that are data-driven and those that are guided by *a priori* literature review. Data-driven approaches range from those that approach the dataset in a way that is theoretically "blind", such as many machine learning (ML) algorithms, to those that use some *a priori* predictor selection and then are driven by statistical analysis and significance testing (for example, elastic net) (Riley et al., 2019b).

The main approaches Identified In the review were 'traditional' regression-based techniques, classification approaches (such as discriminant function analysis (DFA)), and ML approaches. It is important when designing prognosis research to be mindful of the relative benefits and disadvantages associated with different methodological approaches. Regression-based approaches to prognostic model development use regression models to describe an outcome variable (i.e., relapse) in terms of one or more explanatory variables (predictors). This approach allows the modelling of categorical and continuous predictors (and outcomes) and a strength is its ability to produce models that are transparent, explainable and easy to apply in new contexts (Riley et al., 2019b).

DFA is a statistical method used to identify which continuous variables (predictors) best discriminate between two or more groups (in this case, relapse or stable remission (Backs-Dermott, Dobson and Jones, 2010)). DFA is used to answer the same questions as logistic regression but can be used only for continuous (not categorical) predictors (Tabachnick, 1996).

Significance testing (for example, using Wilks' lambda) is used to identify which variables are most discriminatory. A limitation is that the results are not probabilistic but instead present a categorisation that assumes equal utility for all participants without the necessary and important net benefit approach. Regression techniques are generally more appropriate for prognostic model development to present probabilities which can then be used, along with cost-effectiveness information and qualitative data, to assign risk categories (Riley et al., 2019b).

ML approaches generally incorporate regression and classification techniques, among others, and offer the potential of greater predictive performances than more traditional approaches (Tiffin and Paton, 2018). However, this not always the case, as some studies (Tate et al., 2020) have shown. It can also be criticised for lack of interpretability, and variable reporting standards, although the forthcoming TRIPOD-AI may encourage greater consistency in this regard (Collins et al., 2021).

For the purpose of this study, the aim was to develop a prognostic model for use by GPs in a primary care setting. The priorities were that the model be easily interpretable (i.e., GPs and people with depression can understand why it provides the particular risk estimates), have face validity for GPs and people with depression (i.e., include predictors that seem appropriate) and driven by *a priori* prognostic factor research (to ensure important predictors were not inappropriately excluded following data-driven predictor selection). I selected key predictors on the grounds of best available evidence and clinical acceptability, as well as practical reasons related to data availability. Chapters Two and Three both present a comprehensive review of the extant literature with respect to prognostic factors and prognostic models that have guided the methodology of the quantitative workstream. The list of resulting evidence-based predictors was of an appropriate length so I decided to avoid predictor selection techniques during model development and include all predictors regardless of their statistical significance ("full model" approach) (Harrell Jr, 2015). This approach has the advantages of not being overly data-dependent and avoids

the risk of removing clinically important predictors from the final primary model (Harrell Jr, 2015). Prognosis research benefits from also being exploratory in nature (Riley et al., 2019b) and so, while I outlined a pre-registered confirmatory primary analysis, I also detailed a planned secondary (exploratory) analysis including less robustly-evidenced predictors. The aim of doing so was to guide future prognosis research in this area and to highlight any potentially important omissions from my primary analysis.

The other Important methodological consideration with respect to the quantitative component was the fact that the dataset was formed from individual participant data from multiple sources (explained further in Chapter Five). It is essential when using clustered data for prognostic model development that one account for the clustering in the analysis (Bouwmeester et al., 2013). The main feature of clustered data is that participants from the same cluster are likely to be more similar to each other than participants from different clusters. Failing to control for this in the analysis leads to predictions that are poorly calibrated and do not generalise well to other data. I used multilevel regression analysis to account for clustering within the different sources of IPD (details in Chapter Five).

Prognosis research has grown as an area over recent years (Riley et al., 2019b) and, with the development of the PROGRESS initiative, there are now standards and guidelines for conducting (Steyerberg et al., 2013), reporting (Moons et al., 2015) and appraising (Wolff et al., 2019) prognostic model studies. This guidance has been followed throughout the thesis and the quantitative study is reported in line with the TRIPOD (Moons et al., 2015) and TRIPOD-Cluster (Debray et al., 2023) statements.

4.3. Qualitative methodology

Objective three was "to explore the perspectives of people with lived experience of depression and GPs on relapse of depression, relapse risk prediction and relapse prevention interventions". Qualitative methodology

was used to address this objective, to understand and interpret the meaning underlying the experiences and perspectives of people with lived experience of depression and GPs.

Qualitative research can be thought of as multiple interpretive techniques which seek to understand phenomena in the social world. These typically use descriptive and coding tools to come to meaning via methods such as interviews, focus groups and participant observation (Maanen, 1979). Qualitative research is argued to have much to contribute to health service and health policy research by developing depth of knowledge and a more comprehensive theory base in which the '*why?*' questions are dealt with as well as the '*how?*' questions (Sofaer, 1999). Qualitative research requires careful, prospective design to ensure methodological integrity, meaning that the research design should support the research goals and should be consistent with the adopted theoretical paradigms and assumptions as well as the phenomena under investigation (Levitt et al., 2017). Data analysis should be appropriate to the research question and the method of data collection should generate data appropriate to the method of analysis (Willig, 2008).

4.3.1. Theoretical considerations

4.3.1.1. Qualitative research philosophy

As outlined at the start of this chapter, my philosophical approach throughout this thesis (and particularly for the qualitative study) was grounded in a critical realist ontology and contextualist epistemology. For data analysis, I opted to use a thematic analysis method to enable me to incorporate my own biases and experiences into the qualitative interviewing, my understanding of the data and the generation of themes (Braun and Clarke, 2021b).

4.3.1.2. 'Generalisability' and trustworthiness

The question of whether qualitative research findings are generalisable is debated. Qualitative research does not aim to be statistically generalisable in the quantitative sense of inferring numerical effects within a population from analysis within a representative sample [or "statistical-probabilistic generalisability" (Smith, 2018)]. It is, however, unhelpful to assert that the findings of qualitative research are in no way generalisable to the wider population (Smith, 2018). In this study, my aim was to explore the views of the participants with a view to informing clinical pathways in a way that would affect the wider population of GPs and people with depression in the real-world NHS. Therefore, it was important that the results were relevant and meaningful to the wider population of GPs and people with depression.

While it is true that generalisability does not apply in the same way as for quantitative research, qualitative research must be methodologically rigorous if it is to be useful. Trustworthiness is recognised as an important goal for qualitative research and criteria for demonstrating trustworthiness in qualitative research are established (Nowell et al., 2017). I will briefly consider the four trustworthiness criteria developed by Lincoln and Guba (1985) here and how they apply in this study. The four criteria are: credibility, transferability, dependability and confirmability (Lincoln and Guba, 1985).

Credibility describes whether the findings of the research are recognisable and make sense to co-researchers and readers (Nowell et al., 2017). It has been operationalised throughout this study by ensuring regular debrief and discussion of my results with my multidisciplinary supervisory team and the PAG. Transferability is the criterion that captures the quality of generalisability to the wider population and was also described by Smith (2018) as inferential generalisability. This concept of transferability is consistent with the ontological and epistemological positions I have adopted and allows for inferences to be made to guide care for the wider population. The approach I have taken is to rigorously reflect on and challenge whether the findings from this research are applicable and transferable to the wider

context of UK primary health care and general practice (Braun and Clarke, 2021a). I have done this by, first, ensuring an adequate number of interviews (discussed under "sampling considerations" below, and accepting that sample size in qualitative research does not ensure generalisability in the same way as in quantitative research); second, by triangulating my qualitative findings with findings from the quantitative workstream and my own and supervisors' experiences and reflections; and third, by writing a report that explores the findings in a rich and analytic way, accounting for contradictory accounts, comparisons and contrasts between the experiences of GPs and people with lived experience of depression.

Dependability relates to how well-documented and transparent the research procedures are, and confirmability is dependent on the preceding three criteria being adequately met and clear to the reader (Nowell et al., 2017). To ensure my study is dependable and confirmable, I have been explicit about my assumptions underlying the analysis and described in detail my approach to data collection and the analysis itself. I kept a reflexive journal throughout all interviews and I have presented my reflections on the role of reflexivity and the research process and in Chapters Seven and Ten respectively.

4.3.2. Sampling approach and considerations

To meet the objectives of the study, I aimed to interview people with lived experience of depression and GPs. Purposive sampling was used to ensure I had information-rich cases (Green and Thorogood, 2018). I used a maximum variation approach, to cover the widest range of "types" that are likely to be found in the larger population. I considered it less important that these different types were represented in our sample in the same proportion as they are in the population. My aim in choosing this sampling approach was to ensure variability across the group of people accessing primary care for depression and GPs had been captured within my sample. This was to ensure that as much useful information as possible could be gathered in

relation to the research and not in an effort to generalise, in a statistical sense, to the whole population (Sandelowski, 1995). I also incorporated a snowball sampling strategy (Green and Thorogood, 2018; Sedgwick, 2013) when recruiting GPs, utilising contacts of the GPs who agreed to be interviewed and also my own professional network. I avoided recruiting and interviewing GPs with whom I am closely acquainted, as this would have increased the risk of participants sharing incomplete accounts with me (Mcconnell-Henry et al., 2009).

For people with lived experience of depression, I aimed to ensure maximum variability across gender, age, socioeconomic status (measured as index of multiple deprivation of home postcode) and ethnicity. To ensure validity of findings, I considered it important that depression was the main condition for which people had sought healthcare and that their experiences were not primarily the result of an alternative, significant medical or psychiatric diagnosis. Patients identified through screening of their electronic primary care medical record by a GP had a clinically validated code for a diagnosis. For participants who approached the research team directly having seen the study advertisement, I could not be certain of the validity of their clinical diagnosis as I did not have access to their health record. However, while a clinically-validated diagnosis is important, and whilst the severity of depression can be assessed using validated tools such as PHQ-9 (Kroenke, Spitzer and Williams, 2001), ultimately it is the person's story and experience of their symptoms which is of most significance. For the small number of participants not recruited via GP invitation, I explored and verified the self-reported diagnosis as part of the qualitative interview, using the submitted semi-structured topic guide.

For GPs, I prioritised ensuring maximum variability across the variables of: years of experiences as a GP, capacity as a GP[6] (partner,

---

[6] GPs work in three main contractual capacities in the UK: GP partners (self-employed business-owners who hold a contract to deliver NHS services); salaried GPs (GP who are employed by GP practices or other services to provide clinical

salaried, locum), ethnicity and socioeconomic status (measured as Index of Multiple Deprivation at practice level). Following discussion with my supervisory team, we thought that the factors described here were most likely to have impacted on GPs' and patients' experiences.

Formal sample size calculations are inconsistent with many of the assumptions of qualitative research (Braun and Clarke, 2016). A common approach to determining adequate sample size in qualitative research is the concept of data saturation. This is usually described as the point at which sufficient data have been collected and that no new themes are apparent upon the collection of further data. This has some conceptual and operational limitations with respect to thematic analysis generally and my specific theoretical assumptions. Saturation implies that it is possible for interpretation of data to cease at a certain point and that all meaning exists in the data, awaiting discovery by the researcher. As such, it is aligned with a post-positivist epistemology and is less theoretically consistent with the critical realist ontology that I have adopted (Braun and Clarke, 2019b). An alternative to data saturation for guiding sample size is information power (Malterud, Siersma and Guassora, 2016). The concept of information power indicates that the more relevant information (which is relevant to the aims of the study) that a sample holds, the lower the number of participants needed overall.

I considered both data saturation and information power when determining adequate sample size for the qualitative study. I reflected on the richness of the information in the dataset and the extent to which this helped me meet the aims and objectives of the study as I was generating and analysing the data. I also used a consideration of whether new ideas and meanings were arising to help determine whether a sufficient number of participants had been interviewed.

---

care); and locum GPs (self-employed GPs who contract their services to providers on a temporary basis).

4.3.3. Data generation and analysis

I used semi-structured interviews (Green and Thorogood, 2018) to generate data. This method was chosen because there were key areas I wanted to cover in the interviews. To guide this, a semi-structured topic guide was initially developed from the literature and in discussion with my supervisory team, then modified iteratively as data generation and analysis progressed. The analytic method I chose to use was thematic analysis (TA), as developed by Braun and Clarke (Braun and Clarke, 2006). TA involves the identification of themes, which go beyond the categories and subject headings and are interpretative, aiming to capture common meanings, concepts or "stories" across the wider dataset. Braun and Clarke draw a distinction between TA that is "small q" (incorporating some post-positive concepts) and "Big Q" (adhering strictly to a qualitative methodology, which usually adopts a constructivist epistemology). They make the point, however, that TA is not a "ready-made" methodology (with a predetermined theoretical framework) but has some flexibility, which necessitates design thinking on the part of the researcher (Braun and Clarke, 2021b). The specific approach to TA in this study was determined prior to data generation, with support from my supervisory team.

There are several recognized approaches to TA: coding reliability, reflexive and codebook TA (Braun and Clarke, 2021b). A coding reliability approach is a deductive approach, where themes are often formed early in the analytic process and typically align closely with topic summaries or a pre-existing theoretical framework, rather than capturing a pattern of shared meaning identified inductively from across the data. Reflexive TA is a more inductive approach and fits better with a constructivist philosophical approach. I wanted to take a critical realist orientation to the data and found that a form of codebook TA, which incorporated elements of small q and Big Q approaches, best suited my research question.

Principles of constant comparison were used to guide the TA, wherein new data were systematically compared to previously analysed data, looking

for similarities and differences. The Constant Comparative Analysis Method (CCA) was initially developed as part of a Grounded Theory methodological approach to qualitative research (Fram, 2013). There is good evidence, however, supporting the use of CCA as an analytic method outside the Grounded Theory paradigm (Fram, 2013). It has been used extensively with TA in qualitative health research (Oh et al., 2020; Mughal et al., 2021). Adopting principles of constant comparison within a TA approach helped ensure that all data were systematically compared to other data in the dataset. It also allowed a consideration of data saturation, as discussed in the previous section. A further benefit of using constant comparison is that all data produced are analysed rather than being disregarded on thematic grounds and the approach can be adapted for either deductive or inductive approaches to analysis (Fram, 2013).

Semi-structured interviews (as opposed to narrative or unstructured interviews) can tend to lead to a small q, post-positive framework, where codes are generally developed early in the analysis and tend to align with the headings in the topic guide. I did, however, adopt some facets of what Braun and Clarke describe as a Big Q approach (Braun and Clarke, 2022). I reject the premise of an unbiased, objective researcher and much of the report of the qualitative study includes my reflections on reflexivity and my acknowledgement of how this has shaped the analysis, rather than having been "controlled for". In the case of my qualitative study, initial analysis was mostly inductive and was undertaken concurrently with data collection. Although I brought my knowledge and experiences as a GP to the data analysis, I otherwise approached the data items without a pre-existing or theoretically-driven coding framework, and began to assign initial codes. Therefore, the analysis was carried out primarily within an inductive framework, and codes and then themes were generated "bottom up" guided by the data. The exception to this was the analysis of the data around use of prediction models and communication of risk in general practice (Chapter Nine). For this part of the study, questions had been developed with a view to providing shorter and more focussed answers, with analysis being more deductive.

While I used a codebook approach to TA, I made the decision that all the data did not need to be double-coded. The double-coding approach is sometimes taken by researchers adopting TA, sometimes alongside statistical measurement of inter-coder reliability. Braun and Clarke (2021) suggest that this is an inappropriate projection of a post-positivist viewpoint onto TA. In my study, the data were however analysed in collaboration with my multidisciplinary supervisory team (GP, academic clinical psychologist, and psychiatrist) and PAG. This approach allowed a sharing of different perspectives throughout data interpretation and is one of the approaches taken to increase the trustworthiness of my analysis (Henwood and Pidgeon, 1992). A random sample of interview transcripts were allocated to my supervisors and we coded and discussed several of these together. My supervisors discussed my analysis with me and challenged some of my assumptions and coding structures, to ensure that the coding and thematic framework that I was constructing was as nuanced, rich and considered as possible, given the data. Throughout the analysis, I was cognizant of my own subjectivity and how this influenced my interpretation and findings. I reflected on how this affected my conclusions through the process of reflexivity.

I also took the decision not to use member checking (a process whereby transcripts are shared with interview participants for checking), which is also a realist approach which makes ontological and epistemological assumptions that are not consistent with my approach to TA (Tracy, 2010). There is also no evidence that member checking enhances the credibility or trustworthiness of qualitative research (Thomas, 2017). Discussions with the PAG aimed to ensure that the analysis made sense to people with lived experience of depression.

Finally, Braun and Clarke outline a further distinction, between experiential and critical qualitative research (Braun and Clarke, 2013). The approach in this study was in part experiential; my aim was to present an interpretive account of the perspectives and experiences of people with lived experience of depression and of GPs and locate this account within the NHS and its wider social context. However, I also aimed to adopt a critical

orientation to the data with a view to understanding some of the reasons for those experiences. I wanted to understand participants' experiences in the wider context of the health service and explore why certain of the topics covered are the way they are and what could be modified to facilitate change. In that respect, it was useful to compare and contrast the views within groups, but also the ways in which the views of people with lived experience and GPs offered different, sometimes contradictory, perspectives on the same phenomena. I have discussed this further when reporting the findings from the qualitative study (Chapters Eight and Nine).

4.4. Mixed Methods Approach

Having discussed the methodological approaches to the quantitative and qualitative components of the study, I will now describe the rationale and approach to adopting mixed methods in this study. Mixed methods is an established research paradigm in primary care research (Creswell, Fetters and Ivankova, 2004). Mixed methods research has been described as the "integration of qualitative and quantitative research methods in a sustained programme of inquiry" (Fetters, 2020). It is differentiated from multi-methods research in that, in mixed methods research, integration of data and analyses precedes reaching conclusions about the question under investigation (Bazeley, 2018).

A benefit of using mixed methods Is the application of two methodological approaches to the same problem. Qualitative and quantitative are often not sufficient to answer a research question in isolation whereas, by integrating the approaches, one can end up with a deeper and more thorough understanding of the problem (Creswell, Fetters and Ivankova, 2004). Quantitative research is often less able to answer the "how?" and "why?" questions necessary to understanding a complex system like a health service. Qualitative research is more suited to understanding situations and problems but often limited by smaller sample sizes and potentially less generalisability that quantitative research. Whereas the

quantitative workstream in this study was looking to predict an outcome (relapse) in a sample with a view to generalising the findings to the wider primary care patient population, the qualitative work sought to understand the meaning of peoples' perspectives and experience in their specific context. This made it possible to explore areas of nuance and contradiction, for example where views on the same subject diverged between patients and GPs, or even within groups. This provided interesting material which cast more light on the complexities of the individual human experience of being a patient (or a GP), whereas the quantitative work aims to identify a pattern across a broad dataset which was as focussed and parsimonious as possible. By employing a mixed methods approach and integrating these findings, I aimed to benefit from the strengths and mitigate the weaknesses of both approaches, to ensure that the conclusions drawn were grounded in what is meaningful to patients and healthcare professionals, and to generate an understanding of the problem that is more than a sum of its parts (Tariq and Woodman, 2013).

Integration has been described as the interaction between the quantitative and qualitative components of a study and is an essential part of mixed methods research (O'Cathain, Murphy and Nicholl, 2010). Data integration can occur at two different points during the research process: the analytic phase or the interpretative phase. I have adopted a convergent mixed methods design, meaning that the quantitative and qualitative analyses took place concurrently were mixed primarily at the interpretation stage (Tariq and Woodman, 2013). In this thesis, the data from the two studies were initially analysed separately and then combined using triangulation techniques, which occurs at the interpretation phase of the study rather than the analysis phase. Triangulation involves comparing the findings from the different methods and looking for agreement (convergence), complementarity or discrepancy (O'Cathain, Murphy and Nicholl, 2010). Because the workstreams were carried out concurrently, there were some instances where the analysis of one methodological component directly influenced the analysis of the other component, but this was not the principal approach to data integration.

There is some criticism or concern from researchers about combining two distinct, ostensibly incompatible research paradigms (i.e., quantitative and qualitative) within one study (Murphy et al., 1998). I have taken a pragmatic stance in my approach to mixed methods in this study and think the potential benefits and opportunity for increased understanding as a result of combining two different approaches outweigh these important yet more abstract concerns. Indeed, pragmatism has been described as a "philosophical foundation for mixed methods research" which allows "convergence on an epistemological middle ground" (Yardley and Bishop, 2015). The methodological approaches and methods I have outlined have been carefully designed with the aims and objectives in mind and I will reflect on the limitations of both components and the integration separately in the Discussion chapter.

The quantitative and qualitative components are presented separately; quantitative results are presented in Chapter Six and qualitative findings in Chapters Eight and Nine. The integration of findings is reported in Chapter Ten.

4.5. Patient and Public Involvement and Engagement

Public and Patient Involvement and Engagement (PPIE) has underpinned this work from its conception, prior to writing the NIHR Fellowship application, and continued throughout. PPIE is increasingly recognised and encouraged as a way of making health research more valuable for people with lived experience of the conditions being researched and increasing research impact. The NIHR provide specific guidance on the implementation of PPIE (formerly known as INVOLVE). They describe PPIE as "research being carried out *with* or *by* members of the public" rather than "*to, about* or *for*" them (NIHR, 2021). INVOLVE guidance (as it was known at the time) was adhered to throughout the project.

PPIE shaped the project from its early stages and was an important part of the Fellowship funding application. I was awarded a grant from the Yorkshire and Humber Research Design Service in July 2017 and used this to fund and facilitate an initial consultation focus group with six lay people with lived experience of depression, recruited through the Tees, Esk and Wear Valleys Trust. We discussed initial ideas for the proposal and used the event to identify research priorities and shape the methodology. They provided specific comments on the funding application, particularly the plain English summary. This group has remained involved throughout and formed the PAG, who have guided me on all parts of the study.

Since beginning the PhD, I built on this earlier work by establishing a collaboration with the University of York's PPIE network, Involvement@York. This enabled me to be guided by in-house, specialist PPIE advice and to advertise more widely to broaden the experience and demographic make-up of the group. The Involvement@York Network Manager attended a number of PAG meetings throughout the project, as well as providing support and expertise, which was helpful for me to evaluate my own professional development in terms of successfully implementing PPIE.

Hughes and Duffy (2018) identified five operational definitions of PPIE: undefined involvement; targeted consultation; embedded consultation; co-production; and user-led research (Hughes and Duffy, 2018). I have employed a combination of embedded consultation and collaboration/co-production throughout. Embedded consultation is where members of the public are consulted regularly throughout the research cycle, giving feedback on research ideas and dissemination plans. Co-production is where members of the public contribute to key decisions and findings as part of a steering group. Members of the PAG have co-authored a peer-reviewed editorial (Moriarty et al., 2020) and a blog article (Moriarty et al., 2021a) with me as patient representatives, and all participants have been involved in writing lay summaries to communicate findings to service users as they arise. They were also involved in drafting qualitative research materials and in reviewing and revising the semi-structured topic guides.

Hughes and Duffy (2018) also identified the following features as being important to successful and meaningful PPIE: clear and agreed meaning and purpose of involvement, reciprocal relationships and value and recognition of the expertise of those involved. Throughout this project, all participants were remunerated for time and expenses in line with INVOLVE guidance (NIHR, 2021) from the Fellowship budget. Mutual goals and expectations have been agreed at each meeting through discussion with participants. Evaluation forms completed by participants have reported that they have felt valued and able to contribute to key decisions. I have supported PAG members to undertake specific research training within the University where they have expressed an interest to do so.

In summary, the PAG were involved in this study in four key ways:

1. Commenting on drafts of protocols and materials for ethics applications, and developing patient information materials for the qualitative study.
2. Advising on logistical aspects of the research methods, from the perspectives of patients with experience of depression.
3. Interpretation of findings and meaning for people with depression.
4. Informing plans for dissemination, including how best to communicate the results of the work to patients and ensuring materials were in accessible language and format for publication in service-user literature.

4.6. Summary

This chapter has provided an overview of the methodological approach taken throughout this study. What follows is a mixed methods study, incorporating a convergent approach to integrating findings from a systematic review of prognostic models, a quantitative prognostic development and validation study, and a qualitative study with people with lived experience of depression and GPs. In the next chapter, I will describe the methods used in the quantitative study.

Chapter Five

Quantitative Methods

5.1. Introduction

This chapter[7] outlines the methods for the development and validation of a novel prognostic model to predict an individual's risk of relapse of depression in a primary care setting (the PREDICTR[8] study). The model was intended to be implemented in clinical practice for use by primary care health professionals to enable optimal shared decision making with patients. The priorities, therefore, were that the prognostic model be accurate, generalisable and, although ultimately beyond the scope of this current study, effective (i.e., result in demonstrably improved outcomes for patients). In order to be implemented in practice, it also needed be clinically credible and have face validity to healthcare professionals and patients.

As reported in Chapter Three, we currently lack evidence-based tools to assist clinicians with risk predictions of depressive relapse in any clinical setting. While there have been some previous attempts to develop relapse prediction models for depression, these pre-existing prognostic models have some drawbacks with respect to successfully predicting relapse in a primary care context. Critical appraisal of these studies found that the majority of these studies were at high overall risk of bias. The most significant limitations were

---

[7] This chapter is adapted from the published, pre-registered protocol article (Moriarty et al., 2021d), see Appendix 5.1. All necessary permissions have been sought to reproduce this. The study and analysis plan were also pre-registered at ClinicalTrials.gov ID: NCT04666662

[8] The development and validation of a prognostic model to PREDICT Relapse in primary care.

inadequate sample size, inappropriate handling of missing data, and presentation of inappropriate performance statistics (calibration and discrimination not assessed). Furthermore, the developed models have either demonstrated insufficient predictive performance on validation (Klein et al., 2018), or they could not be feasibly implemented in a primary care setting due to the large number and type of included predictors (van Loo et al., 2020).

5.2. Objective

To develop and validate a multivariable prognostic model[9] to predict relapse within six to eight months in patients with remitted depression in primary care.

5.3. Methods

The methods for this study were developed in accordance with the PROGnosis RESearch Strategy (PROGRESS) initiative (Riley et al., 2019b; Steyerberg, 2019). The study is reported in this thesis in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) Statement (Moons et al., 2015) and the recently published TRIPOD-Cluster guidance for reporting prediction models developed or validated using clustered data (Debray et al., 2023).

The study used IPD from RCTs and a cohort study, therefore elements of the Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA-IPD) statement were also relevant (Stewart et al., 2015). However, this study is not a systematic review and the aim is not to provide a summary of a complete body of

---

[9] An equation to estimate an individual's probability of relapse within 6-8 months, conditional on the values of several predictors.

research and so not all PRISMA-IPD items are applicable (Appendix 5.2). The PAG inputted to several aspects of this study, including selecting predictors and their measurement (for example, commenting on the acceptability of validated diagnostic instruments for depression and anxiety symptoms), definition of outcome, target patient population and clinical application. The study was registered prospectively on clinicaltrials.gov (available: ClinicalTrials.gov ID: NCT04666662).

5.3.1. Source of data

A dataset (the "PREDICTR dataset") was formed using combined IPD from UK primary care-based studies. Cohort studies and RCTs are recommended sources of data for the development of prognostic models (RCTs are essentially cohort studies, usually with excellent data quality at baseline and over follow-up) (Pajouheshnia et al., 2019).

I had IPD readily available in a pragmatic sample of three RCTs [CASPER Plus (Bosanquet et al., 2017), REEACT (Gilbody et al., 2015), and REEACT-2 (Gilbody et al., 2017)] and a cohort study [the West Yorkshire Low Intensity Outcome Watch (WYLOW) (Ali et al., 2017), a longitudinal cohort study following up patients after low intensity cognitive behavioural therapy (LiCBT) through the Improving Access to Psychological Therapies (IAPT) service]. These were all studies carried out within the Mental Health and Addictions Research Group at the University of York.

In order to increase the sample size available for model development, I identified further studies: first, by searching all NIHR-funded RCTs of primary care-based interventions for depression and second, by reference to the search results from a recent IPD meta-analysis of RCTs of depression interventions (this meta-analysis had searched for studies that had used the CIS-R as a measure of baseline severity and provided a recent search of relevant studies) (Buckman et al., 2020).

To be eligible for inclusion in this study, I specified that RCTs must:

- Include adult patients (18 years and over) with depression, and measure depressive symptoms at a minimum of three time-points (to enable diagnosis of depression, remission, relapse/no relapse). I excluded RCTs in patient groups with significant psychiatric or medical comorbidity. I also excluded feasibility studies, as these typically have smaller sample sizes and shorter follow-up periods;
- Have sufficient follow-up after participants reached remission to allow the identification of relapse (or no relapse) within at least six months;
- Use the Patient Health Questionnaire (PHQ-9) as a measure of depression.

This search added three further RCTs: COBRA (Richards et al., 2016), CADET (Richards et al., 2013) and the Healthlines Depression RCT (Salisbury et al., 2016). I received IPD for all studies identified. All of the included studies had pragmatic and unrestrictive inclusion criteria, and so the participants were deemed representative of the target general population. In summary, the final PREDICTR dataset is derived from all arms (control and intervention) of six RCTs of low-intensity primary care-based interventions for depression (CADET, CASPER Plus, COBRA, Healthlines Depression, REEACT, and REEACT-2) and one observational cohort study (WYLOW). See Table 5.1 for details of the final included studies.

Table 5.1: Summary of primary sources of IPD

Study	N	Study type	Inclusion criteria	Length of Follow-up	Follow up Points	Mean age (SD)	Gender (% Female)	RCT Intervention	Duration of RCT Intervention
CADET	581	RCT	Adults with depression	12 months	0, 4,12	44.4 (13.3)	71.9	Collaborative care	14 weeks
CASPER Plus	485 (358 at 12 m)	RCT	65 years or older with depression	18 months	0, 4, 12 and 18 months	Intervention group: 71.9 (6.03) Control: 71.6 (5.96)	Intervention group: 59.1 Control: 63.1	Collaborative care	8-10 weeks
COBRA	440	RCT	Adults with depression	18 months	0, 6, 12, 18	43.5 (14.1)	66	Behavioural Activation vs CBT	16 weeks
Healthlines Depression	609	RCT	Adults with depression	12 months	0, 4,8,12	Intervention group: 49.1 (12.9) Control: 50 (12.8)	Intervention group: 69 Control: 68	Complex intervention (Integrated telehealth)	12 months
REEACT	461	RCT	Adults with depression	24 months	0, 4, 12 and 24 months	39.86 (12.65)	67	cCBT	6 weeks

REEACT-2	369	RCT	Adults with depression	12 months.	0, 4 and 12 months	40.6 (13.8)	64.5	cCBT	4 months
WYLOW	439	Longitudinal observational cohort study	Adults with depression	12 months (Start-point = Remission)	Monthly	41.28 (14.59)	59.7	None (cohort were followed up after LiCBT)	NA

5.3.2. Participants

Adult participants (aged 18 years and over) with depression. The included participants did not have significant psychiatric comorbidity (e.g., schizophrenia, bipolar affective disorder).

5.3.3. Setting

All data are from UK-based primary care or community-based settings.

5.3.4. Start-point (remission)

There are three important time-points in the data for defining the start- and end-points for the prognostic model (Figure 5.1):

- Baseline: the point at which participants were depressed (all studies) and randomised (for RCTs);
- Follow-up 1 (FU1): four months after baseline; to diagnose remission. All participants in WYLOW met this criterion. This is t=0 (point of prediction) for this prediction model study.
- Follow-up 2 (FU2): t=1 (occurs at either 6 or 8 months after t=0). This is where the predicted outcome occurs; participant either relapses or does not relapse.

The PHQ-9 is a screening tool for major depressive disorder and a cut-off of 10 or more is used to detect clinically significant depressive symptoms (Kroenke, Spitzer and Williams, 2001). In all RCTs, the majority of participants met criteria for a diagnosis of depression at baseline. Any participants identified to have a baseline PHQ-9 less than 10 were excluded from the analysis. In WYLOW, all participants in the dataset received had reached remission, and details of depression at baseline were provided.

	CADET	CASPER Plus	COBRA	Healthlines Depression	REEACT	REEACT-2	WYLOW
Total number in dataset received	n=581	n= 485	n= 440	n= 609	n= 691	n= 454	n= 439
Number confirmed depressed at baseline	n=537	n= 372	n= 414	n= 609	n= 691	n= 454	N=327
Number of participants remitted (t=0)	n=161	n= 118	n= 185	n= 118	n= 250	n= 173	n= 327
Number with outcome data at Follow-up 2 (t=1)	n=158	n= 101	n= 169	n= 110	n= 221	n= 159	n= 326

Baseline

Follow-up 1

Follow-up 2 (6-8 months later)

Final PREDICTR dataset (n=1,244)

Figure 5.1: Flow diagram of participants in PREDICTR dataset

The start-point (or time of intended prediction) was FU1, the point at which a participant, who started treatment with case-level depression, had entered remission. Remission was identified as a participant who had case-level depression at baseline (a PHQ-9 score of 10 or more, as above) having: i) a post-treatment PHQ-9 score below the established cut-off of 10 at 4 months after trial baseline [this is consistent with clinical recovery (Clark et al., 2009; Ali et al., 2017)] and ii) an improvement of ≥5 points on the PHQ-9 [which aligns with the established reliable change index used to identify those with "reliable improvement" (McMillan, Gilbody and Richards, 2010)].

5.3.5. End-point (relapse)

Participants were coded as relapsed if they fulfil the following criteria within six to eight months post-remission (by FU2): i) PHQ-9 score above the diagnostic cut-off (10 or more) and ii) ≥5 points greater than their symptom score at the time of remission. As above, this is consistent with accepted criteria for reliable and clinically significant deterioration (Jacobson and Truax, 1991; McMillan, Gilbody and Richards, 2010).

The main reason for specifying the prediction horizon at six to eight months rather than a single time point is pragmatic and based on the available data (the time between FU1 and FU2 is eight months for six of the seven RCTs and six months for COBRA). As discussed in the Chapter Two, relapse is most commonly operationalized as occurring between six- and twelve-months post-remission (Beshai et al., 2011) and the majority of people who do relapse do so within the first six months (Ali et al., 2017), so the time-frame used in this study was thought to be appropriate.

5.3.6. Predictors

Predictors were identified *a priori* (Harrell Jr, 2015), following a literature review and on clinical grounds (based on discussion with my multidisciplinary supervisory team and the PAG supporting the study). Umbrella reviews (reviews of other systematic reviews and meta-analyses) are one of the highest levels of evidence for determining associations between predictors and outcomes when selecting predictors for inclusion in a prognostic model (Fusar-Poli et al., 2018). An umbrella review of prognostic factors associated with increased risk of relapse and recurrence guided the selection of candidate predictors for inclusion in the model (Buckman et al., 2018). A further systematic review of prognostic factors, published after the umbrella review, supported those findings and was also used to guide the included predictors (Wojnarowski et al., 2019). In addition to this, as reported in Chapter Three, I reviewed all existing prognostic models for predicting relapse or recurrence to explore other predictors used.

All candidate predictors are based on self-report or clinical information and I did not include, for example, biomarkers and in-depth neuropsychological testing in an effort to ensure that the model is acceptable and usable in a primary care setting (Kessler, 2018). Categorisation of continuous predictors was avoided where possible in order to avoid loss of information and power to detect an association between predictors and outcomes (Riley et al., 2019b).

5.3.6.1.   Predictors in primary analysis

The following variables have robust evidence for their role as relapse predictors and were included in the model:

5.3.6.1.1.   *PHQ-9 score at remission (residual depressive symptoms)*

Residual depressive symptoms are strongly established as a predictor of relapse (Buckman et al., 2018; Wojnarowski et al., 2019). This predictor

was operationalized in this study using the PHQ-9 score. The PHQ-9 is routinely used in primary care (Kroenke, Spitzer and Williams, 2001). As above, remission was defined as a PHQ-9 score below 10 (and 5 or more points below the score when depressed) and residual symptoms are defined as a PHQ-9 at remission of between 5-9 (McMillan, Gilbody and Richards, 2010). Per the inclusion criteria for my study, all participants met criteria for remission (i.e., PHQ-9 score of below 10); PHQ-9 score at remission (0-9) was modelled as a continuous variable rather than binary (e.g., presence or absence of residual symptoms).

5.3.6.1.2. Number of previous episodes of depression

There is strong evidence that this is a significant relapse predictor (Buckman et al., 2018; Wojnarowski et al., 2019), albeit slightly weaker than for residual symptoms. I modelled this as a dichotomous predictor. The coding of this variable in the original RCTs was variable (i.e., a combination of continuous and dichotomous), and so it was not possible to model as a continuous variable in this study. While there is some weak evidence that the relapse risk increases with each successive depressive episode, the prognostic effect of previous episodes on relapse is strongest when comparing any number of previous episodes to no previous episodes (Buckman et al., 2018). This finding from the pre-existing literature is likely to be helpful for a primary care based prognostic model, as there may be potential difficulty in achieving a precise number of previous episodes in clinical practice. In this study, I modelled this predictor as a dichotomous variable (0=no previous episodes, 1=one or more previous episodes) and accepted participant report, GP report or documentation in GP records (depending on how it had been captured in the original studies).

5.3.6.1.3. Comorbid anxiety

There is good evidence that comorbid anxiety predicts relapse or recurrence of depression and it was included as a predictor in the model (Buckman et al., 2018; Wojnarowski et al., 2019). The GAD-7 is a valid tool for screening and assessing severity of Generalised Anxiety Disorder in

clinical practice (Spitzer et al., 2006). Pre-treatment symptoms (i.e., those at baseline) seem to be more predictive of relapse than those at depressive remission (Wojnarowski et al., 2019). The pre-registered protocol stated that pre-treatment GAD-7 score would be used where it was available in the IPD datasets; otherwise, I planned to use the GAD-7 at remission (t=0). GAD-7 score was planned to be modelled as a continuous predictor.

Inspection of the data received showed that the GAD-7 was not used in one of the original studies (REEACT). In REEACT, the Clinic Interview Schedule-Revised (CIS-R) was used as an alternative anxiety measure. The CIS-R is a computerised score that establishes the nature and severity of neurotic symptoms and the presence of a depressive episode according to ICD-10 criteria. Each section scores a particular type of neurotic symptoms, ranging in severity between 0 and 4 (Lewis et al., 1992). The CIS-R includes a specific anxiety subscale (see Appendix 5.3 for full details).

To avoid losing data for analysis, I retained REEACT in the analysis. Because the GAD-7 and the CIS-R anxiety subscale are both composite scales, they can be combined and used to create a composite score for comorbid anxiety, using standardised scores, or z-scores. Z-scores use the mean and standard deviation of the measure within the population (in this case, the original full dataset) to calculate a standardised score for each individual participant within the sample (using the formula $Z = (x – M) / SD$, where M=population mean, x=the individual measurement and SD=population standard deviation) (Andrade, 2021). This was done for each GAD-7 score within each original dataset, to ensure that each participant was described with respect to their original cluster. The sum of the scores on the CIS-R anxiety subscale was calculated for REEACT and this was also converted to a z-score. The data from the seven clusters were then combined to form the final IPD dataset. A sensitivity analysis was performed (excluding REEACT and using GAD-7 scores as originally planned) to assess the impact of this decision on the final results.

5.3.6.1.4. Severity of episode

There is reasonable evidence that the baseline severity of the *index* episode is a prognostic indicator of greater odds of relapse (Buckman et al., 2018). This was measured using the PHQ-9 score at baseline (pre-treatment) rather than that at the point of prognostication (remission). The PHQ-9 score at the point of depression diagnosis was modelled as a continuous predictor.

5.3.6.1.5. RCT Intervention

Because the data are drawn from RCTs, I thought it important to be mindful of the fact that approximately half of the participants have received a treatment (above usual care) and the other half have not. Where such treatments have been found to be effective, not modelling the effect of different treatments can lead to unreliable risk predictions when the model is validated in a different population. Excluding the treated individuals would have meant losing half of the available data, and so a preferable option was to explicitly model for treatment effect when developing the prognostic model (Groenwold et al., 2016; Pajouheshnia et al., 2017).

The treatments In all RCTs were acute-phase psychological treatments rather than relapse prevention interventions, and therefore their effects specifically on relapse outcomes is unclear. One of the studies [Healthlines Depression (Salisbury et al., 2016)] did include an element of relapse prevention beyond the acute phase treatment (advisors phoned the participants every two months to check how they were getting on and encourage them to keep following the intervention advice). The interventions were also heterogeneous and so they may have affected relapse outcomes in different ways. To avoid overcomplicating the model, I coded the presence or absence of an effective intervention as a dichotomous variable. I defined an e'fective Intervention by whether Individual participants entered remission after receiving an RCT intervention that was demonstrated to be effective (based on the results of the RCT) (code=1) or whether they entered remission after receiving a control or ineffective intervention (code=0). Note

that this predictor is intended to control for the intervention as part of the model building process only. When making predictions in real-world general practice, the intention is that this predictor would always be set to zero (i.e., no experimental intervention present).

Table 5.2: Summary of selected predictors for primary analysis

Predictor	Type of data	Method of measurement	Range of values and coding of predictors
PHQ-9 score at point of remission (residual depressive symptoms)	Continuous	PHQ-9 score at remission (t=0)	Range from 0-9
Number of previous episodes of depression	Categorical	Participant or GP report (No previous episodes vs any previous episodes)	No previous episodes=0; 1 or more previous episodes=1
Comorbid anxiety	Continuous	z-score (GAD-7 or CIS-R anxiety subscale)	Not applicable
Severity of episode	Continuous	PHQ-9 score at baseline	10-27
RCT Intervention	Categorical	Presence or absence of effective treatment	Remission after receiving control or ineffective intervention=0; Remission after receiving effective intervention=1

5.3.6.2.   Exploratory predictors

The following are less well-evidenced predictors of relapse or recurrence of depression and I included them as part of an exploratory secondary analysis:

5.3.6.2.1.   *Age*

While epidemiological studies have found that age can be associated with increased prevalence of depression (McManus et al., 2014), there is not strong pre-existing evidence that it is a significant predictor of relapse or recurrence of depression (Buckman et al., 2018; Wojnarowski et al., 2019).

5.3.6.2.2.   *Relationship status*

The literature suggests that, while marital status is associated with onset of depression, it is not a significant predictor of depressive relapse (Buckman et al., 2018; Wojnarowski et al., 2019; Burcusa and Iacono, 2007). In my systematic review of prognostic models, discussed in Chapter Three, one included study found that marital status contributed to prediction of recurrence of depression over three years (Wang et al., 2014) and another included "having a partner or not" as one of two significant predictors in a model to predict relapse over 12 months (with a regression (beta) coefficient of -2.14 (95% CI 0.02 to 0.64) p = 0.01) (Johansson, Lundh and Bjärehed, 2015). As discussed, both of these studies had methodological weaknesses that identified them as high risk of bias. Given the strength of the association reported, particularly in the latter study (Johansson, Lundh and Bjärehed, 2015), I identified this as a predictor for exploratory analysis.

5.3.6.2.3.   *Multimorbidity*

Multimorbidity is defined by NICE as the presence of two or more long-term conditions, which can include physical or mental health conditions (NICE, 2016). The combination of depression and another long-term condition has the most adverse effect on quality of life and leads to worse

physical health outcomes (Moussavi et al., 2007). The extant literature suggests that physical co-morbidity/multimorbidity is not a significant predictor or relapse or recurrence (Kok et al., 2013; Buckman et al., 2018). It is, however, of Interest and relevance to primary care to better understand the association between multimorbidity and relapse. The presence of one additional long-term condition meant that participants met the NICE criteria for multimorbidity, given that all included participants already had a diagnosis of depression.

5.3.6.2.4. *Employment status*

Socioeconomic status is associated with onset of depression (Burcusa and Iacono, 2007) but has now been fairly robustly excluded as a statistically significant predictor of relapse of depression (Buckman et al., 2018; Wojnarowski et al., 2019; Ali et al., 2017; Burcusa and Iacono, 2007). Similarly, the literature available suggests that employment status is not a significant predictor of depressive relapse (Wojnarowski et al., 2019). However, a recent prognostic model study found tentative evidence that employment status is a predictor of relapse (Lorimer et al., 2020). The study in question used machine-learning techniques on a small sample size and included people with mixed depression and/or anxiety, which may limit its applicability to a depression-only cohort. I included employment status as a predictor in the secondary analysis. I used the definition of unemployment as being those of working age who do not have a job and are actively seeking one (Bartley and Ferrie, 2001).

5.3.6.2.5. *Gender and ethnicity*

The literature suggests that gender and ethnicity are not significant predictors of relapse (Buckman et al., 2018; Wojnarowski et al., 2019; Burcusa and Iacono, 2007). Gender and ethnicity data are routinely collected as part of RCTs and are often included in prognostic models (Riley et al., 2019b). I included them as exploratory predictors in this study to guide future prognostic model research in this area.

5.3.6.2.6.	*Current antidepressant use*

This predictor was included in the study as an exploratory predictor after the registration of the protocol, but prior to statistical analysis. The qualitative study identified current antidepressant use as being perceived to be an important factor associated with relapse, particularly by the GPs interviewed. As part of the convergent mixed methods approach, I decided to explore its statistical association with the outcome of relapse in the quantitative study. This was operationalised as a binary variable and included any antidepressant medication, prescribed for the indication of depression, which was being taken by participants at the point of prediction (remission).

Table 5.3: Re-categorisation of categorical variables for analysis

Variable	Original categories (in RCTs)	New categories (PREDICTR)
Ethnicity	White	White
	Mixed	Other
	Black	
	Asian	
	Chinese	
	Other	
Employment status	Employed (Full time or part time)	Employed/not seeking employment
	Student	
	Retired	
	House-person	
	Unemployed due to ill-health	
	Other	
	Unemployed and seeking work	Unemployed
Relationship status	Married/civil partnership/cohabiting/relationship	In a relationship
	Single	Single
	Separated	
	Divorced	
	Widowed	
Multimorbidity	None	No long-term physical health condition
	Mental health only	
	Diabetes	One or more long-term physical health conditions
	Asthma or COPD	
	Degenerative or inflammatory arthritis	
	Heart Disease	
	Stroke	
	Cancer	

5.3.7. Sample size

Ensuring an adequate sample size allows for more accurate estimation of regression coefficients and reduces the potential for overfitting. Rules of thumb for calculating required sample size for prediction models with binary outcomes (such as ten Events Per candidate predictor Parameter (EPP)) are now considered too simplistic to provide robust estimates of minimum required sample size (Van Smeden et al., 2016). The actual required sample size is context-dependent and is informed by several factors (see below). I used the *pmsampsize* package (available online: https://riskcalc.org/pmsamplesize/) to calculate the required minimum sample size (Riley et al., 2020). The presence of clustering and between-study heterogeneity does not significantly impact on minimum sample size requirements (Wynants et al., 2015) and was therefore not considered in the calculations.

The Cox-Snell R^2 is a measure of overall model fit and based on the methods for sample size calculation (Riley et al., 2020), an anticipated Cox-Snell R^2 must be specified when calculating sample size, usually based on previous studies of similar patient groups/outcomes. No previous prognostic model study predicting relapse of depression identified so far has reported a Cox-Snell R^2 and so, to ensure an adequate minimum sample size, I used the recommended conservative estimated Nagelkerke R^2 of 15% (Riley et al., 2018). This corresponds to a Cox-Snell R^2 of 0.0945, assuming an overall outcome proportion of 0.2, which again is a conservative estimate based on the literature (Ali et al., 2017). I targeted an expected shrinkage factor (*S*) of 0.9 (to reflect small optimism in predictor effect estimates and thus low overfitting), as recommended (Riley et al., 2018).

To include all predictors, I required eight predictor parameters (*P*), which corresponds to PHQ-9 score at remission; Previous depressive episodes; Co-morbid anxiety; Severity of index episode; and RCT Intervention (including two parameters for each continuous predictor to account for potential non-linear trends). Therefore, the minimum required

134

sample size (*n*) is 722 (with 145 events) for these predictors. The actual sample size of the PREDICTR dataset (n=1244, with 261 events) exceeded this and, therefore, results were expected to be stable, with less uncertainty in the estimates obtained.

5.3.8. Missing data

To avoid loss of power and precision and to reduce the risk of biased estimates, complete case analysis (excluding participants with missing data) was avoided and missing data were handled using multiple imputation with chained equations (MICE) (Hughes et al., 2019; White et al., 2011). Missing values were imputed based on the values of all other predictors and the outcome, under a *missing at random* assumption[10].

Data were imputed using chained equations which included linear models for continuous predictors (residual symptoms, severity, comorbid anxiety) and logistic models for binary predictors [number of previous episodes, RCT Intervention, outcome (relapse/no relapse)]. Multiple copies of the dataset were created with identical known information and different imputed values, reflecting the uncertainty associated with imputation. Imputation was undertaken for each study separately, to preserve the clustering of participants within studies and any between-study heterogeneity in predictor effects and outcome prevalence. Each imputed dataset was then analysed separately using the same statistical methods, and the estimates were combined using Rubin's rules, to produce an overall estimate and measure of uncertainty of each regression coefficient (Rubin, 1987; Debray et al., 2023). I used thirty imputations, based on the percentage of participants with one or more missing values and in line with current

---

[10] *Missing at random* assumes that missing values occur according to a certain probability that depends on the observed data. Multiple imputation creates multiple versions of the dataset and imputes values of the predictor with missing values, drawn from their predicted distribution using the other observed data (in this case, the other predictors and the outcome).

guidance (which specifies at least twenty imputations, as long as this is greater than or equal to the percentage of participants with one or more missing values) (Riley et al., 2019b; White, Royston and Wood, 2011).

5.3.9. Statistical analysis methods

5.3.9.1.   Data pre-processing

Anonymised data were requested and data from all seven studies were acquired. Data were shared using data sharing agreements, transferred securely (using OneDrive or other password-protected links) and stored in line with the University of York's data storage policy (Smith et al., 2015). I am the named data custodian. Members of the original study teams providing data remained involved as collaborators on the project and were available to answer data queries and clarify issues as they arose. Data pre-processing involved merging the data into one storage system (technical harmonisation) and integrating datasets into a logically coherent entity (semantic harmonisation) (Debray et al., 2023). Data were received in the form of either Excel CSV files or Stata data files. The datasets were combined and harmonised to ensure consistency across trials (Dewidar et al., 2021). For the purpose of technical harmonisation, all data were opened in Stata and saved as data files for cleaning, while the originals were stored and unmodified. For the purpose of semantic harmonisation, a master codebook was created (Appendix 5.3) in line with the pre-specified harmonisation procedure outlined in the study protocol (Appendix 5.1). A study ID was created for each source of IPD to maintain clustering for data analysis.

All data received were summarised and checked against publications (e.g., with reference to study flow diagrams) for key features such as number of participants (total and in each study arm), demographics (age and sex) and primary outcomes of study, where applicable. I checked that I had data for all expected follow-up points and the key outcome data required at each outcome point. I ensured key predictor information was present, including a

variable identifying treatment and control arms (for the RCTs). Any discrepancies or irregularities were clarified through communication with the original authors. Validity of data values were checked on data inspection; invalid data-points were again clarified with study authors. Finally, missing data were quantified and compared with trial publications.

5.3.9.2.   Data integrity checks (risk of bias)

To assess IPD integrity, I compared numbers of participants in each treatment and control arm with those reported in the primary references. I checked the relapse rate within each arm and compared these across datasets. To define the quality of the IPD for prognostic modelling, I performed risk of bias assessment on the included datasets using the PROBAST risk of bias tool for prognostic model studies (Moons et al., 2019). Only the participants, predictors and outcome domains were pertinent; the analysis domain is used for assessment of prognostic model development and validation studies which does not apply to the RCTs and cohort study included in this study. The analysis methods used in those original studies do not impact on this prognostic model study and so the analysis domain of PROBAST was not used.

Once remission was identified this was represented as time t=0. Relapse was then coded at t=1 as 0=no relapse, 1=relapse, as described in "End-point" section (see Figure 5.1). Descriptive statistics were produced for all predictors and outcome data. Exploratory univariable analysis was performed to evaluate the unadjusted relationship between each predictor variable and the outcome variable, but not for the purpose of informing predictor selection.

5.3.9.3.   Model development (primary analysis)

Using data from multiple clusters (in this case, the individual studies from which IPD were acquired) has some advantages, in particular allowing

larger sample sizes by combining data from multiple clusters and reducing the risk of overfitting that can occur when developing a model in a single cluster. Penalisation and shrinkage[11] can be used to reduce the risk of model overfitting, although these approaches are less effective when used in datasets with small sample sizes (Riley et al., 2021). A further advantage of using data from multiple clusters for prognostic model development is that it can increase the generalisability and transportability of the model by including different but related source populations. The main point of relevance from a data analysis perspective when using clustered data is the importance of accounting for clustering. Even after harmonising and combining the IPD, data drawn from the same clusters will generally be more similar than that from other clusters. Within-cluster participants have been generally exposed to the same healthcare (or trial) staff, procedures and activities (Debray et al., 2023). Clusters can also differ in outcome occurrence, predictor effects and participant characteristics. Failing to account for clustering can result in suboptimal predictions (Falconieri et al., 2020).

The model was developed using multilevel multivariable logistic regression, with a binary (relapse/no relapse) outcome. Model parameters were estimated via 138npenalized maximum likelihood estimation, and then penalised post-estimation using a uniform shrinkage factor (see later). The modelling preserved the clustering of participants within studies, by having a random effect on the intercept, a random intervention effect (Heisig and Schaeffer, 2019), and also allowed for between-study correlation in these effects. The rationale for having a random slope on the intervention predictor was to account for the heterogeneity between the interventions delivered as part of the RCTs. As described earlier, these were acute-phase interventions occurring before the point of prediction. However, where these had been proven effective (or superior to usual care) within the RCTs, we could not be

---

[11] These are methods used to shrink predictor effect estimates towards the null and therefore reduce the mean-square error of predictions when the model is applied in new individuals (Riley et al., 2021).

certain they would not have had a differential effect on the outcome of relapse.

Stepwise methods for predictor selection are not generally recommended for prediction models as this has been reported to remove judgment of the analyst from the process of model development as well as leading to overfitting (because the performance of the model is estimated after testing for statistical significance of predictors in the same data) (Fusar-Poli et al., 2018). As described, I had selected the key predictors on the grounds of best available evidence and clinical acceptability, as well as practical reasons related to data availability. The list of predictors was felt to be of appropriate length so I avoided predictor selection techniques during model development and included all predictors regardless of their statistical significance ("full model" approach) (Harrell Jr, 2015). Multi-collinearity is not known to be an issue for prediction purposes (as the focus is on the model prediction, not the predictor effect estimates themselves) and I only planned to consider the need to exclude predictors due to collinearity if this were preventing convergence of the estimated model (Riley et al., 2019b).

I explored non-linear relationships in the modelling process using multivariable fractional polynomials[12] (MFPs), a flexible and recommended approach for modelling continuous predictors in medical datasets. The other recommended method for modelling continuous predictors is the use of restricted cubic splines, and while these two methods often result in similar models, there is some evidence that MFPs perform better than restricted cubic splines in the presence of simpler relationships and medium amounts of information (Binder, Sauerbrei and Royston, 2013; Steyerberg, 2019). I factored in two predictor parameters (beta coefficients) per continuous variable to account for this approach, as described in the "Sample size"

---

[12] Fractional polynomials use a limited, flexible set of transformations (e.g., powers of -2, -1, -0.5, 0, 0.5, 1, 2, 3) to describe predictor effects, rather than assuming a linear trend (Riley et al., 2019b; Royston and Altman, 1994).

section. The continuous predictors were also mean-centred during this process.

Discrimination (the ability of the model to differentiate between those who do and do not relapse) was assessed using the C-statistic. The C-statistic assesses the extent to which the model assigns a higher probability of relapse to an individual who did eventually relapse in contrast to an individual who did not. A value of 0.5 suggests no discrimination beyond chance, and a value of 1 indicates perfect discrimination. Calibration is a measure of the agreement between predictions from the model and observed outcomes. It was assessed by estimating calibration-in-the-large and calibration slope. Calibration was also assessed visually by producing calibration plots (plots comparing observed outcomes against predictions) and curves. These are summarised and defined in Table 5.4.

Predictive performance metrics (C-statistic for discrimination; calibration slope and calibration-in-the-large for calibration) were calculated for the final developed model. They were calculated within each cluster in turn, to quantify the extent of heterogeneity between clusters, and then pooled using random effects IPD meta-analysis (IPD-MA) to summarise the model's average performance. Prediction intervals were also constructed to calculate the model's likely performance in new but similar settings (Debray et al., 2023).

Table 5.4: Summary and definitions of predictive performance metrics used
in this study

Discrimination	
C-statistic	A measure of how well a model identifies those who will and will not develop the outcome of interest. A C-statistic of 0.5 indicates the model does this no better than chance, while a C-statistic of 1 indicates perfect discrimination.
Calibration	
Calibration slope	The extent to which observed and predicted risk of the outcome agree across the whole range of predicted values. Ideal value of one.
Calibration-in-the-large	A measure of mean calibration; the average predicted risk is compared with the overall event rate. This tells us whether the algorithm over- or underestimates risk in general. Ideal value of zero.
Calibration plot (with calibration curves)	A visual (graphical) representation of how closely observed and predicted risks agree.

7.2.3.2.   Internal validation

The optimism of the developed model was then assessed. Optimism describes the risk of obtaining misleading measures of predictive performance when this is assessed in the same dataset used for model development, mainly due to overfitting. Internal validation can be used to provide optimism-adjusted performance statistics to mitigate for this effect. Using internal validation techniques for this purpose has the advantage, for example over a single split-sample approach, of allowing all of the data to be used in model development.

In this study, optimism was measured using non-parametric bootstrapping. Bootstrapping is a means of resampling the original dataset (with replacement), to avoid the need for new validation data. One hundred bootstrap samples were produced from the original dataset (with each bootstrap sample also stratified by study).

Within each bootstrap sample, the same modelling procedures were used as for model development; continuous predictors were modelled using MFPs and multiple imputation was used to address missing data. As for model development, multiple imputation was done within-study to preserve clustering (with thirty imputations for each study) and results averaged across the thirty imputed datasets using Rubin's rules.

The model estimated from each bootstrap sample was applied in both the same bootstrap sample (apparent performance) and in the original (imputed) dataset (test performance). Each time, average performance measures were calculated by pooling within-study metrics, using meta-analysis.

Optimism was calculated as the difference between apparent and test performance; this process was repeated one hundred times and the average difference between the bootstrap (apparent) and test performance for each performance statistic provided the estimate of overall optimism for that statistic. Optimism-adjusted performance statistics (C-statistic, calibration slope and calibration-in-the-large) were derived.

The uniform shrinkage factor (in this study, the optimism-adjusted calibration slope) was applied to all of the original estimated beta coefficients (to shrink them toward zero to address overfitting) to produce a penalised logistic regression model. Finally, the intercept was then re-estimated (whilst constraining the penalised predictor effects at their shrunken value) to ensure overall calibration (calibration-in-the-large) was maintained. This formed the final model.

5.3.9.4.   Internal-external cross-validation

External validation is the assessment of a model's predictive performance in data from a different source, that was not used in the development process. It is a measure of a model's generalisability and performance in a range of populations and settings. To conserve information and to allow for all data to be used for model development, I did not perform a conventional external validation as part of this study. I did, however, have IPD from multiple studies and, therefore, generalisability and heterogeneity of the model performance was examined using internal-external cross-validation (IECV) (Royston, Parmar and Sylvester, 2004), as follows.

I excluded data from each primary study in turn and developed the risk prediction model in the remaining data, using the same model development approach as detailed (without shrinkage, as the purpose was primarily to explore the generalisability of the model). I then externally validated the developed model using the data from the excluded study. This process was repeated, each time omitting a different study, until the model had been fitted excluding each study once. Random effects meta-analysis was then used to summarise the performance across studies, to obtain summary measures of the model performance and estimates of heterogeneity in performance across studies. I ensured that each cycle of the IECV approach retained sufficient sample size for model development; in this manner, each cycle retained the majority of the available IPD for model development, and so the models produced in each cycle were likely to be similar to each other. A consistent model development strategy was used in each cycle of the IECV approach (Steyerberg and Harrell, 2016).

Predictive performance metrics (C-statistic for discrimination; calibration slope, calibration-in-the-large and visual inspection of calibration plots with LOESS-smoothed calibration curves for calibration) were calculated for the final developed model in each "external" validation (that is, when the model was applied in the study that had been left out). These were

calculated within each cluster in turn, to quantify the extent of heterogeneity between clusters, and then pooled using random effects meta-analysis to summarise the model's average performance. Using the meta-analysis results, 95% prediction intervals were also constructed to calculate the model's likely performance in new but similar settings (Debray et al., 2023).

5.3.9.5. Sensitivity analysis

A sensitivity analysis was performed measuring predictive performance statistics omitting REEACT and using GAD-7 as the measure of comorbid anxiety (per protocol). I also planned to conduct a sensitivity analysis omitting any studies which were not assessed as being at overall low risk of bias (using PROBAST), but this was not required as no such studies were identified.

5.3.9.6. Secondary analyses

The full model approach described has the advantages of not being overly data-dependent and avoids the risk of removing clinically important predictors from the final model (Harrell Jr, 2015). In the protocol, I had also planned an exploratory secondary analysis, with data-driven predictor selection for the less robustly-evidenced predictors (age, relationship status, employment status, multimorbidity, gender, ethnicity, and antidepressant use). This planned secondary analysis was not possible due to the number of systematically missing exploratory predictors across the clusters. A secondary analysis was instead undertaken using univariable analysis. Where univariable analysis found statistically significant associations (after accounting for multiple significance testing), the impact on predictive performance of adding these variables to the model developed in the primary analysis was measured.

Data pre-processing and integrity check

- Data checked and cleaned, including integrity check
- Semantic and technical harmonisation to form final PREDICTR dataset

Model development and assessment of apparent performance

- Multiple imputation by chained equations (x30) for each cluster
- Modelling of continuous predictors using MFP
- Multilevel logistic regression in thirty imputed datasets and estimates combined using Rubin's rules
- Linear predictor calculated for each individual
- Apparent performance (discrimination and calibration) assessed for each cluster
- Average (pooled) predictive performance estimated (across clusters) using random effects meta-analysis

Internal validation

- 100 bootstrap samples created from original dataset
- Model estimated (as for model development) within each bootstrap sample
- Each model then applied in same bootstrap sample (apparent performance) and original dataset (test performance)
- Overall optimism estimated (average difference in apparent and test performance, over 100 comparisons)
- Derivation of optimism-adjusted performance statistics (C-statistic, calibration slope, calibration-in-the-large)

Internal-external cross-validation

- Model developed in 6 of 7 studies, using approach described under "model development"
- Estimated model then applied in the missing ("validation") study
- Process repeated seven times, with each study missing once
- Average (pooled) predictive performance (across validations) estimated using random effects meta-analysis

Shrinkage and estimation of final model

- Uniform shrinkage factor (optimism-adjusted calibration slope) applied to original regression (beta) coefficients
- Shrunken intercept re-estimated with fixed shrunken coefficients to provide final shrunken model

Figure 5.2: Flow diagram outlining primary analyses

145

5.3.10.   Changes from protocol

The full protocol and statistical analysis plan, as pre-registered and published, is presented in Appendix 5.1. Following inspection of the data, some changes to the pre-registered analysis plan were necessary.

1. The protocol included a further (eighth) study [COINCIDE (Coventry et al., 2015)] for inclusion in the PREDICTR dataset. The IPD from COINCIDE could not be used in the analysis as there was data only from baseline, 4 months and 24 months and, as a result, I could not define the outcome of relapse within 6-8 months for this study.

2. The IPD from REEACT did not use the GAD-7 to measure anxiety. Instead, the authors had used the Clinical Interview Schedule-Revised (CIS-R). Rather than discard the IPD from REEACT, I made a decision with my supervisory team to use the anxiety subscale from the CIS-R as a measure of comorbid anxiety. I converted this to standardised scores (z-scores) and used this to model this predictor, along with z-scores for GAD-7 from the other six studies. Z-scores were calculated within each cluster for the whole dataset, prior to removing those who had not reached remission. For the primary analysis, therefore, comorbid anxiety was measured as z-score rather than GAD-7 as planned. To test the validity of this, I conducted a sensitivity analysis removing REEACT from the analysis and using GAD-7 to assess the impact of this decision (Section 6.4.5).

3. Given the number of systematically missing exploratory predictors (see Table 6.1 for distribution of predictor information in sources of IPD), the pre-planned exploratory analysis, using data-driven predictor selection, could not be performed. However, I was able to measure univariable associations between these exploratory predictors and relapse. Where this association was statistically significant, I

measured the effect on predictive performance of including the
predictor in the model (Section 6.5). The Bonferroni method was used
to account for multiple significance testing to provide adjusted p-
values for use as thresholds for significance (Bland and Altman,
1995). This was done to reduce the risk of false positive significant
associations after multiple testing during the exploratory analysis.

4. The planned sensitivity analysis omitting WYLOW and COBRA were
   not deemed necessary as the IECV included a development analysis
   omitting WYLOW and COBRA (and used these as validation sets).

5. The definition of "unemployed" changed from the protocol, prior to any
   analysis. I adapted the categorisation of employed to include those
   unemployed but not seeking work due to ill health (Bartley and Ferrie,
   2001).

6. I planned to calculate sensitivity, specificity and positive and negative
   predictive values for the model at risk thresholds considered
   potentially clinically relevant by the multidisciplinary research team
   and PAG group during model development. I also planned to explore
   the net benefit of the model at particular thresholds using decision
   curve analysis, comparing the net benefit of the model to treat-all and
   treat-none decisions across a range of thresholds (Vickers, Van
   Calster and Steyerberg, 2016). Given the model's suboptimal
   predictive performance, I did not undertake these analyses but I have
   discussed a suggested approach to this in the discussion.

5.3.11.   Ethics approval

The University of York's Health Sciences Research Governance
Committee confirmed that this study was exempt from full ethical approval as
it entails the secondary analysis of anonymised data from studies that had
already received ethical approval (see Appendix 5.4).

Chapter Six

Quantitative Results

6.1. Introduction to chapter

This chapter reports the results of the quantitative prognostic model development and validation study. First, I summarise the included data and report descriptive statistics. I then report the results of the primary analysis: model development (apparent performance), internal validation (and model shrinkage) and internal-external cross-validation (to assess generalisability). Finally, I report the results of the sensitivity and secondary analyses. The results are reported in accordance with the TRIPOD-Cluster statement (Debray et al., 2023).

6.2. Descriptive statistics

6.2.1. Sources of data

Chapter Five outlined the process for identifying relevant sources of IPD for this study. Data were received for all studies identified (six RCTs and one cohort study). Therefore, data from these seven individual studies were included in the final PREDICTR IPD dataset.

6.2.2. Missing data and descriptive statistics

This section presents the descriptive statistics and an overview of the extent to which data for the relevant variables and predictors was available for analysis. There was only a small amount of missing data (Table 6.1). The maximum missing for any of the variables of interest was for number of previous episodes (10.2% overall; 30% were missing for this variable in the WYLOW study).

Table 6.1: Availability of variables and missing data in individual participant data

Variable	Variable present in study							Total number with predictor*	Total number missing (%)
	CADET	CASPER Plus	COBRA	Healthlines Depression	REEACT	REEACT-2	WYLOW		
Previous episodes (% missing)	✓ (None)	✓ (None)	✓ (9.5)	✓ (11.8)	✓ (None)	✓ (None)	✓ (30)	1244	127 (10.2)
Residual symptoms (no missing data)	✓	✓	✓	✓	✓	✓	✓	1244	0
Severity (no missing data)	✓	✓	✓	✓	✓	✓	✓	1244	0
Comorbid anxiety: GAD-7 (% missing)	✓ (1.3)	✓ (1)	✓ (None)	✓ (0.9)		✓ (None)	✓ (None)	1023	4 (0.4)
Comorbid anxiety: CIS-R (% missing)					✓ (0.5)			221	1 (0.5)
Age-continuous (% missing)	✓ (None)	✓ (None)	✓ (None)		✓ (0.5)	✓ (None)	✓ (None)	1134	1 (0.09)
Age-categorical				✓				110	0
Gender	✓	✓	✓	✓	✓	✓	✓	1244	0
Ethnicity (% missing)	✓ (None)	✓ (None)	✓ (None)	✓ (None)	✓ (None)	✓ (0.6)	✓ (4.9)	1244	17 (1.4)
Employment status (% missing)	✓ (0.6)		✓ (None)	✓ (0.9)	✓ (None)	✓ (None)	✓ (None)	1143	2 (0.2)
Relationship status (% missing)	✓ (None)		✓ (None)		✓ (16.3)	✓ (0.6)		707	37 (5.2)
Multimorbidity (% missing)	✓ (None)	✓ (None)	✓ (None)				✓ (10.1)	754	33 (4.4)
Antidepressant use at remission (% missing)	✓ (None)	✓ (17.8)	✓ (None)	✓ (1.8)		✓ (16.4)	✓ (0.9)	1023	49 (4.8)

✓ = variable present in study; *Number of participants in the combined studies with predictor available (including those with missing data)

Table 6.2 presents the descriptive statistics for the IPD. Appendix 6.1 gives more detail around the re-coding and harmonisation of the categorical variables, including the coding in the original sources of IPD. The percentage relapsed was on average 21%, which was in line with the estimate of 20% used in the sample size calculation. WYLOW had a higher outcome frequency than any of the other sources of IPD (32.8%).

Table 6.2: Descriptive statistics for IPD

Variable	Study							
	CADET	CASPER Plus	COBRA	Healthlines Depression	REEACT	REEACT-2	WYLOW	Combined PREDICTR IPD dataset
Total number in published study	581	485	440	609	691	369	439	Not applicable
Total number available	581	485	440	609	691	454	439	Not applicable
Total number at 12-month follow-up	498	358	364	516	484	341	Not applicable	Not applicable
Total number included in analysis (remission)	158	101	169	110	221	159	326	1244
Number relapsed at 6-8 months (%)	32 (20)	28 (27.7)	19 (11.2)	24 (21.8)	34 (15.4)	17 (10.7)	107 (32.8)	261 (21)
Number with one or more previous episodes of depression (%)	112 (70.8)	82 (81.9)	104 (61.5)	89 (80.9)	154 (69.7)	159 (100)	124 (38)	824 (66.2) (1117 with data about number of previous episodes)
Mean PHQ-9 at baseline (SD)	17.35 (4.20)	15.33 (3.39)	17.27 (4.05)	16 (3.70)	15.96 (3.92)	15.77 (3.54)	15.67 (4.17)	16.17 (3.98)

Mean PHQ-9 at remission (SD)	4.70 (2.69)	5.19 (2.50)	4.08 (2.56)	5.97 (2.30)	4.50 (2.61)	4.53 (2.62)	3.52 (2.31)	4.40 (2.60)
Mean GAD-7 (SD)	12.52 (4.83)	8.25 (4.85)	12.41 (4.96)	11.80 (4.44)	GAD-7 not used in REEACT (CIS-R anxiety measure**)	12.97 (4.44)	13.62 (4.63)	12.43 (4.92)
Mean age (SD)	43.2 (12.9)	71.8 (5.16)	45.3 (13.9)	Not applicable (age is categorical*)	40.3 (13.1)	43.3 (14.7)	41.7 (13.8)	45.1 (15.63) (1133 participants with age as continuous variable)
Gender (% Female)	71.5	64.3	59.2	72.7	67	64.2	58.3	64.1 (No missing data)
Ethnicity (% White)	86.7	99	97.6	97.2	97.7	97.5	94.2	95.4 (1227 participants with ethnicity data)
Employment (% Employed)	82.8	Not collected	78.1	95.4	96.4	93.7	65.6	82.5 (1141 participants with employment data)
Relationship status (% In a relationship)	43.7 (Married/living as married)	Not collected	62.7 (Married, cohabiting or civil partnership)	Not collected	75.7 (Married, cohabiting or in a relationship)	58.9 (Married or cohabiting)	Not collected	60.9 (670 participants with relationship data)

Multimorbidity (% with multimorbidity)	55.7	85.1	56.8	Not collected	Not collected	Not collected	29.7	49.5 (721 participants with multimorbidity data)
Current antidepressant use (% on antidepressants at remission)	71.5	41	75.7	90.7	Not collected	41.4	40.2	572.9 (974 participants with data around antidepressants at remission)

*See Appendix 6.1 for detail of Age (categorical) for Healthlines Depression study

**Appendix 6.2 presents a summary of the CIS-R anxiety subscale data for the REEACT study

6.2.3. Risk of bias assessment (PROBAST) of IPD

Risk of bias and concerns around applicability were both low for the included sources of IPD, probably reflecting the purposeful and systematic approach to identifying relevant studies. See Appendix 6.4 for full risk of bias assessment.

Figure 6.1: Risk of bias of IPD (PROBAST)
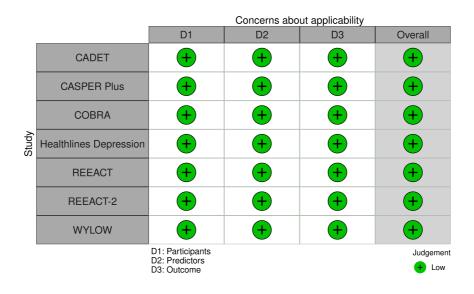
Figure 6.2: Concern regarding applicability of IPD (PROBAST)

In summary, the final PREDICTR IPD dataset included 1,244 participants (18 years old and over), drawn from seven clusters, who had entered remission after a confirmed depression diagnosis (Table 6.3).

Table 6.3: Overview of included data in quantitative study

Population	Adults (18 years old and over) who have entered remission after confirmed depression diagnosis
Data source	IPD from multiple studies
Total sample size	1244
Setting	Primary care/community
Number of clusters	7
Heterogeneity in study design	6 Randomised controlled trials; 1 longitudinal cohort study
Heterogeneity in study populations	Different interventions delivered in RCTs (no experimental intervention in cohort study)
Heterogeneity in data quality	All low risk of bias using PROBAST (Domains 1-3)

6.3.   Primary analyses

6.3.1. Univariable associations

Though not part of the model building strategy, univariable associations were summarised using multilevel logistic regression analyses. Univariable associations were estimated accounting for clustering, incorporating random intercept and random slope (RCT Intervention) terms. Residual symptoms and severity were both associated with relapse in a statistically significant way (an increase in both residual symptoms and

severity on the PHQ-9 led to an increased odds of relapse). Number of previous episodes and comorbid anxiety were not.

Table 6.4: Univariable associations (unadjusted) between predictors and relapse within 6-8 months

Predictor	Unadjusted odds ratio (95% CI)	p-value	Number of participants (N)
Number of previous episodes	1.19 (0.84 to 1.72)	0.319	1117
Residual symptoms	1.13 (1.07 to 1.20)	<0.001	1244
Severity	1.07 (1.04 to 1.11)	<0.001	1244
Comorbid anxiety (z-score)	1.04 (0.90 to 1.20)	0.589	1239
Comorbid anxiety (GAD-7)	1.00 (0.97 to 1.03)	0.943	1019
RCT intervention	0.99 (0.60 to 1.66)	0.981	1244

6.3.2. *Model development and apparent predictive performance*

Missing data for the variables number of previous episodes and comorbid anxiety (see Table 6.1) were handled using MICE. Thirty imputed datasets were generated, using the predictor variables (severity, residual symptoms, RCT interventions) and the outcome variable (relapse). Clustering by study was accounted for during the MICE process by imputing separately for each study.

To check the validity of the multiple imputation, I inspected the values of the 30 imputed datasets for each imputed predictor to ensure plausibility of multiple imputation and consistency in summary statistics across imputed datasets (and with the original, unimputed dataset) (Appendix 6.4).

6.3.2.1.   Modelling of continuous predictors

MFPs were used to model continuous predictors and explore non-linear relationships within the imputed datasets. Fractional polynomial functions were selected within the imputed datasets and prior to fitting the random effects model for computational reasons. The fractional polynomial fitting algorithm converged after one cycle. The transformations of covariates (including correction for mean-centring) used for model analysis are detailed below (Table 6.5).

Table 6.5: Transformations and mean-centring of continuous predictors following MFP modelling

Predictor	Transformation
Residual symptoms	X^-0.5 – 0.4302546311 (X = (residual_symptoms+1))
Severity	Severity – 16.16969453
Comorbid anxiety	Comorbid_anx_zscore + 0.118103204

Table 6.6 presents the results of multivariable multilevel logistic regression analysis.

Table 6.6: Results of multilevel multivariable associations (adjusted) between outcome and predictors (before shrinkage)

Predictor	Regression beta coefficient (95% CI)	p-value
Number of previous episodes	0.13 (-0.24 to 0.50)	0.500
Residual symptoms	-2.11 (-3.07 to -1.14)	<0.001
Severity	0.09 (0.04 to 0.13)	<0.001
Comorbid anxiety	-0.13 (-0.30 to 0.03)	0.936
RCT intervention	0.03 (-0.59 to 0.65)	0.936

Intercept (baseline risk): -1.55 (95% CI: -2.12 to -1.00)

Standard deviation of random effect on intercept: 0.43 (95% CI: 0.14 to 1.38)

Standard deviation of random effect on slope (RCT Intervention): 0.49 (95% CI: 0.11 to 2.16)

Correlation between random effects: -0.23 (-0.93 to 0.84)

6.3.2.2.   Model apparent performance

Predicted risks were calculated for each individual from the developed model (prior to shrinkage), using the average (mean) intercept and regression coefficients (Table 6.6). Figure 6.3 shows the distribution of predicted risks (probabilities) from the model in each cluster. The predicted risks were then used to estimate apparent discrimination and calibration, first by calculating average within-cluster predictive performance statistics, and then by calculating pooled apparent performance statistics (C-statistic, calibration slope and calibration-in-the-large). This process allowed me to understand overall (average) performance across clusters and also within-cluster performance to understand heterogeneity in performance. Prediction intervals were also calculated (Figure 6.4 – 6.6 and Table 6.7).

1) CADET   2) CASPER Plus   3) COBRA   4) Healthlines Depression

5) REEACT   6) REEACT-2   7) WYLOW
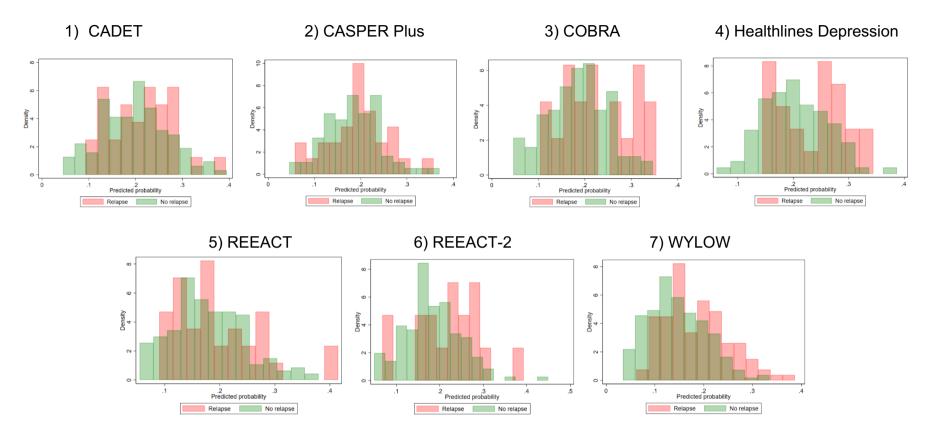
Figure 6.3: Range of predicted probabilities (apparent performance) and observed outcomes in each cluster
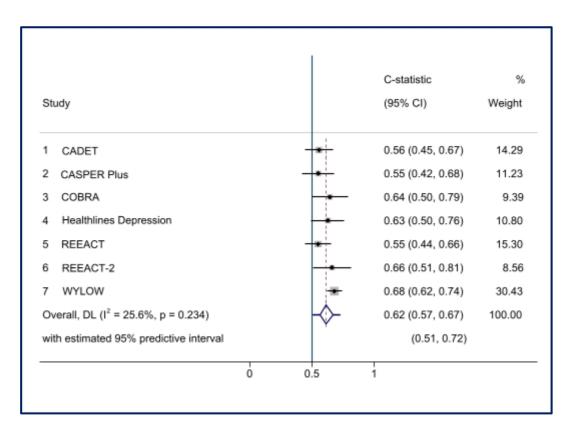
Figure 6.4: Forest plot showing within-cluster and pooled C-statistic (apparent performance)
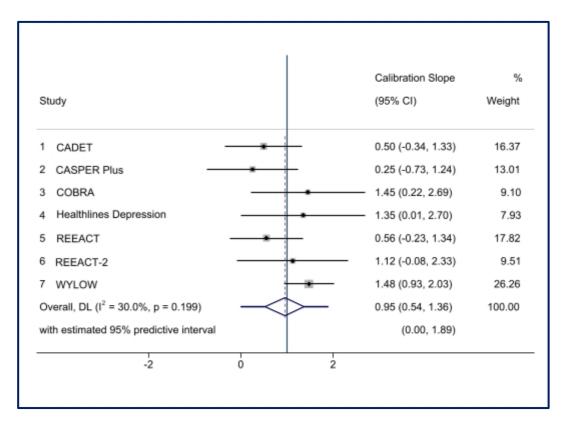
Figure 6.5: Forest plot showing within-cluster and pooled calibration slope (apparent performance)
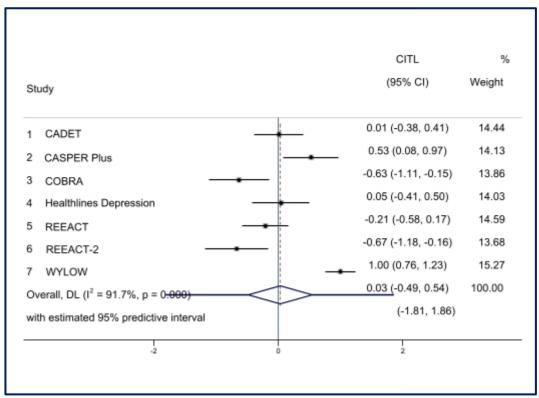
Figure 6.6: Forest plot showing within-cluster and pooled calibration-in-the-large (apparent performance)

Calibration plots with calibration curves were produced, illustrating the apparent performance within each of the seven studies (Figure 6.7).

1) CADET   2) CASPER Plus   3) COBRA   4) Healthlines Depression

5) REEACT   6) REEACT-2   7) WYLOW
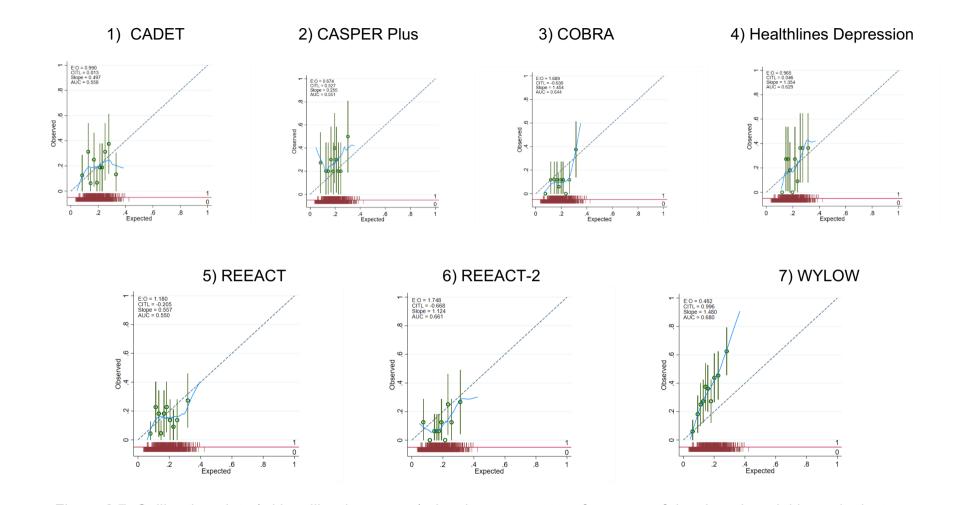
Figure 6.7: Calibration plots (with calibration curves) showing apparent performance of developed model in each cluster

The developed model, prior to shrinkage, had an average apparent performance of: C-statistic 0.62 (95% CI: 0.57 to 0.67), calibration slope of 0.95 (95% CI: 0.54 to 1.36), and calibration-in-the-large of 0.03 (95% CI: -0.49 to 0.54) (Table 6.7). There was heterogeneity in apparent performance across individual studies (clusters).

Table 6.7: Within-cluster and pooled (apparent) predictive performance statistics

Study	N total (N relapsed)	C-statistic (95% CI)	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)
CADET	158 (32)	0.56 (0.45 to 0.67)	0.50 (-0.34 to 1.33)	0.01 (-0.38 to 0.41)
CASPER Plus	101 (28)	0.55 (0.42 to 0.68)	0.25 (-0.73 to 1.24)	0.53 (0.08 to 0.97)
COBRA	169 (19)	0.64 (0.49 to 0.79)	1.45 (0.22 to 2.69)	-0.63 (-1.11 to -0.15)
Healthlines Depression	110 (24)	0.63 (0.50 to 0.76)	1.35 (0.01 to 2.70)	0.05 (-0.41 to 0.50)
REEACT	221 (34)	0.55 (0.45 to 0.66)	0.56 (-0.23 to 1.34)	-0.21 (-0.58 to 0.17)
REEACT-2	159 (17)	0.66 (0.51 to 0.81)	1.12 (-0.08 to 2.33)	-0.67 (-1.18 to -0.16)
WYLOW	326 (107)	0.68 (0.62 to 0.74)	1.48 (0.93 to 2.03)	1.00 (0.76 to 1.23)
Pooled results	1244 (261)	0.62 (0.57 to 0.67)	0.95 (0.54 to 1.36)	0.03 (-0.49 to 0.54)

6.3.3. Internal validation of model and shrinkage

6.3.3.1.   Internal validation of model

Table 6.8 presents the results of the internal validation using bootstrapping (see Section 5.3.9.3 for reminder of methods). Optimism-adjusted performance statistics were: C-statistic 0.60, calibration slope 0.81, and calibration-in-the-large 0.03 (note 95% confidence intervals are not applicable for optimism-adjusted performance statistics).

Table 6.8: Average (pooled) predictive performance – apparent performance and internal validation (primary analysis)

Measure of predictive performance	Average apparent performance (95% CI)	Optimism-adjusted performance
C-statistic	0.62 (0.57 to 0.67)	0.60
Calibration slope	0.95 (0.53 to 1.36)	0.81
Calibration-in-the-large	0.03 (-0.49 to 0.54)	0.03

6.3.3.2.   Shrinkage and final equation

The original beta regression coefficients, estimated from the model in the previous section (Table 6.6), were multiplied by 0.81 (the value of the optimism-adjusted calibration slope). The shrunken intercept was then re-estimated to provide the final shrunken model, as presented in Table 6.9.

Table 6.9: Final shrunken model for predicting risk of relapse in 6-8 months

Intercept and Predictors	Shrunken intercept and regression coefficient
Intercept (baseline risk)	-1.49
Number of previous episodes	0.11
Residual symptoms	-1.71
Severity	0.07
Comorbid anxiety	-0.11
RCT intervention	0.02

To calculate the risk of relapse within 6-8 months, using this model, one would use the following formula: exp(person's risk score) ÷ [1 + exp(person's risk score)]

Where person's risk score (linear predictor) = -1.49 + 0.11 (Number of previous episodes[13]) – 1.71 (Residual symptoms[14]) + 0.07 (Severity[15]) – 0.11 (comorbid anxiety[16]) + 0.02 (RCT Intervention[17])

So, as an illustration, the average (mean) risk of relapse, for a person with no previous episodes of depression and not receiving an effective RCT Intervention (i.e., where these predictor variable = 0) = exp(-1.49) ÷ [1 + exp(-1.49)] = 0.19

---

[13] No previous episodes = 0; One or more previous episodes = 1

[14] $X^{-0.5}$ - 0.43 (where X = (residual_symptoms+1)) – this is the adjustment for non-linear transformation and mean-centring

[15] Severity – 16.17

[16] Comorbid_anx_zscore + 0.118

[17] This would be zero when applied in clinical practice, outside the context of an RCT

6.3.4. *Internal-external cross-validation*

Generalisability of the model was assessed using IECV. The method used for IECV was described in Section 5.3.9.4. As a reminder, the model was developed in six of the seven studies, while the remaining study was left out and used as a validation sample. The developed model was then applied in the left-out (validation) sample to assess external performance. This process was repeated seven times, with each study left out once.

Calibration plots were compared for each validation in each of the different clusters (Figure 6.8). These demonstrate inadequate calibration and heterogeneity across clusters (for example, the calibration plot for REEACT-2 demonstrates over-prediction and the one for WYLOW shows under-prediction of outcome). The distributions of predicted probabilities for each cluster are displayed in Appendix 6.5.

1) CADET

E:O = 1.002
CITL = -0.003
Slope = 0.246
AUC = 0.531

2) CASPER Plus

E:O = 0.619
CITL = 0.638
Slope = 0.130
AUC = 0.522

3) COBRA

E:O = 1.867
CITL = -0.760
Slope = 1.179
AUC = 0.617

4) Healthlines Depression

E:O = 0.970
CITL = 0.040
Slope = 1.172
AUC = 0.619

5) REEACT

E:O = 1.267
CITL = -0.297
Slope = 0.485
AUC = 0.550

6) REEACT-2

E:O = 2.018
CITL = -0.853
Slope = 1.111
AUC = 0.666

7) WYLOW

E:O = 0.424
CITL = 1.134
Slope = 1.703
AUC = 0.666
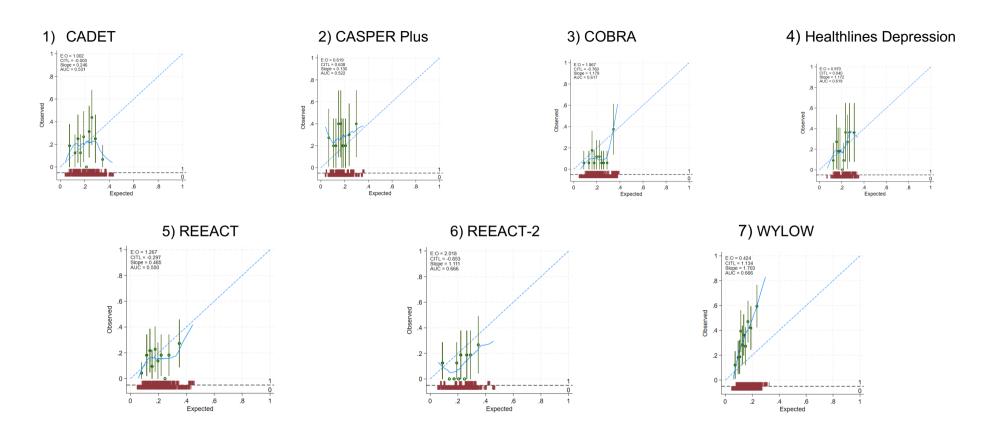
Figure 6.8: Calibration plots in each internal-external cross-validation ("external" validation within each study)

Random effects meta-analyses were performed to summarise the performance statistics from each validation within each round of IECV. The pooled summary performance statistics are presented in Figures 6.9 to 6.11 and Table 6.10.
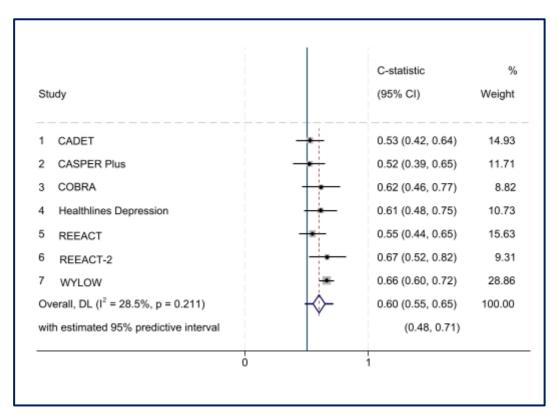
Figure 6.9: Forest plot showing C-statistic for each validation and pooled C-statistic in IECV
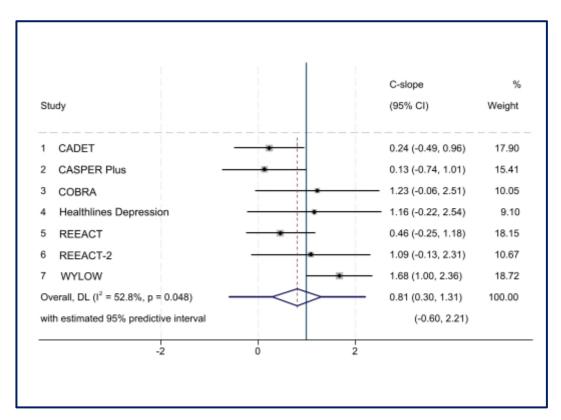
Figure 6.10: Forest plot showing calibration slope for each validation and pooled calibration slope in IECV
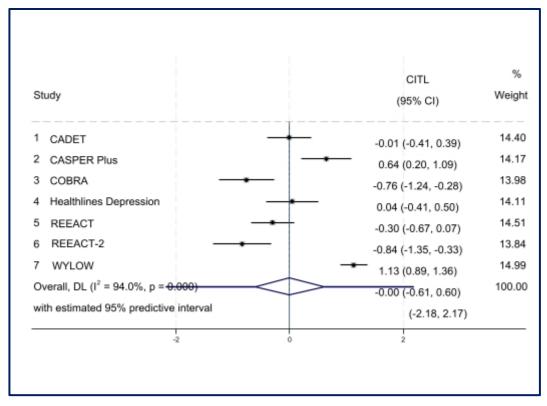
Figure 6.11: Forest plot showing CITL for each validation and pooled CITL in IECV

Table 6.10: Summary of performance statistics in each validation (IECV)

Study	N total (N relapsed)	C-statistic (95% CI)	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)
CADET	158 (32)	0.53 (0.42 to 0.64)	0.24 (-0.49 to 0.96)	-0.01 (-0.41 to 0.39)
CASPER Plus	101 (28)	0.52 (0.39 to 0.65)	0.13 (-0.74 to 1.01)	0.64 (0.20 to 1.09)
COBRA	169 (19)	0.62 (0.47 to 0.77)	1.23 (-0.06 to 2.51)	-0.76 (-1.24 to -0.28)
Healthlines Depression	110 (24)	0.61 (0.48 to 0.75)	1.16 (-0.22 to 2.54)	0.04 (-0.41 to 0.50)
REEACT	221 (34)	0.55 (0.44 to 0.65)	0.46 (-0.25 to 1.18)	-0.30 (-0.67 to 0.07)
REEACT-2	159 (17)	0.67 (0.52 to 0.82)	1.09 (-0.13 to 2.31)	-0.84 (-1.35 to -0.33)
WYLOW	326 (107)	0.66 (0.60 to 0.72)	1.68 (1.01 to 2.36)	1.13 (0.89 to 1.36)
Pooled	1244 (261)	0.60 (0.55 to 0.65)	0.81 (0.31 to 1.31)	0.00 (-0.61 to 0.60)

6.3.5. Sensitivity analysis

A sensitivity analysis was performed with and without REEACT (see Appendix 6.6 for detailed results and analysis). For the analysis without REEACT, comorbid anxiety was modelled as a predictor using the continuous GAD-7 rather than z-scores. In the sensitivity analysis, comorbid anxiety was associated with relapse in a statistically significant way [regression coefficient (95% CI): -0.04 (-0.08 to 0.00), p=0.047)]. This predictor was not statistically significant on univariable regression analysis (Table 6.4).

Despite this, the sensitivity analysis did not significantly change the study conclusions, although it did suggest there was potentially less predictive power from using z-scores across clusters (from GAD-7 and CIS-R anxiety subscale), and that comorbid anxiety measured by GAD-7 alone may be a better predictor of depressive relapse. Calibration and discrimination [Pooled C-statistic (95% CI): 0.65 (0.61 to 0.69); pooled calibration slope (95% CI): 0.98 (0.65 to 1.32); pooled calibration-in-the-large (95% CI): 0.03 (-0.56 to 0.62)] were both marginally improved on this analysis, compared to the primary analysis. Following discussion with my multidisciplinary supervisory team, this is not thought likely to be a clinically significant improvement in predictive performance of the model. However, when considering the clinical usability of the model, it is likely that using the GAD-7 score in practice would be more acceptable and useful than using a z-score.

6.4. Secondary analyses

As outlined in Chapter Five (Section 5.3.9.6), due to the number of systematically missing exploratory predictors across the datasets, the pre-registered exploratory analysis (using data-driven predictor selection procedures) was not possible. I performed univariable multilevel logistic analysis with all exploratory predictors to explore their association with relapse (Table 6.11).

Table 6.11: Univariable associations (unadjusted) between outcome and predictors (secondary analysis)

Predictor		Unadjusted odds ratio (95% CI)	p-value	Number of participants (N) with predictor
Gender (female)		0.87 (0.62 – 1.10)	0.196	1244
Ethnicity (white)		1.59 (0.86 – 2.93)	0.138	1227
Age (continuous)		1.01 (1.00 – 1.02)	0.198	1133
Age - categorical (years old)	<40	Reference category		1243
	40-49	1.16 (0.80-1.68)	0.433	
	50-59	1.31 (0.88-1.95)	0.180	
	60-69	0.85 (0.51-1.43)	0.543	
	70+	1.93 (1.04-3.59)	0.037	
Employment (employed)		0.76 (0.52 – 1.11)	0.161	1141
Relationship status (in a relationship)		0.43 (0.28 – 0.67)	<0.001	670
Multimorbidity		1.31 (0.90 – 1.90)	0.158	721
Current antidepressant medication		0.97 (0.70 – 1.35)	0.853	974

.

Relationship status was noted to be a highly statistically significant predictor on univariable analysis. This was true after adjusting the significance level to account for multiple significance testing using the Bonferroni adjustment (adjusted p-value of 0.005 (0.05 / 10). Univariable association between the variable age (categorical) and relapse demonstrated a small statistically significant association between being 70 years old and over and increased risk of relapse (Table 6.11). This association does not robustly withstand accounting for multiple significance testing using the Bonferroni correction.

To explore relationship status as a relapse predictor further, I repeated the model development procedures used for the primary analysis for the studies with the variable relationship status (CADET, COBRA, REEACT, and REEACT-2). See Appendix 6.7 for full details. Table 6.12 shows the results of the multivariable analysis, demonstrating that relationship status remained associated with relapse in a statistically significant way, after adjusting for other prognostic factors.

Table 6.12: Multivariable analysis in secondary analysis model (including relationship status as predictor)

Predictor	Beta coefficient (95% CI)	P value
Number of previous episodes	-0.15 (-0.66 to 0.37)	0.582
Residual symptoms	0.07 (-0.01 0.15)	0.081
Severity	0.07 (0.01 to 0.13)	0.020
Comorbid anxiety	-0.12 (-0.37 to 0.14)	0.363
Relationship status	-0.79 (-1.23 to -0.34)	0.001
RCT intervention	-0.40 (-0.84 to 0.04)	0.076

To assess the impact of including this variable as a predictor on overall predictive performance, I developed the model in the four applicable studies both with and without relationship status, to provide a like-for-like comparison. Table 6.13 shows the results of these analyses, as well as summarising the predictive performance statistics across the different analyses described in this chapter. There was a small improvement in both discrimination (C-statistic) and calibration (slope) from including relationship status as a predictor compared with the model without this predictor. As for the sensitivity analysis, it is unlikely that this improvement in prognostic model performance is clinically meaningful, but the prognostic value of this association warrants further exploration.

Table 6.13: Summary of predictive performance for primary, sensitivity and secondary analyses

Measure of predictive performance	Primary analysis		Sensitivity analysis (without REEACT)	Secondary analysis exploring relationship status as a predictor	
				Without relationship status	With relationship status
	Development (apparent performance)	Internal validation (optimism-adjusted performance statistics)	Apparent performance	Apparent performance	Apparent performance
Number of participants	1244	1244	1023	707	707
C-statistic (95% CI*)	0.62 (0.57 to 0.67)	0.60	0.65 (0.61 to 0.69)	0.60 (0.54 to 0.66)	0.63 (0.57 to 0.70)
Calibration slope (95% CI*)	0.95 (0.53 to 1.36)	0.81	0.98 (0.65 to 1.32)	0.94 (0.37 to 1.51)	0.96 (0.56 to 1.36)
Calibration-in-the-large (95% CI*)	0.03 (-0.49 to 0.54)	0.03	0.03 (-0.56 to 0.62)	0.01 (-0.25 to 0.27)	0.01 (-0.20 to 0.23)

*95% CI presented, where applicable (not for optimism-adjusted statistics)

6.5. Discussion and summary of quantitative results

Here, I will provide a brief summary and discussion of the results of the quantitative study. A more detailed discussion of the results in the context of the wider literature and mixed methods integration, and also the strengths and limitations of the study, is included in Chapter Ten.

This study has developed a model with suboptimal predictive performance and cannot be recommended for implementation in primary care in its current form on this basis. Residual symptoms and severity of depression were associated with relapse in a statistically significant way. Number of previous episodes and comorbid anxiety were not. The univariable association between relationship status, an exploratory predictor, and relapse was also statistically significant. Inclusion of this predictor within the multivariable model marginally improved the predictive performance of the model, although probably not enough to be clinically meaningful. Older age may be associated with relapse but this association could not be robustly confirmed or refuted in this study. It might be that this warrants further investigation beyond this study, but the lack of granularity in the PREDICTR dataset meant that further exploration within this study was not feasible.

As the IECV in this study demonstrated, the external performance and generalisability of the model was poor. While the heterogeneity across clusters was accounted for through the use of multilevel analysis, it is worth considering where this could have impacted on model generalisability. The studies from which IPD were drawn were generally comparable in terms of severity of depression at baseline and at remission (measured by PHQ-9). The studies were also comparable in terms of baseline demographic variables such as age, gender and ethnicity. The most notable exception to this was CASPER Plus, where all participants were aged 65 years old or over. The COBRA study was notable for being the only study for which relapse was identified at six months (rather than eight months) after baseline.

WYLOW was qualitatively different to the other studies in that it was a cohort study rather than an RCT.

The calibration in three of the validation studies (COBRA, REEACT-2 and WYLOW) was particularly poor. There are several potential explanations for this. The developed model under-predicted when applied in the WYLOW cluster. WYLOW had a higher outcome prevalence (thirty per cent) than the average; as a consequence, the model was calibrated to expect a lower risk of relapse (based on the average baseline risk, which is around twenty per cent). Going forwards, local recalibration could be explored for this study, by adjusting the intercept to account for the increased outcome prevalence. Similarly, in COBRA and REEACT-2 (where the model over-predicted relapse), the outcome prevalence was lower than average. A similar approach to local recalibration could be attempted in these studies, but by adjusting the intercept to account for a lower, rather than higher, outcome prevalence. If a model were developed with promising predictive performance following internal validation, then IECV would ideally be undertaken using the shrunken model, developed following bootstrapping. In this study, given the suboptimal predictive performance, it sufficed to validate the un-shrunken model with the intention of exploring generalisability and heterogeneity in performance (per the protocol).

Chapter Ten includes a more detailed discussion of the findings from this study in the context of the literature and as part of the overall mixed methodology used in this thesis. In that chapter, I also discuss further possible strategies for local recalibration, assessment of clinical usefulness and other suggestions for further research.

Chapter Seven

Qualitative Methods

The methodological approach of the qualitative research, and the rationale for adopting a mixed methods approach throughout this thesis, was discussed in Chapter Four. This chapter outlines the methods used for the qualitative study.

7.1. Aims of study

The aim of the qualitative study was to better understand the perspectives of people with lived experience of depression and GPs around relapse of depression and the role of risk prediction and relapse prevention interventions.

The objectives were:

- To understand the experiences of people with lived experience of depression who have sought treatment in primary care;
- To understand GPs' experiences and current practices when diagnosing, treating and following up patients with depression, including relapse prevention;
- To explore peoples' understanding of relapse, perceptions of relapse risk (and communication of risk more generally) and the extent to which people with lived experience of depression wish to discuss relapse in primary care;
- To understand the GP perspective on depressive relapse and how relapse risk is currently assessed and managed in practice;

- To understand GPs' current practice with respect to risk prediction models and to explore the ways in which they are currently used and communicated to patients;
- To explore the potential implementation of a prognostic model to predict relapse risk in primary care from the perspectives of people with lived experience of depression and GPs, and what the enablers and barriers to this might be;
- To assess the acceptability and future feasibility of primary care-based relapse prevention interventions and to ascertain patient and GP preferences around these.

7.2. Methods

7.2.1. Ethics Approval

An Integrated Research Application System (IRAS; project ID 292780) application was approved by the University of York's Department of Health Sciences and submitted to the NHS's Health Research Authority (HRA) in December 2021. I was invited to attend a Research Ethics Committee (REC) meeting led by the Black Country REC, via video conferencing, on 31st January 2022. Final ethical approval was granted on 11th March 2022 (REC reference: 22/WM/0022; Appendix 7.1).

7.2.2. Setting and participants

7.2.2.1. People with lived experience of depression

Participants in this group were adults (aged 18 years and over) with lived experience of depression (self-report or clinical diagnosis coded in their primary care health record). Participants had to have capacity to provide informed consent; this was assessed while obtaining informed consent on the basis those participants could understand and relay their understanding of the study to me as the researcher. Participants that had had or were

having secondary care involvement for their depression were asked to focus mainly on their experiences with their GP in primary care during the interviews. Exclusion criteria were: history of severe mental illness (e.g., bipolar affective disorder, schizophrenia).

7.2.2.2. General practitioners

Participants in this group were practising GPs, working within the NHS in any capacity (partner, salaried, or locum GP).

7.2.3. Recruitment strategy

Recruitment was supported by the NIHR Clinical Research Network (CRN). The study opened to recruitment in April 2022, with study information and initial requests for expressions of interest sent to all research active primary care sites across Yorkshire and the Humber. Practices were then selected to include a range of patient demographics (in particular, practices were selected to achieve maximum variation in participants' ethnicity and socioeconomic status). 15 general practices across Yorkshire and the Humber signed a Participant Identification Centre (PIC) agreement to support the study.

7.2.3.1. People with lived experience of depression

People with lived experience of depression were recruited in two ways: database search of GP electronic health record and study advertisement through participating general practices (Appendix 7.2). These two recruitment methods were used to optimise the chances of recruiting a sufficient number of eligible participants. General practices were asked to conduct a database search and eligibility check to identify and invite a minimum of 50 patients per practice with a history of depression and without any of the exclusion criteria. Following a database search to identify eligible people, a letter of invitation (Appendix 7.4) was sent by the practice (on

practice-headed notepaper) to prospective participants. Potential participants were asked to contact the research team directly (by emailing me) to express interest in taking part in an interview. At this point, a comprehensive study information sheet was shared (Appendix 7.5) and they were given the opportunity to ask questions or clarify any issues with me.

Each participating site also shared a study advertisement poster (Appendix 7.2) in their waiting rooms and on their social media platforms, as they felt appropriate. For participants recruited via advertisement (and not via a review of their clinical record), there were additional considerations. I undertook an additional eligibility check in such cases at the beginning of the interview and, to ensure that I could respond to risk in the same way as for participants recruited via clinical record assessment, these participants were asked in advance to provide details of their general practice and give permission for us to contact them in the event that it is felt necessary per the risk protocol. If potential participants were unwilling to grant this permission prior to the interview, they would not have been invited to participate in an interview.

7.2.3.2. General practitioners

Each participating GP surgery (PIC) was asked to share study information with all of their GPs, and interested GPs were asked to contact the research team directly. The REC-approved social media advertisement (Appendix 7.3) for GP recruitment was shared on Twitter in April 2022. GPs identified through a snowballing method were asked to approach me directly or were approached using a participant invitation letter (Appendix 7.6). A detailed participant information sheet (Appendix 7.7) was provided and written informed consent taken (Appendix 7.9) prior to any research activities.

7.2.4. Expenses/reimbursement

All participants were reimbursed for their time in the form of a voucher, in line with NIHR INVOLVE guidance (NIHR, 2021) (for people with lived experience of depression) and British Medical Association (BMA) guidance for GPs. All reimbursements to participants were made in the form of vouchers. Participating GP practices were reimbursed study support costs (database search, eligibility check and GP recruitment) and research costs (mail-out) in line with the "Attributing the costs of health and social care research and development (AcoRD)" guidance (Department of Health, 2012).

7.2.5. Patient and public involvement and engagement

The PAG contributed to the design of participant-facing materials and in refining the language of these materials and the topic guides. The advertisement posters, invitation letters and information sheets underwent a series of iterations over two online workshops with the PAG, to make sure that the wording and visuals were acceptable and clear. I also undertook pilot interviews with two members of the PAG separately and the topic guide was modified with their input, to ensure that the questions asked were clear, sensitive and logically sequenced. The PAG also contributed to reviewing the analysis; initial analysis was shared with the PAG at a workshop in September 2022 and initial impressions of the analysis sought from the group. A further workshop was held in April 2023 where themes were finalised and plans for lay dissemination discussed. A summary of PPIE throughout the whole mixed methods study is presented in Figure 7.1 (see Appendix 7.10 for more detail).

Pre-PhD September 2017
- Overview of study proposal and group discussion
- Idea-sharing and prioritising research objectives
- PAG comment on draft NIHR application

October 2019
- Review and development of systematic review and quantitative protocols
- Preliminary discussions around qualitative workstream
- Discussion of training opportunities for PAG members

October 2020
- Review of Plain Language Summary for Cochrane Review by all PAG members
- Review and finalisation of prognostic model protocol
- Further preliminary planning for qualitative workstream

June 2021
- Workshop for co-production of participant-facing qualitative materials prior to HRA ethics submission
- Initial feedback from PAG on semi-structured topic guide

July 2021
- 2 x Pilot interviews with PAG members
- Feedback and revision of topic guide

September 2022
- Presentation of initial data analysis
- Discussion of preliminary themes and framework

April 2023
- Finalisation of thematic analysis
- Agreement on final themes
- Discussion of next steps including lay dissemination of findings

Figure 7.1: Overview of PPIE activities throughout study

7.2.6. Data generation

I interviewed GPs and people with lived experience concurrently, from May-October 2022. I was supported in this when needed by my supervisor (also a practising GP, professor of general practice research and experienced qualitative researcher). Interviews were guided by semi-structured topic guides (Appendices 7.11 and 7.12). An initial version of the topic guide was informed by the pre-existing literature, developed in partnership between myself, my supervisors and the PAG and was approved by the REC. The topic guide was then developed and modified iteratively as data were collected and analysed. Areas covered in the topic guides were, for people with lived experience: depression history and understanding of relapse; understanding of relapse prevention and ideas/preferences around treatment of depression, relapse prevention interventions; perspectives on the prognostic model and its face validity and acceptability (Sekhon, Cartwright and Francis, 2017); and ideas around other important potential predictors not included in the model. Interviews with GPs were more focussed and specifically assessed understanding of depression remission and relapse, the acceptability and feasibility of the prognostic model (and use of prediction models generally in general practice), and the range of primary care-based relapse prevention interventions.

At the start of the interviews, demographic information (such as age, ethnicity, gender, household composition, occupation, educational attainment, comorbidities) from people with lived experience depression were collected to contextualise the data and support the description of the sample in publications. For GPs, information collected included their age, gender, job roles (GP partner, salaried or other; whether they have a special interest; and information about other relevant clinical or managerial roles) and years of experience working in general practice.

7.2.7. Data management

The General Data Protection Regulation (GDPR) was adhered to and data were stored in line with the University of York's data management policies. Interviews were conducted using a secure video conferencing platform (Microsoft Teams), except for two interviews (one person with lived experience and one GP) which were conducted by telephone and recorded through Teams (participants dialled into Teams using their mobile telephones). All interviews were digitally recorded using Microsoft Teams, fully transcribed and anonymised, and securely stored at all stages on the password-protected, secured shared drive hosted by the Department of Health Sciences, University of York. The transcription was done by an external, commercial professional audio-transcription company in the naturalized "intelligent" verbatim, which has the advantage of presenting the data using written norms and is considered an acceptable method of transcription for health sciences research (McMullin, 2021). Data were analysed in NVivo and only anonymised transcripts were shared between supervisors at the University of York and Keele University. I am the named long-term custodian of the data, which will be stored securely and preserved for ten years.

7.2.8. Ethical considerations

Prior to data collection, I obtained written informed consent, using a consent form (Appendices 7.8 and 7.9). Participants were given a minimum of 48 hours with the participant information leaflets, and were given the opportunity to ask questions before signing the consent forms. I had completed Good Clinical Practice (GCP) training and am trained in the assessment of mental capacity and management of risk through my clinical role. Participation in the study was kept confidential and data were anonymised through the use of a unique study code.

I was mindful that interviews may have contained sensitive information about experiences of distress. If, during interviews, participants became upset or distressed they were offered the opportunity to debrief at the end of the interview or to stop the interview altogether at any point. At the end of every interview, I checked on the participant's wellbeing and explain the next steps. I ensured a robust risk protocol (Appendix 7.14) was in place for if the participants disclosed suicidal ideation or intent. This was intended to guide decision-making should the safety of participants or others be identified as at risk. As a qualified GP, I am experienced with assessing and managing risk. If a second opinion or additional support was required, I planned to discuss the case with my supervisor (although in the event this was not necessary).

Reimbursement of participants for time is an ethical consideration and the rate of compensation was set after deliberation with my supervisory team. The INVOLVE rate was used for compensation of people with lived experience as this was felt to be fair reimbursement of their time without being an inducement to participate. The BMA rate was used for GPs in recognition of the time pressures on primary care and the need to either adequately backfill practice time spent contributing to the study, or reimburse GPs' time at a rate that was commensurate with the usual value of their time.
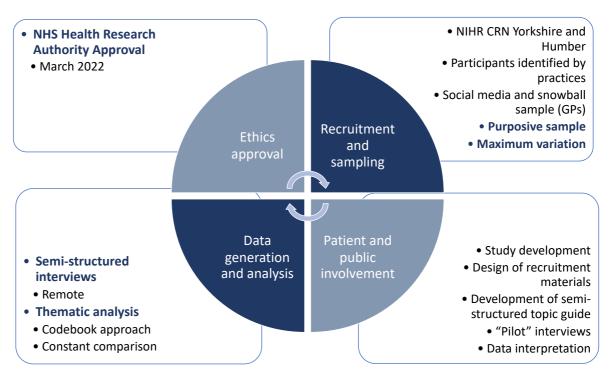
Figure 7.2: Summary of qualitative methods

7.2.9. Data analysis

Below, I outline the approach I took to the six phases of TA, as set out by Braun and Clarke:

7.2.9.1. Data familiarization

Familiarization began at an early stage as I had conducted all of the interviews myself. After each interview, I documented my thoughts and reflections in a reflexive journal, to refer back to later. Once the data had been transcribed, I initially listened back to every recording with the transcript in front of me, both to check the transcripts for accuracy and to note initial points of interest. Once I was happy that the transcripts were an adequate representation of the interviews, I read and re-read the transcripts critically several times. During this process, I made a note of possible codes and also of my reactions to the data, to add to and compare with my initial reflections

in the reflexive journal.

7.2.9.2. Generating codes

I then began to code the data inductively, initially using the "Notes" function in Microsoft Word and then using NVivo 12, a computer-assisted qualitative data analysis system. The codes were intended to be both semantic (capturing surface meanings) or latent (capture underlying ideas and implicit meaning) as appropriate, and aimed to capture what was analytically interesting about the data. I undertook at least two cycles of this phase for all data items, ending with a list of codes and the data relevant to each code.

7.2.9.3. Generating initial themes

I generated themes by reviewing the codes and identifying patterns of shared meaning, which were united by an underlying core concept [or "central organizing concept" (Braun and Clarke, 2019a)]. I used paper and pen and computer-generated mind maps (see Appendix 7.18 for an example) to consolidate codes and help develop the themes.

7.2.9.4. Reviewing initial themes

I then reviewed the initial themes with reference to the codes and also to the overall dataset. The main question at this phase was whether the initial themes captured the most important features of the dataset and whether together they helped to tell a story that was consistent with the data collected. This part was undertaken in collaboration with my multidisciplinary supervisory team and the PAG. I held a PPI workshop in September 2022 to share the preliminary data and results of initial analysis and the PAG inputted into shaping these and guiding the focus of the ongoing analysis.

7.2.9.5. Defining and naming themes

I assigned names to the themes and began by writing a short description that summarized the meaning of the theme. At this stage, I began to consider how the themes contributed to the overall story and how they in turn related back to my research aims and objectives. Sub-themes were assigned where these captured a particular important facet of a theme. I held a further PPI workshop in April 2023 and shared a further iteration of the thematic framework with the PAG. This workshop was used to further refine the themes and to arrive at final theme names.

7.2.9.6. Producing the report

Chapters Eight and Nine report the analytic commentary and the final generated themes, along with illustrative extracts from across all of the data items. Chapter Ten discusses how these themes relate back to the research questions and also how the findings relate to pre-existing research literature and the broader context of national health policy.
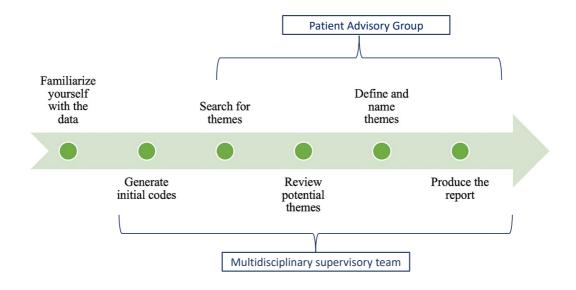
Figure 7.3: Stages of thematic analysis in this study[18]

---

[18] Adapted from the six stages of thematic analysis (Braun and Clarke, 2006)

7.3. Statement on reflexivity

Qualitative research, particularly that which adopts a Big Q approach, is recognised as a subjective pursuit, where the researcher brings a perspective that shapes the data collection and analytic processes (Braun and Clarke, 2021b). This is better thought of as an important resource for research rather than a problem to be controlled, although it is important for the researcher to reflect critically on the way that the subjectivity has affected the research (Gough and Madill, 2012). This process of critically reflecting on the process and content of knowledge produced through research is known as reflexivity, or a situated account of the analysis by the researcher (Trainor and Bundon, 2021). This is particularly important when adopting a contextualist epistemology, as the process of reflexivity is a way of being transparent about the contexts relevant to data analysis and knowledge production. I took several approaches to reflexivity throughout: I kept a reflective journal which I filled in after each interview; I had regular reflective meetings and discussions with my supervisors and I have reported a reflexive account here, prior to presenting the findings from the qualitative work.

Reflexivity can be functional (the way in which the research tools and approach shape the research) or personal (the way in which the researcher affects the research) (Wilkinson, 1988). I have incorporated reflective accounts of both functional and personal reflexivity in this research. In terms of functional reflexivity, all interviews were done remotely due to restrictions imposed by COVID-19 pandemic. This was enabled by the increased familiarity and computer literacy of people as a result of the pandemic, in particular a willingness to use Microsoft Teams. This opened up a much wider participant pool and worked well from a flexibility point of view. However, Braun and Clarke (2013) discuss drawbacks of remote interviews, which can sometimes mean that non-verbal communication is less well conveyed and can affect data generation and interpretation. The use of virtual interviews may have limited the amount of information people were willing to share. There were only technical problems with two of the

interviews and these were minor (slight screen freezes or parts where the audio froze or cut out). There were also occasional interruptions (doorbell ringing, participant received a phone call) and some of the GPs were in their consulting rooms which created the impression that they had limited time or even that they could be interrupted for a clinical reason, which changed the dynamic very slightly. I do not think these affected the data collection and caused only minor disruptions overall. The selection of semi-structured interviews as my data collection tool meant that the conversation progressed according to my agenda rather than the participants, although there was plenty of time for me to allow participants to talk and explore tangential avenues of conversation. There was only one interview where I found it very difficult to keep on track due to the patient being extremely talkative and wishing to discuss her life more generally and secondary care presentations more than the subjects I was intending to explore. This interview certainly yielded useful (and, actually, fascinatingly contradictory data) but was the only interview where I felt I had to explicitly ask if we could focus more on the topics of my choosing. The participant was very understanding about this and I do not think this affected the rapport we had developed over the hour.

In terms of personal reflexivity, there are certain values and characteristics that I possess that situate me as a researcher. First, from a socio-demographic perspective, I am male, white and, of particular relevance, I am a GP and therefore viewed the data from the perspective of one group of participants (other GPs, with whom I was an "insider" researcher), but not the other (people with lived experience of depression, where I was more of an "outsider"). Furthermore, I brought a research background to this qualitative project that was more quantitative in nature – the majority of my previous research projects have been carried out within a scientific, positivist (or at least post-positivist) paradigm. My natural inclination is to think in terms such as bias, generalisability and reliability and orienting myself to a qualitative methodology was not an immediate process but one that took some introspection, discussion with my supervisors and lots of reflection and reading. Finally, many of my mannerisms and approach to interviewing were inevitably influenced by my clinical training, and particularly

the training I have had in taking a medical history. A commonly used framework in medical training, and one that I was trained using, is the Calgary-Cambridge framework, where open questions are used initially and become gradually more focussed as the consultation progresses. I undertook qualitative research training as part of my Fellowship to ensure that I was not leaning too heavily on my clinical interview training and was adopting a suitable approach for qualitative research.

I also hold particular political views with respect to the governing of the country generally but, more relevantly, regarding the funding and resourcing of primary care. My background as a GP Partner and a member of my Local Medical Committee (LMC, the local statutory GP representative body) mean I am regularly engaged in political discussions and aware of medico-political advancements and changes. In particular, I am aware of how pressured and under-resourced primary care is at the current time. Some of the GPs I interviewed shared similar views as me, and my knowledge in these areas led to interesting conversations about primary care mental health provision and the implementation of relapse prevention in primary care. However, my view as a GP was not compatible with some of the expectations expressed by people with lived experience and I had to be more vigilant of my biases where discussions around GP resource were raised; my purpose in these interviews was to learn about patients' experiences and views and not to challenge these unduly.

I felt it was important to explain to all participants at the outset that I am a GP. I tried to remain cognizant of the ways in which this might have affected people with lived experience of depression. For example, I am aware of the potential power dynamics that could be perceived and felt by people with lived experience. There is also a reverse power dynamic in that people with depression may view me as someone who represents a profession who is supposed to serve and help them and that there may be some criticisms of the care that they have received from their own GP. My clinical background also meant that there was a risk that my interpretation was overly clinical and less related to the meaning in the data.

Discussing my clinical role as a GP was particularly interesting in the case of the GP interviews. Some of the answers given by the participants gave the impression that they felt as though they were being tested: for example: "I don't because I…I might be wrong here, you can instruct me afterwards, but…" [GP9]. This fits with previous evidence on the subject (Chew-Graham, May and Perry, 2002), who wrote about the experiences of GP researchers interviewing GPs. In keeping with this previous literature on the subject, my sense was that being a GP interviewing GPs had some benefits in terms of building rapport and generating rich data. Chew-Graham described the role of a GP researcher interviewing GPs as that of a "professional peer and private confidante" (Chew-Graham, May and Perry, 2002). There was a sense in which some of the conversations could begin at a point of exploring the important questions rather than too much explanatory dialogue or context-setting (on either part). There was also a feeling that the interview participants understood that I had some pre-existing insight into some of the clinical issues and, in some case, I think the interviewees opened up more than they might have done if I had been a non-clinical researcher.

Where this had perhaps a less positive effect was in cases, particularly in earlier interviews, where I realised on listening back to the interviews or reviewing the transcripts with my supervisor that I had colluded with the participant rather than remaining objective and exploring topics in more detail. In the later interviews, I made an effort to remind myself to step back from this and not to take the participants' statements for granted. I asked them to clarify where there was any ambiguity, even when their implication and expectation appeared to be that I would know what they meant. As Chew-Graham (2002) reflects, the assumption of a shared understanding between the interviewer and interviewee in qualitative research is problematic. The aim of this study has been to understand the ways that GPs manage patients and try to make sense of clinical decision making. It is difficult to explore the more nuanced and less explicit factors influencing this if much of it is assumed to be obvious. There were also some instances where the GPs being interviewed asked me questions, as though I

were an expert in the clinical area. There were others where GPs may have felt judged, or as though they were revealing something about their practice that they felt I would not approved of: for example, by prefacing answers with "I'll be honest…"

The qualitative work has really enabled me to contextualise the quantitative aspects of the project and helped me understand what they mean and what the implications are moving forwards (Creswell et al., 2011). Because my previous research exposure has largely been quantitative, it has been a really interesting experience and learning point for me to come to understand that the interviewer can be actively involved in generating data and constructing meaning from the interviews rather than passively "collecting" it. Reassuring interviewees about confidentiality and adopting a non-judgemental approach was essential in ensuring that the data generated were as honest as possible, although the data must still be interpreted on the basis that some interviewees may have been holding some things back or even, potentially, fabricating answers or modifying their responses to give what they felt was the "right answer".

I found it difficult in the first few interviewers to remain dispassionate and was worried about leading the interview participant or sharing my own thoughts. I found as the interviews went on that I became more comfortable and confident in sharing my own views, where appropriate, and sometimes engaging in a peer-to-peer conversation rather than an interview. I found that this helped to build rapport and engage the participant more. While I found it challenging when lay participants criticised or expressed negative views of GPs, I think it was important to recognise and accept that some patients have had less than satisfactory experiences with GPs and recognise that the aim of the research is hopefully to improve this.

Chapter Eight

Qualitative Findings Part 1: Perspectives on relapse risk and prevention in general practice

8.1. Overview of chapter

This chapter presents the findings from the thematic analysis of the qualitative data pertaining to relapse and recurrence of depression in primary care. The findings relating to the use of prognostic models and communication of risk in primary care are presented in Chapter Nine.

8.2. Participants

I interviewed twenty-three people with lived experience of depression and twenty-two GPs (see Figure 8.1 for a summary and Appendix 8.1 for more detail of individual participants for contextualisation). I achieved the aim of ensuring maximum variation in GPs. People with lived experience were an even mix of male and female, ages ranged from 24-75 years old and covered the range of Index of Multiple Deprivation (IMD; participants' home postcode). The average length of interview for people with lived experience of depression was 55.9 minutes (range 34-80 minutes) and for GPs 53.3 minutes (range 39-76 minutes).

People with lived experience of depression (n=23)

- 11 Female, 12 Male
- Age Range: 24-75 years old
- Socioeconomic status (Least deprived to most deprived (IMD))
- Majority White British (n=21)
- Mean interview length: 55.9 minutes (range 34-80 minutes)

General Practitioners (n=22)

- 12 Female, 10 Male
- Range of job roles (11 Partner, 10 Salaried, 1 Locum)
- Range of years of experience as fully qualified GP: 1-30 years
- 2-11 clinical GP sessions per week
- Mean interview length: 53.3 minutes (range 39-76 minutes)

Figure 8.1: Summary of included participants in qualitative study

8.3. Thematic framework

I generated three themes, outlined in Figure 7.2 and discussed in detail, with the aid of illustrative examples from the data, in the rest of this chapter. Where illustrative examples from the data are presented, the participants are detailed in the form: Type of participant (P for person with lived experience or GP for general practitioner; gender (M or F); and age)[19]. Detailed demographics and participant information are available in Appendix 8.1 for additional context, although I have limited the amount of detail presented to ensure the anonymity of participants is protected.

---

[19] For example, P10-M-29 is a 29-year-old, male person with lived experience of depression.

Theme 1: Perceived determinants of depression course

- Social and environmental (external) factors
- Personal (internal) factors
- When is it depression?

Theme 2: Relapse risk and prevention

- Relapse is important – but limited discussion in general practice
- Relapse, remission and recovery: interpretation and conceptualisation

Theme 3: Relationships and communication

- Importance of active listening and empathy
- Importance of continuity of care
- Ongoing care and support
- The patient at the centre of care within general practice
- GP consultations are a limited resource

Figure 8.2: Overview of thematic framework

8.3.1. Theme 1: Perceived determinants of depression course

This theme captures the thread running through the majority of interviews about what the main determinants and predictors of relapse and/or poorer longer-term outcomes of depression are. There was a strong recurring theme throughout the interviews that life stress, adverse life events and childhood adversity are associated with depression and relapse of depression. The PAG thought it was important that, within this theme, it was made clear that there is a distinction between "external" factors (described here as social and environmental factors) that are outside the person's control, and "internal" factors (described here as personal factors) that describe personality types and coping styles that are more innate.

8.3.1.1. Social and environmental (external) factors

This sub-theme described the *extrinsic* or external factors that affect depression course. These included: work and financial issues; relationship and family issues; and childhood adversity, trauma and educational under-performance. Increased social-connectedness and support were viewed as protective. GPs described the perceived effects of social determinants of health and the relationship between these and depression course.

> *I think the social determinants of health are really important. So, things like poverty, lack of work, financial struggles, that's a big aspect. And then social networks, you know. If people tend to be, say, in a stable relationship, children, family, community…you know, well ingrained within the community or a faith network or something, you know, they tend to do better.*
>
> GP6-M-33

People with lived experience of depression had a similar view on this, citing work and family factors as being linked with longer-term poor depression course.

> *I: What do you think makes somebody more likely to have longer-term depression or relapses in the future?*
>
> *R: Life, generally I'd say…You know, worry, lousy workplace, I suppose, that working long hours, you know, if, for instance at the moment until things are sorted, we do 12 hour shifts seven days a week. And also, if you've got problems in, you know, at home, you know, people who...like I've got my dad, or you've got nightmare kids…*
>
> P7-M-48

Childhood adversity was raised by many people with lived experience of depression as an underlying contributor to an unfavourable depression course. The following data extract describes the extent to which such adverse experiences can pervade one's adult life and have an impact on mental health on an ongoing basis.

I had a lot of like father problems, my parents divorced when I was very young and my dad's an alcoholic and I didn't really have a father figure. So, growing up and at that age, I guess, that really played into my anxiety and my depression which sent me...it really like went, sort of, downhill from around then, I would say about 14 years old or so.

P10-M-29

Some people with lived experienced of depression explained how they had only more recently realised the impact that adverse childhood events had had on their longer-term mental health.

I think there is an element of history, like I say, I've had some childhood events that occurred, the more I go into my mental health journey the more I realise actually how much they impacted me.

P15-F-37

People with lived experience of depression perceived that other relapse risk factors were particularly salient against a background of an increased propensity towards depression as a result of adverse childhood experiences and drew a parallel with physical health.

Yeah, of course there are individual differences, but I think that a lot of those are created in childhood, and if you have a tumultuous traumatic childhood then your likelihood of feeling the pressures, as well as like a whole host of physiological poor health outcomes…neglect has been linked to diabetes, cancer, dementia, so it's not just mental health, you know, it's…like we're this whole integrated package of people.

P21-F-37

Social support was mentioned by several people with lived experience of depression and GPs as an important protective factor against relapse. Participants explained the importance of GPs capturing this in a meaningful way in practice.

> *I think probably that is one of the questions a GP should ask, do you live alone, have you got…and obviously the first time around I'd got my husband who was a great support with the postnatal depression and, again, family. And, yeah…that's where GPs need to know a background of that particular person and to know if they are alone, living alone, if they've got family with them, if they've got family nearby, if they've got a couple of friends, that kind of thing. That is important to find out about the individual people.*
>
> P13-F-57

The perceived importance of exploring patients' environmental and social contexts in primary care consultations for depression was shared by GPs. GPs commented on the deficiencies in routine primary care records and capturing of data generally, and how we might rethink this going forward.

> *GP records don't capture data traditionally on number of people in household, how many children, what age were you when you had your first child, is there a poor social background to dad or mum that is not on the record…what is your financial situation like. Those bits of data that I think actually make a big difference to somebody's mood, we don't capture that in medical records. So, I think I just wonder whether those pieces of the puzzle, they're important but they're not there in medical records.*

> *I guess it's whether you see depression as a purely medical or potentially partially social construct as to whether all those bits of information are important…I mean I don't know if I feel comfortable asking a patient, so what is your annual household income to put on their medical records…But actually, if you've got*

somebody that has got four children at home and their annual
household income is under thirty thousand pounds, especially
given what's going to be happening over the next few months in
terms of energy prices and stuff, they're going to be feeling the
pinch and that's enough to make anybody feel anxious and low.

GP21-F-33

An important aspect of this sub-theme was the consistent message from both groups of participants who discussed the difficulty of capturing information around these personal, social and environmental factors in a way that is meaningful and useful to GPs and people with depression. Many people with lived experience of depression and GPs reflected on this and explained their scepticism about the way in which this information is currently measured in practice. There was a feeling from both groups of participants that one must be careful when interpreting some measures of social and environmental factors. The following extract describes the difficulty of using, for example, employment and relationship status as measures of presumed protective factors for relapse. The assumption would be that these two factors are protective against relapse, but actually they may prove to be a stressor or trigger for some people with depression.

It's tricky 'cause employment, like, doesn't cut it now, does it, like,
with everyone working from home and stuff. There could be some
people that just never leave the house, and get everything delivered,
and you're working from home as well. That would be a big flag for me
actually, that where are they getting their human connections and,
yeah that, kind of, who's watching out for them? So…and yeah, I
suppose relationship status may be a good or a bad thing, might it?

GP1-F-36

This view was supported by people with lived experience of depression. As an illustrative example, P8 reported finding that being married could be a strain and he perceived it as a driver of his own depression:

I would be surprised if that's…because, you know, I was married for a long, long time and, you know, if being married was one of the things that was a positive that said, yes, this person has got support, actually, in truth, it was probably a drain on me rather than a support to me.

P8-M-58

Finally, participants thought that the significance of environmental and social drivers of depression had become more important as a consequence of the COVID-19 pandemic. One GP described the effect the pandemic and its associated restrictions had on his own mental health:

It's been terrible, hasn't it? It's a really interesting one to be honest. You know, on a personal perspective…I think of myself as a strong character and, you know, I can go through a lot and I've had, you know, quite significant challenges in my life that I'd like to think I've been able to cope with…But actually I really struggled in COVID. I really struggled with anxiety around certain things. And that was me as… I'd probably call myself now, you know, middle-class, settled, with wife and children and, you know, a lot of benefits in my life. So, actually for people that don't have that, you know, it was really, really challenging. And, you know, personally I think…the people that suffered most, in my practice, were people with mental health problems...The social isolation was a challenge…So what about people that lived alone? You know, I think they really struggled.

GP6-M-33

People with lived experience of depression agreed with this view and thought that the COVID-19 restrictions had exacerbated depression for many people, highlighting again the importance of social and environmental factors in driving depression.

I know the pandemic was mega serious but so are a lot of other illnesses and I couldn't believe that the amount of publicity mental

health was getting before the pandemic and then we got the pandemic and everybody was told to stay in the house, not see anybody. Hello, mental illness [laughs]…I was with my daughter so I was alright and we'd got FaceTime and my friends were phoning me regularly and I was going for my daily exercise, walking and everything. But if I'd not been able to do that and not been in that position, well, I'm sure the suicide rate went up through the pandemic because of mental illness.

P13-F-57

8.3.1.2. Personal (internal) factors

This sub-theme describes the *intrinsic* or internal factors perceived by people with lived experience of depression and GPs to be important drivers of depression course. The internal factors outlined by participants included: genetic factors, biological predisposition, low self-esteem, personality types and coping styles. People with lived experience of depression perceived family history and genetic factors as important in contributing to depression.

There is an element of hereditary events, I still have the label in my family that, oh, I take after grandma, I'm one of the [family name] genes. There's always a member in the family that's got the mental illness and I was the one that got that crown, shall we say, in our family, so there's always that element of hereditary, nature versus nurture, isn't there?

P15-F-37

Some people with lived experience of depression perceived themselves to have an underlying, biological predisposition to depression. P11 described viewing this from the perspective of being a "chemical" process, and felt that this framing helped to explain his tendency towards relapsing ("to-ing and fro-ing"):

For me, it feels, the way I see it as being more chemical, I know there's science to it and the proof behind all that, because I don't

203

*suffer from trauma, I didn't have an abusive background, and that kind
of side to it all. Whereas I know many do, and many can, and so I feel
well if I've not got that, then what else have I got to be unhappy about
basically? And I feel as though it is more a kind of chemical
imbalance, which would explain the to-ing and fro-ing a bit as well, at
least in my mind if does.*

<div align="right">P11-M-31</div>

P13 had a similar view and made reference to "hormones" and discussed neurological causes to support her view of her biological predisposition to depression and relapses:

*I think it's obviously just how you are, how you are made and
your hormone levels and your…well, I don't know, is it a brain thing,
mental illness? It's the same with any illness, isn't it, what
people…because, touch wood, I haven't had COVID, so there's
obviously something in my genes that I'm able to cope with not getting
that, but I'm not able to cope with not getting depression. Whereas
someone else who had COVID might never, ever get depression. So, I
think just how you are made and what you're more prone to than
others.*

<div align="right">P13-F-57</div>

GPs also perceived that some internal factors, such as coping style and predisposition towards depression were important in driving depression course for some people.

*I: I wonder whether you have any thoughts about what do you
think are the more important factors that would predispose somebody
to be at a higher risk [of relapse]?*

*R: I think it's something that we don't measure. I think it's called
resilience…I compare it with boxing. Some people go out in the boxing
ring and receive a lot of punches on their chin and still stand, while*

other ones get one punch on the chin and they're down. And I think I can tell which patients can take a hit. I think resilience is what I try to find in them, how they're coping, what they're doing…The people that are internally depressed will always be depressed, they do need the medications to just get them to the level where they can do activities, get the shopping, get the food going, get a life going…Other people just sit on top of the wave and if the wave crashes, fine, I'll wait for the next wave. It's two different tribes that are different in different approaches.

GP19-M-61

People with lived experience of depression also described personality factors as important and also drew the distinction between "happy" and "unhappy" people.

I'm not a liked person… I just rub people up the wrong way so I know I'm not liked, so I have no friends… if you're generally an unhappy person you're always unhappy, and it's just how much you can put the mask up sometimes and sometimes you just can't. So sometimes people appear happy when deep down they're not really, is probably the only way to describe it. So are they not…they're not necessarily relapsing, they're always unhappy but sometimes they put a facade on more than others.

P2-F-40

Participants made the distinction between people with a tendency towards and optimistic or pessimistic outlook and thought that those with a tendency towards pessimism were more likely to relapse.

Well, I think, I mean, there's a saying isn't there? It's either half full or half empty. And obviously the people who think it's half empty are the ones that are most likely to relapse.

P14-M-74

205

8.3.1.3. When is it depression?

The previous two sub-themes were interlinked with a broader discussion of depression as a construct and how this is differentiated from associated constructs such as emotional distress. In this sub-theme, I will present and discuss the findings related to this distinction. This is relevant to the subject of my research and can perhaps lend us a more critical lens with which to view the underlying constructs of depression, relapse and remission of depression and, in particular, how these are applied in primary care.

A distinction between depression and distress as a result of life events was made by GPs and points to a feeling by GPs that a clinical diagnosis of depression is not necessarily appropriate in all patients who present with low mood or other mental health symptoms in primary care. GPs implied that depression is a term best reserved for those for whom antidepressant medication is felt to be helpful, whereas distress is best used for those for whom there is a clear, identifiable life event or trigger and that response needs to be something other than medication. This most likely reflects the undifferentiated case-load of people seen by GPs in primary care and "first point of contact" function provided by primary care. An illustrative example from GP14 introduces this distinction and how a consideration of this factors feeds in to a GP consultation. Interestingly, the GP used the phrase "just a negative life event" when describing an extrinsic trigger, and contrasted this with depression as a more endogenous process without an external trigger.

> *My initial approach is really sort of trying to work out what's brought the patient in, and whether or not it is depression or just a negative life event that's happened that's caused them to feel distress. And so a lot of what I'm trying to work out when they first come, is really 'is this depression or is this just stress, or something terrible has happened to them, and they're struggling to cope with that at the moment?' And part of that would be based on the timelines and how long that's been going on for and how severe the patient's symptoms are…I'm trying generally not to rush into medication, unless I think that's really needed at that*

point in time, and use a few weeks to try and see how, to sort of work out what's going on with that patient, maybe review them in a couple of weeks to see how things are going, particularly if there's been life events, 'cause I find that many people say that they're depressed, but what they means is they're distressed at the moment.

<div align="right">GP14-F-45</div>

To elaborate on this further, an illustrative example from GP3 highlights their perceived distinction between a "true depressive illness" and patients being "overwhelmed with life":

I'm thinking of a patient I spoke to this week actually, who had a suicide attempt in January, a true, actual, proper…like it was an accident that he survived, and it was the first time he told anyone about it was this week. So that, I was very happy to say, actually, no, you have depression, he has got clear symptoms of low mood. There's not actually, yes, he's got some life circumstances, but he's not feeling overwhelmed with life, he's feeling low and poor motivation and a true kind of depressive illness. But I don't think I actually see that many true depressive illnesses. It's more that overwhelmed with life thing, I guess.

<div align="right">GP3-F-32</div>

People with lived experience of depression also drew a distinction between depression with and without an extrinsic trigger. However, people with lived experience of depression viewed the two constructs as less distinct than GPs, or at least that the distinction was less important than GPs thought it to be. People with lived experience of depression, on the whole, felt that depression symptoms were similarly problematic for patients, regardless of whether there was an external trigger or not. They thought that extrinsic triggers can trigger an episode of "depression" and can contribute to increasing the risk of relapse and longer-term poorer depression outcomes. The following data extract illustrates this point effectively, by presenting a reasoned account of the different factors affecting depression onset and how both can be equally consequential for patients.

If your depression has a concrete cause, it's not going to be medically treated…like if you've lost a baby, you've lost a family member, you've lost your job, you know, going through any kind of grief or life event, or anything like that, and that's triggered an episode of depression, it's like you just kind of have to work through that, and you're not going to get drugs, because it's not because of the biological imbalances in your brain - it's almost just because it's a normal reaction to a life event…But, I don't think that's actually the case in reality is it, because those life events can almost make your likelihood of relapse and ongoing episodes worse, because there's a reason for you to feel depressed, and there's going to be a constant reminder of that.

<div align="right">P16-F-24</div>

Overall, the distinction between depression and non-depression/distress seemed to be of less importance to people with lived experience of depression, possibly reflecting the fact that people with depression do not necessarily work and think within the constraints of the biomedical model of illness and the constraints of the need for formal diagnosis.

In summary, social, personal and environmental factors are perceived to be valid predictors of a poorer longer-term course of depression by GPs and people with lived experience of depression. Both groups of participants recognised a meaningful distinction between a formal diagnosis of depression and distress following adverse life events. However, it was recognised that it can be difficult to tease these apart and that life events and stressors can also be associated with an episode of depression. It seems that, in primary care, the distinctions between emotional distress and depression with and without a proximate trigger are less clear-cut and less easy to discern. This theme demonstrates the perceived commonalities and differences between depression and its associated constructs and how life

events, ongoing life stressors and personal factors are thought to be inter-linked with them.

8.3.2. Theme 2: Relapse risk and prevention

This theme relates to relapse risk and prevention and is further divided into two sub-themes: first, relapse was perceived as important but there is limited discussion of risk or prevention in practice and, secondly, that there is varying conceptualisation of relapse, remission and recovery among GPs and people with lived experience of depression.

8.3.2.1. Relapse is important - but limited discussion in general practice

This sub-theme presents findings relevant to the significance of relapse to people with lived experience of depression and GPs, and the extent to which relapse risk and/or relapse prevention are discussed routinely in practice. All participants interviewed thought that relapse was important. People with lived experience of depression described their experience of relapse in evocative and emotional ways, and metaphor was used frequently to make this point:

> *I don't know whether it was a psychological thing, thinking, 'oh Christ I hope I don't suffer with that depression again'. And lo and behold…you would start to tumble and it was one of those where you were virtually on the top of the ladder and then the next day you were down on the bottom rung. It just happened as quickly as that and it was stupid things that triggered it off. You know, and then you struggled. You know, to climb back up the ladder it would take three months probably……just like somebody kicking your feet from underneath you and then you're finding that you couldn't walk anymore because obviously you were down on your knees. It was as quick as that.*
>
> P3-M-67

One participant reflected that they had been unprepared for the extent of the recurrence of symptoms that they experienced:

I think it was communicated to me as I was older… and I went back to the doctors, they said that it wasn't really something that I'm going to really ever be able to cure, if that makes sense, it's something that I'm going to battle with my whole life, but they didn't really explain that it would have really big ups and then really big downs. It was, kind of, explained that with the medication I would be on a constant baseline where I'm neither, sort of, happy nor sad it was more of just creating a baseline to begin with. I didn't really expect to relapse as bad as I did.

P10-M-29

People with lived experience of depression also reflected that they worried about the prospect of relapse even when well, with some expressing some hopelessness about the inevitability of experiencing a relapse:

I do [worry about relapse]*, yes. I know how low I've been many times. I know I'll get there again and I know it will keep happening.*

P2-F-40

I think I have…had times when I have felt really well and…I'm always aware that it could come back, and it's horrible when it does.

P4-F-50

Most people with lived experience of depression reported that they had not had a discussion with their GP about their risk of relapsing or about relapse prevention.

I've never had that conversation with them at all, it's never something that was brought up. Obviously, I know that I can relapse, I know that I have and I'm fully aware that there's been times where I've

got better, and then, stuff has just taken a nosedive, either through life circumstances or just through, I guess, brain chemistry going a little bit haywire. But I've never had: 'you can relapse and if this is what happens, then you should do this'. I've never had any briefing on how to deal with it, I've just found my own ways of dealing with it. I've never really been told about it. But I've experienced it, so I know it exists.

<div align="right">P20-M-25</div>

One participant reflected that relapse must be a newly identified phenomenon having never had the discussion with their GP about it before.

I don't know if it's quite a new thing, if they've only just realised that people with depression have relapses or anything, because it's not something that my doctor, or as far as I can remember…anything that we've ever discussed.

<div align="right">P12-M-39</div>

People with lived experience of depression felt that such a discussion would be beneficial and helpful as a means of educating them about their condition and setting expectations around what to expect over the course of illness. There was a view shared by many participants, however, that this would not necessarily be the case for all people with depression and that the appropriateness of such a discussion would be dependent on individual preferences.

I think it's patient dependent but for me personally it would have been important for me I think, I think that would have definitely been a good conversation to have at a young age. And they should have seen that I have a family history of depression and I'm likely going to have it for a long time, so I should be educated on the long-term effects of it.

<div align="right">P10-M-29</div>

Some GPs described feeling a responsibility to discuss the chance of relapsing with patients once they had made an improvement in their mood. The main barrier to this was a concern over whether a patient would want to think about becoming unwell again.

Well, they probably don't particularly want to consider themselves being ill again because they think, oh, if it's treated and it's sorted and things…But I think you probably…I think it would be a bit remiss of you not to have that discussion and at least point that out really.

GP8-F-52

People with lived experience of depression felt that the timing of a discussion about risk of relapse was important, and that some people may not wish to have the discussion around relapse. It was suggested that such a discussion should be individualised and agreement from the patients should be sought before entering the discussion.

That's a difficult one, isn't it? It depends when it was given. If that advice was given early on, I, because of the mental state I was in I would have thought, 'my god no'. Towards the end a bit guidance to say, but we've got your back, you've now improved. We hope you, well, we think you're going to keep on improving, but we're here for you, we are the safety net, and we'll catch you.

P9-M-74

Some people with lived experience of depression felt it would be useful for GPs to explain to patients about the potentially ongoing nature of depression and the risk of relapse. Such a discussion would have the benefits of ensuring patients were better educated on their condition, achieve better concordance with treatment, and setting realistic expectations around longer term depression course.

I think it would help clear up some of the potentially false expectations people can get from being prescribed antidepressants. It helps get rid of that…almost a cure-all attitude that you hoped for. If you set the expectations a little bit better, I feel like maybe people will be better on their medication, and things like that, or better when they've gone to their GP. If their GP then tells them, look, this won't make your depression go away, it'll make it better, it might even make it so you don't even notice it anymore, but there is always a chance it comes back, there is always something that could trigger it off, when that happens, if that happens, you can come back, you're more than welcome to come back, we'll review these medications, you're not by yourself and it's not unusual for this to happen.

P20-M-25

GPs recognised the benefits of discussing early warning signs of relapse and relapse prevention, but that it was not something that was incorporated into their current practice.

Ah, that's really interesting. That's not something that I've talked to them about really, about recognising the signs. Recognising the early warning signs and getting help early. That's a really good point, but not something that's currently something that I do.

GP1-F-36

Where GPs reported discussing relapse, this tended to be in the context of advising what to do if things get worse rather than discussing individualised risk or focussing on strategies for preventing relapse.

Yeah I do [discuss relapse]. *So, you know a little bit like you might with an acute patient where you say to them, look, please bring little Johnny back if he gets any of the following. So, with patients who have improved they often forget how bad they were bad in the first place. So, I'll say to them, look if you get these symptoms again, again do not watch this develop for months and months and months before*

you come and see somebody. If you think this is becoming more… If you think it's becoming a trend then you come back and see us and we do that level of kind of counselling with all of the patients.

<div align="right">GP20-M-53</div>

Some GPs expressed concerns that discussing the risk of relapsing may cause harm by making patients worried about their risk and discussed the concern about this becoming a self-fulfilling prophecy. The rationale for this concern seemed to be largely the worry about pre-empting or increasing the risk of relapse simply by discussing it, causing patients to over-think or focus on risk of relapse and, therefore, ultimately making them more likely to do so.

And, I think, concern over whether that's going to worry a patient or maybe even make that risk higher, because people are concentrating on their risk in a negative way rather than a positive way.

<div align="right">GP14-F-45</div>

8.3.3.2.1. Risk of relapse

GPs were generally not aware of any evidence-based tools to help assess and stratify patients according to their individual risk of relapse, other than, for some, the NICE Guideline (NICE, 2022, 2009) which specifies previous depression as a risk factor for a depressive episode:

I mean, I know there's guidance in terms of whether they've had one relapse before, then obviously they need to be staying on the medication at least probably two years after that or consider whether you're going to give lifetime medication. So that's about the only guideline I would say that I use. And I think, generally speaking, once somebody's had a relapse, they themselves know that they're probably going to need the medication for a lot longer and seem more accepting of that to be fair.

<div align="right">GP8-F-52</div>

Other GPs explained that they drew upon their own clinical experience and learning from practice to guide assessment of relapse risk in practice:

I don't think there's any tools as such, it's more about what I've learned over the years in terms of risk factors, and it's to do with things like whether they live on their own or not, whether they've got other pre-existing conditions, whether they've got alcohol issues, drug issues, any domestic violence issues, it's the usual risk factors for any kind of mental health problems. I think from experience, the biggest risk factors really are pre-existing comorbidities, such as COPD, heart disease and so on, and obviously past history of mental health problems.

GP11-M-40

GPs commonly described using their pre-existing knowledge of patients to help assess risk of relapse in an individual.

I wouldn't say I'm confident, but I think if there are patients that you know particularly well and have had the benefit of consulting with over a long time, so there are patients who I've known for six or seven years and I've been with them through relapses before and I know things that are likely to trigger them, and you can often just see the natural history, in those patients, yes, because the job is made easy because if they've relapsed four times previously, they're probably going to relapse a fifth. With a first episode of the depressive illness in a patient who I don't necessarily know well, I don't think it's easy to recognise the chance of them relapsing.

GP17-M-35

In summary, people with lived experience of depression and GPs thought that relapse is a significant problem, but relapse risk and prevention are not discussed in a consistent or structured way in practice.

8.3.2.2. Relapse, remission and recovery: interpretation and conceptualisation

8.3.2.2.1. Conceptualising relapse, remission and recovery

Relapse and associated terms (remission and recovery) were conceptualised in different ways by both GPs and people with lived experience of depression. Similarly, most people with lived experience of depression explained that the episodic nature of depression implicit in the description of relapse and associated terms is not one they necessarily recognised from their experience. In particular, people with lived experience of depression did not always understand the word "recovery" and felt that low mood was always present to some extent. Many GPs and people with lived experience of depression interviewed speculated that depression would be better thought of as a long-term condition to be controlled rather than cured. The following two data extracts from GPs capture this view of thinking of depression as a long-term condition.

> *And unfortunately for some people, depression just never really leaves them, it will always be there. They'll go through good times and bad times and during the bad times, it's our job to, you know, risk assess them and make sure nothing happens. But during the good times, still be there in a way because they are the ones that are likely to relapse, unfortunately.*
>
> GP2-F-38

> *I possibly think of it as like a flare up almost, of their mental health. Rather than it kind of ever necessarily going away.*
>
> GP7-F-36

Some people with lived experience described their depression as being on a continuum rather than a cyclical or discrete condition.

> *I don't think of them specifically as episodes, I think of it more like a scale from zero to 100, and I'm always somewhere on that scale. I'm*

never at zero, I'm somewhere between maybe 20 and 60 at all times. And if it goes past 60, then I go, things are getting bad again, something needs to be done. But I don't think of it so much as episodes, because episodes would imply that there's an ending to it, and that when you're not having an episode, everything's fine, and to me, depression doesn't work like that. For me, it's, you have depression, stuff makes you sad when it shouldn't, sometimes you just wake up and it's a bad day. Sometimes you'll have bad weeks, sometimes you'll have a bad month, but other times, you'll have a good week. But it's never a spell of depressed, fine. It's, are you this much depressed or are you this much depressed.

P20-M-25

Some people with lived experience of depression felt that the fact they did not necessarily fit within the episodic construct of depression and, therefore, did not recover meant that relapse prevention or ongoing support would not be offered to them.

I suppose, I've never been "better", so I don't think I've ever been eligible for any aftercare.

P10-M-29

GPs and people with lived experience of depression expressed similar views specifically regarding the conceptualisation of relapse, remission and recovery when applied to depression in primary care. One participant described his experience of "functioning better" but not actually recovering.

It's interesting, you know, looking at some of the terminology that you've used in some of the questioning about recovery and relapse. I think when I look back at my history, I'm not sure I've ever actually recovered, if I'm perfectly honest. I think it's just been a case of, you know, I've got to function again and just get on with it, actually…in the nature of the question, it's almost suggesting that you get better, and I'm not sure I ever sorted out any of my issues properly. And I think all

that's ever happened is that I've got into a position where I've been able
to function. So, I would look back now and say I've never recovered
from any of this.

<div align="right">P8-M-57</div>

Other participants felt the same, viewing depression as requiring
"lifelong management":

I don't know 'cause the whole idea of relapse, kind of, makes
me...like I question the whole idea of relapse, I think, maybe, it's just a
management, like a lifelong management. So, it's about like equipping
individuals with like tools, you know, that they can use.

<div align="right">P21-F-37</div>

Similar to people with lived experience of depression, GPs reported
being uncertain about the relevance of these terms to the care of people with
depression in primary care. "Improvement" in mood and function, usually
based on patient self-report, seem to be the principal points of history guiding
assessment of progress and decisions around treatment in practice. The
following data extract contrasts this approach with the "constructs" that
psychologists use and questioned the relevance to primary care.

I use the term 'improved' and talk about their feelings, and I
look at their function, but I think they probably are within the construct
the psychologists would use - we just don't use those phrases, and
perhaps those would be useful phrases to use with patients because
they would understand that there's recovery but there also could
be…it could re-establish itself…So, I think 80 per cent of it is patient
report of that, so we'll talk about whether they're able to work when
they weren't able to work, whether they're connecting with other
people or the outside more as I've talked about, and I'll try and gauge
what their plans for the future are so that I can see that they've got
more hope about things.

<div align="right">GP9-M-45</div>

<div align="center">218</div>

Most GPs described using improvement in the specific symptoms discussed at the initial consultation as a way of assessing overall improvement.

I probably think more in terms of feeling better and I don't really put any more threshold on it than that. I ask about their specific symptoms that they presented with. So, for example, someone presenting with low mood, poor motivation and poor sleep, and then ask about, so 'are you managing to get to work and you're feeling like you're quite enjoying your work at the moment and how's your sleep?' I'm going to ask about those specific symptoms that they said were bad before. And then I'm going to see where those things hit, but usually people are just kind of like, 'yeah, I'm feeling a lot better. My sleep's not quite there yet, but I've got this plan in place or whatever it is.' So, I tend to think more in terms of feeling better than a specific, that means that they are in remission or have recovered.

GP3-F-32

Some GPs reported using validated assessment tools for measuring improvement in symptoms but they did not seem to use these in a consistent or standardised way.

I say, I think, you know, things are improving…You know, if I've been using the GAD-7/PHQ-9[20], I'll say they have improved from that point of view. But I generally go off patients' experience…I've never really used those kind of terms [relapse and remission] *actually.*

GP5-M-30

Where the terms are used by GPs, their usage still does not align with their usage in the psychological literature or their original definitions.

---

[20] GAD-7: Generalised Anxiety Disorder Assessment; PHQ-9: Patient Health Questionnaire: screening and severity measure for generalised anxiety disorder and depression, respectively.

One GP felt that the term "recovery" was most appropriate to use after a first episode of depression whereas "remission" is more appropriate for patients with a history of relapses:

> *I probably think of remission more for someone who might have a long-term history of mental health illness, where I really suspected that at some point, this is not going to last forever, and they're likely to relapse again. For someone who's never had problems before, and it's a first presentation, maybe precipitated by life events, I think more in terms of recovery unless there's something that I know is going to be likely to bring this back again. Yeah, I see a lot of people who have one episode of depression on their records, and then that's it. So, yeah, I tend to think more of recovery in those, and then for people who've had life-long problems, much more I think in terms of remission.*
>
> GP14-F-45

In terms of assessing level of improvement, or identifying remission, GPs described the use of a more subjective approach, asking patients how they feel. In the preceding data extracts GPs describe using the term "improved", "improving", or "feeling better" and talking about "function" and being back to "normal self". Where validated diagnostic instruments (e.g., PHQ-9 (Kroenke, Spitzer and Williams, 2001)) are used, the results are not used in a consistent or pre-defined way and GPs did not report using any pre-defined criteria for magnitude of change but instead looked for a trend.

8.3.2.2.2. Role of medication

The findings in the following section are drawn from GP data only. Another way in which the conceptualisation of relapse was described was that GPs associated relapse with antidepressant use and specifically the stopping of antidepressant medication. GPs suggested "relapse" was a more useful term in the context of reducing antidepressant medication and used a discussion of relapse as a warning against weaning antidepressants too quickly.

So I think if somebody was coming off medication, so they were having a conversation about wanting to come off, then I would definitely, kind of, go into that [discussion of relapse]*. Sometimes people are really apprehensive about starting medication and then at that point I explain to them that there is a good chance that once you come off of the medication, the symptoms might come back. So that's another point at which I've, kind of, explained that. I don't think I specifically mentioned relapse in other situations, but I usually tell people that if you suddenly start to get worse or things stop being on track, then come back to talk again.*

<div align="right">GP18-F-34</div>

Some GPs explained they used a discussion of relapse to encourage longer-term medication concordance:

So, it's not in a kind of scaremongering way, I suppose, you know, it's educating them. But I suppose when patients are well, it's a good time to use that word to kind of almost, you know, make them realise, right, they need to keep doing what they're doing because it's almost like the memories of how bad they were are quite recent. So, if you say the risk of relapse, they're going to say, okay, I don't want to do that, let's keep going down the prescribed route, as it were.

<div align="right">GP2-F-38</div>

Not all GPs thought relapse was principally associated with reducing or stopping antidepressants, although reported that medication reductions were still commonly the time when relapse was mentioned and discussed.

I don't think relapse is necessarily linked with stopping medication. So, patients on medication could deteriorate, and that's when, obviously, you review them and look to see if you need to increase the dose of their medication.

<div align="right">GP12-F-43</div>

GP participants reflected that a relapse prediction tool may be most clinically useful around the time of reviewing medication and making decisions around reducing or stopping medications, and that this might be useful for both guiding medical decisions and also for helping to communicate either reasons to maintain the current dose or to encourage them to reduce the medication.

I think sometimes I think it is a conversation, certainly when I've done reviews on patients, I don't know, some patients may not have been reviewed for eight months a year or whatever, that that bit has got lost, that six-month bit. That bit where they're most likely to relapse or they'll try to come off in the winter and that will be a terrible time to do it. So, I think having a formal tool at six months where you identify these specifics, like what's changed, what tools have you used? What parameters have you noticed, what insight do you have if something changes. A tool of that sort at that point would be really useful.

GP4-F-37

8.3.2.2.3. Relapse is a negative term

Finally, there were some negative connotations of the word "relapse" described by both patients and GPs, particularly with respect to its use in the field of substance misuse, which may form a barrier to its routine use in a general practice setting.

I think relapse you associate with like addiction and stuff, don't you, you really do associate the word relapse with like alcohol and drug addictions.

P18-M-34

It might have connotations of other things like, you hear addicts, you hear like alcoholics or drug users relapsing, because they're trying to stay sober and then they relapse. So, I could understand maybe if someone didn't fully understand what the word

meant they might just automatically assume that, well is this person
trying to say I'm an alcohol or I'm a drug user, with relapse? But I
understand what that word means, yeah.

<div align="right">P23-M-37</div>

One GPs interviewed felt the word "remission" was more appropriately associated with physical health conditions, such as cancer or inflammatory conditions, rather than depression.

But, you know, I just use the word improving, or whatever. I don't
think I've ever used the word remission, I've got to say, in mental health
kind of…Yeah, so I suppose it is a right word to use, you know? But, I
don't know, I just don't tend to really. I don't know why, I always think
about remission being cancer related really. I would never sort of put,
they're in remission from their polymyalgia rheumatica or their giant cell
arteritis or whatever. But, you know, I suppose it's a correct word to
use, but I just don't really. I don't know.

<div align="right">GP13-M-46</div>

In summary, the terms relapse, remission and recovery are used and assessed inconsistently in practice. The framework developed by Frank et al. (1991) describing the change-points does not necessarily apply directly to patient or GP views on depression course in primary care, and there are some negative connotations specifically associated with the term "relapse" to be mindful of. It may be that re-framing of these terms in a way that resonates more in practice may improve the extent to which it is addressed. Finally, it may be that a tool to stratify patients according to relapse risk around the point of reducing or stopping antidepressant medication, or to guide medication withdrawal, has clinical utility and would be of use to GPs. I explored current antidepressant use as a relapse predictor in the qualitative workstream (Chapter Five) as a result of the findings from this qualitative work.

8.3.3. Theme 3: Relationships and communication

Within this theme, five sub-themes were generated exploring the role of the GP-patient relationship in the longer-term care of people with depression. In the section that follows, along with illustrative data extracts, I will explain the findings demonstrating that: listening and demonstrating empathy is key to a good GP-patient relationship; GPs and people with lived experience both value continuity for depression management; and that both groups of participants recognise the time and resource constraints in primary care. There are a range of current approaches of following people up throughout the continuation and maintenance phases of depression, which are of varying degrees of acceptability to patients and GPs. General practice was seen by GPs and people with lived experience of depression in this study as being at the centre of care. Both groups of participants felt that relapse prevention would sit most appropriately in primary care, but that there were barriers to this that would need to be addressed.

I will begin this section with a discussion of the way in which COVID-19 is perceived to have impacted on primary care by GPs and people with lived experience of depression. This impacts on some of the subsequent sub-themes and discussion. GPs reported an increased focus on triage and remote consulting, whereas people with lived experience noted it being harder to receive an in-person consultation when needed. GPs and people with lived experience of depression thought face-to-face assessment was important for people with mental health problems. Some GPs thought that, on the whole, the move to remote consulting over COVID-19 had impacted negatively on patients:

> *Yeah, I mean, there's different arms to that isn't there. There's the arm of before COVID we were completely patient facing. And I think for mental health, that is a massive change because there's a lot of hidden cues, you don't see the cues from patients on the telephone or in a message, you don't get the same feedback, and sometimes a*

lot of your worry in terms of risk is gauged by seeing them. So, I'd say from that point of view, I think COVID has been detrimental for mental health, in terms of management from health professionals.

<div align="right">GP4-F-37</div>

Some GPs felt that remote consulting had worked well during the COVID-19 restrictions and that patient choice around preferred modality of consultation is important moving forwards.

Prior to COVID, if anybody presented with anything with mental health, I would say I needed to see them face-to-face and I was quite blanket strict about that… I think COVID really changed our perspective on things and I used to do a fair bit of video calls and I used to do quite a lot over the phone…I'm moving back towards face-to-face now certainly and often if somebody presents as a new presentation, particularly to have not met them before. So, if I know them, if I've already met them, if I saw them three weeks ago for, I don't know, menorrhagia or if I saw them a few weeks ago for something else, I might say, look I know you, do you want to do this over the phone or do you want to come and see me. I guess sometimes give them more of a choice but if they're feeling really, really low, then I tend to insist that they come and see me.

<div align="right">GP21-F-33</div>

The same GP, however, outlined the benefits of face-to-face consultations, particularly for mental health presentations, such as the ability to identify non-verbal cues and assess mental state.

I think there is just so many things you pick up on, on face-to-face, that you don't get over the phone, just looking at how well kempt somebody is and somebody's demeanour and the way they walk through the door and their speech and their eye contact when you're speaking to them face-to-face, it's so different, there are so many cues

<div align="center">225</div>

that are non-verbal, that we miss over the phone, that I'm always keen to try and see, can we pick up on them at the beginning.

<div align="right">GP21-F-33</div>

People with lived experience of depression recognised that there were some potential benefits of remote consulting.

I know for me I'd much, much prefer face-to-face but a lot of people, whether it's like a mobility issue, or an introversion issue, or they don't want to leave the house, you know, like a lot of...like I do think that virtual reality has a definite place for a certain demographic of people, and they will respond much better to it, I think. So yeah, it's valuable but it needs to be offered as well as [face-to-face].

<div align="right">P21-F-37</div>

Some people with lived experience of depression agreed that remote consulting should remain an option for people and highlighted that removing the option may be have a negative impact on some people:

I think the virtual options are acceptable, for sure, especially when it's stuff like mental health, maybe not like, I've got a big gash on my arm, can I send you a text message about it. But when it comes to mental health, where people might not have the right mental strength to go to a place and go to speak to a person, if they're stuck in their bed, depressed, having the option to send a message or a phone call, I mean, it could be the difference between them not getting help and getting help. If they've not ever got the option, and their option is only to go face-to-face and they don't feel like they're up to it, then it's never going to happen, they're not going to get the help they need…It's one of those where, if there's an option there, if there's an option to have the option, it should be there, kind of thing.

<div align="right">P20-M-25</div>

On the whole, GPs felt that some of the changes to practice systems arising from COVID-19 are beneficial moving forwards, or that the practice systems had not changed significantly even when the modality of many consultations had.

I don't think it's impacted on our follow-up and monitoring particularly much actually in our practice. I am able to book people in. We have appointments set aside for on the day type things, and things are embargoed until certain times like most practices have. But if I feel like a patient really needs follow-up, as long as I don't do it for every single patient I see, you know, for every condition, then I'm able to over-ride those, as long as we still have some appointments available during the day. So, most of the GPs in our practice will over-ride some embargoes to make sure that people who need to be seen are followed up without the patient having to ring in the morning and things like that. So, I don't think that's particularly changed with COVID. If anything, it's carried on the same, that's how we tended to work pre- and post-COVID.

GP14-F-45

8.3.3.1. Importance of active listening and empathy

Active listening and empathy are factors that contribute to a strengthening of the GP-patient relationship and contribute to shared decision making around care. In this section, I will describe, with the use of illustrative vignettes, the benefits of listening and empathy and how they help to build a trusting and open GP-patient relationship.

In this section, I will outline the findings describing perspectives from GPs and people with lived experience about the positive outcomes associated with listening and empathy, as well as some negative encounters and experiences and how this can impact patients in the longer term. When people with lived experience of depression describe GPs demonstrating active listening and empathy, it seems to transform the descriptions of

consultations. People with lived experience of depression reporting having felt as though somebody was interested as being important in helping them to open up.

Most people with lived experience of depression explained how seeing a GP that they felt was listening and empathising facilitated shared decision-making around treatment.

I think just having a level of empathy as well. Because I think a lot of times, and again, doctors are very, very, very short on time, but I think it almost feels to me in my experience has felt quite rote and okay, so you've got this that's okay, I'm going to give you this then off you go. Whereas it felt like she was an actual person and not some kind of robot. No offense. But she had the empathy and just, I think the fact that she listened, she took time to explain things. She said that she was going to follow up with leaflets and information, which she did…It just felt like it wasn't just a, I'm going to listen to you and then I'm going to tell you to do this and then I'm not going to follow up. It was, I'm going to listen to you, I'm going to signpost, I'm going to give you the empathy that you obviously need because you sound like you're losing your shit a little bit, excuse my language. And then following that up afterwards, and it felt much more collaborative as well.

P17-F-36

GPs shared the view that they valued being able to have time to listen and support patients, sometimes without offering medication or other medical interventions.

So, it's nice to feel that you're able to support somebody and it doesn't always involve a prescription, sometimes it is time and listening and just being there and being that person, they can trust, that actually does that for them.

GP21-F-33

228

Conversely, a few people with lived experience of depression described negative encounters with GPs and how this adversely impacted on them seeking help again in future. Some described how the thought of seeing the GP and having a similar experience becoming a source of stress and anxiety in itself.

She couldn't be less interested and it just, kind of, puts you off even making an effort to bother ringing them back. Because I just think I just don't feel like I'm heard. I was talking and she kept going mm, mm, mm. I mean, you know somebody is not listening to you…My friends keep saying will you please ring the doctor. I'm like, yeah, yeah. And it's just not wanting to be not heard I think that puts you off… I'm already full of anxiety. So that anxiety gets worse having to think of even just ringing them. I just feel really anxious about it…you can tell when somebody empathising with you, can't you? And you can tell when they're understanding…It just felt like I was a box being ticked. It would be nice to have felt a bit more…the word care is coming to my head. But I just mean, yeah, just a bit more empathy really.

P1-F-45

Some people with lived experience of depression described losing trust as a result of a previous negative experiences with GPs after not feeling listened to:

I didn't feel validated or listened to, like I say, I felt very much an inconvenience and went through the system, ten minutes done, I felt very…I lost a lot of trust…I think for me it's about being able to sit down and actually know that somebody's listening and having that active listening makes all the difference to somebody when they're struggling and opening up about quite difficult subjects and difficult conversations…the other GPs that I've had a very good relationship with they're always, sit down, and they always come to you and it's very much about that non-verbal interaction of being present, being there,

that empathy, isn't it, of, I'm listening, I'm so sorry you're going through
this, well, let's see what we can do to help you.

<div align="right">P15-F-37</div>

8.3.3.2. Importance of continuity of care

GPs and people with lived experience of depression, without exception, described how continuity of care was important, particularly in the context of mental health presentations. The main perceived benefits of this were the fact that there is less need to explain the background and history in detail, rapport-building, and the ability of the GP to make an assessment regarding the change in a patient's presentation and progress over time. This applied in the case of both physical and mental health presentations and was felt to be particularly important in the earlier stages of a depression presentation compared with later stages when more stable.

I think for me…if you've invested that amount of time in somebody in the initial appointment, often follow-ups are significantly quicker and you can understand, when you're trying to touch base on things with people, you can say, look last time we talked about this, what have you thought about it and you can do a bit of a recap, exactly what happened previously. I like the continuity, I think it's quite doctor-friendly to have continuity as well as patient-friendly because they've got somebody coming in, you can see them on your list. So, it's nice to feel that you're able to support somebody and it doesn't always involve a prescription, sometimes it is time and listening and just being there and being that person, they can trust, that actually does that for them.

<div align="right">GP21-F-33</div>

GPs suggested that the need for continuity of care became less important as patients entered the longer-term follow-up phase of depression.

I think it's one of the main things. And I think, you know, it should be preserved at all opportunities really, because, what I'm finding at the

moment, is that people are getting very frustrated by, I suppose, seeing someone and then seeing someone else. And I always make a big effort for them to come back and see me particularly, because they're almost starting all over again aren't they, do you know what I mean?

And I think once people are into, you know, they're getting better, they're on sort of maintenance therapy, and it's six months we've got to see them, and I don't know, I think it's less important, but certainly at the beginning, developing that therapeutic relationship and sort of, so you've got a common understanding is really important, I think.

GP13-M-46

The other advantage to continuity as expressed by GPs was the ability to provide a subjective assessment of progress and improvement between the initial assessment and follow-up appointment.

I think continuity is incredibly important, especially with mental health problems. I think patients value it and I think it's much easier for the GP as well. When you talk about how do you know there's been improvement, that sort of objective measure can only really come from a clinician that's reviewed them previously, else you don't have a benchmark. So yeah, I think continuity is really key, especially in mental health problems.

GP17-M-35

People with lived experience of depression also described the importance and benefits of continuity and its role in building relationships between doctors and their patients.

I think it's very important, extremely, extremely. For me I wouldn't make an appointment unless it was with the GP that knew my situation, because there's been plenty of times other GPs don't know the situation with my back and they question the medication I'm on and the strength I'm on. And you know, it's...yeah, it's so hard

because, you know, when you pour your heart to someone and you feel like somebody knows you personally, and then you go into a room with a different GP it's, yeah, it's not the same, there's a coldness there, you know, not intentionally but out of just not knowing your personal situation.

<div align="right">P10-M-29</div>

Some people with lived experience perceived that continuity of care with their GP actually enabled them to anticipate a relapse and treat this pre-emptively.

I think that was very important because it was a continuation. I mean, he didn't have to look at the notes at that time because he recognised you as a person…The beauty about going to the doctors and me seeing the same one, you know, he knew exactly that I was going to be going down. He gave me the tablets and you knew that you could probably ring him up if things were getting worse…You knew that you could come out the other end.

<div align="right">P3-M-67</div>

8.3.3.3. Ongoing care and support

Analysis of the data from GPs and people with lived experience of depression suggested that acute management of depression is generally consistent. This sub-theme focuses on the longer-term care and follow-up of patients beyond this. The two primary focuses of this sub-theme were the type and quality of follow-up offered to patients after the acute phase (continuation phase) and also the longer-term appointments and follow-ups (in particular medication reviews) offered to patients in the maintenance phase. There was a tension in the data around the most appropriate ways to follow up patients, between what I will call patient-initiated follow up (patients are advised when and how to follow up if needed and retain responsibility for initiating this) and proactive follow-up (the GP retains responsibility for arranging follow-up for patients). Medication reviews, as a way of monitoring

depression, were generally inconsistently provided and were perceived as being superficial in their exploration of depression. Finally, there was a feeling from GPs that patients with depression who had antidepressants prescribed by their GP generally receive more intensive follow-up from primary care than patients who receive only psychological therapies from NHS Talking Therapies.

8.3.3.3.1. Proactive vs patient-initiated follow-up

A tension was evident between two principal approaches to following up patients in the acute, continuation and maintenance phases of depression. Most GPs described their usual practice as arranging follow-up for patients during the acute phase, particularly for the second appointment which usually took place within two to four weeks. The reason for this, as expressed by GPs, was primarily to try to ensure continuity of care. GPs generally thought that patient-initiated follow-up was appropriate in the continuation and maintenance phases, with patients given the responsibility to arrange follow-up.

GPs reflected on some of the practical barriers to being able to offer proactive follow-up to patients beyond the acute phase (mainly related to practice appointment booking systems) and what some solutions to this might be (the use of "scheduled tasks" - automated reminders to the GP to follow up the patient). GPs described giving the patient "responsibility" and makes use of consultation skills such as safety-netting[21] to advise patients when to re-establish contact.

> I think you balance that with this sort of idea of patient – what
> do they call it in the hospital – patient-initiated follow-up. So…I am
> giving some responsibility to them because my calendar doesn't book

---

[21] Safety netting is a consultation technique used commonly in primary care; it has been defined as "a consultation technique to communicate uncertainty, provide patient information on red-flag symptoms, and plan for future appointments to ensure timely re-assessment of a patient's condition" (Jones et al., 2019)

ahead that far. I can send myself a scheduled task to follow up on people, but I don't tend to do that with everybody, I tend to give them some ability to come back and safety net that if things are getting worse rather than better, then they would come back sooner.

<div align="right">GP9-M-45</div>

Most GPs described taking a similar approach:

It depends really. If they're improving, I'd say great, I'd do it at another four weeks to confirm that. And if they said yeah, no, I'm doing really well, that's fine… I would leave the ball in their court and I'd just say you know where I am, any problems let me know.

<div align="right">GP2-F-38</div>

People with lived experience of depression reported some concerns and adverse experiences arising from this approach. They described feeling as though they had been forgotten about or had not felt well-supported. They also described barriers to patient-initiated follow-up, which included low patient motivation as a result of depressive symptoms and system barriers including limited GP capacity and difficulty navigating practice appointment systems. People with lived experience of depression preferred a more proactive approach to follow-up, particularly during the acute phase and ideally beyond, although did recognise limitations on the ability of practices to offer this to patients. Some participants explained how a lack of proactive approach to follow-up can feel as a patient. One participant explained how impaired executive functioning and decreased motivation arising as a result of depression can impact patients' ability to initiate their own follow-up.

My previous experiences, apart from the last couple of times, have been very much falling into the latter category of being left to it…one of the main things for me, and I'm sure many other if not all people with depression are the ability to perform like normal executive functioning…I think leaving people to sort out their own follow-up when they're in a state where they can't even be arsed to go out and

get a shower…I think that's not the kind of task that you should be
leaving up to people. And definitely it should be standard that it is the
GP that does the follow up.

P17-F-36

Other participants reported similar experiences, with some discussing
the role practice appointment-booking systems can play in being a barrier to
patient-initiated follow-up:

So, it's a great idea but a) if you have no initiative because
you're feeling too ill and b) you ring up and they say you can't have an
appointment for three weeks or six weeks and you can't have an
appointment with the same GP. They've put so many barriers in place
that keep you away from the GP…by the time they do get to you, any
hope of being able to help or be proactive or follow things up or make
people feel cared for has gone.

P4-F-50

Some people with lived experience of depression suggested that
patients should be able to arrange their own follow-up, although this was a
minority view among this participant group:

I'm of the opinion that I shouldn't be putting onto GPs, I should
put it onto myself to get in touch again. I know you're very busy, I
appreciate, your time's more valuable than mine.

P14-M-74

Other people with lived experience of depression agreed with this
view, stating that patients should be willing to drive their own follow-up:

I think a lot on it does have to be down to patient. Don't get me
wrong a doctor, I suppose, every now and again, could at least do a
follow up. But I think when it comes down to it, you have to be main
driving force behind it because it's about you. So, if you aren't willing to

take the steps that you need to carry on, then nobody else is going to because in fairness, why should a doctor do something for somebody if they're not willing to do it for themselves?

<div align="right">P12-M-39</div>

There was an understanding by GPs that patients in the early stages of a depressive illness often lack the drive and motivation required to arrange their own follow-up and many GPs and patients explained the importance of reaching a plan for follow-up that was patient-centred, shared and clearly communicated to avoid any confusion around how to re-establish contact.

Yeah, I mean, the whole of primary care really has to rely on patient-initiated follow up, doesn't it? You add to that with those patients you're more concerned about, more worried about where you will follow them up regularly. And there are those that will then tell you when they feel they've improved enough that they don't require your regular type of follow up. They can now go to patient-initiated. So, again it's back to the bit that medicine isn't a one size fits all. You might start off with a regular level of follow up and then go to patient initiated, and then they have a relapse and your back to regular follow-up for a bit. It swings between one and the other.

<div align="right">GP20-M-53</div>

When discussing follow-up for people with depression, GPs frequently drew parallels with physical health conditions, where patients are expected to follow up their own results and arrange their own appointments. There was generally an understanding that in the acute phase, GPs may need to consider the need to proactively arrange follow-up, particular if depression is having a significant effect on function or if there were risk identified on the history. When relying on patient-initiated follow-up, GPs reported thinking that communication around how and why to establish follow-up needs to be explicit and understood by the patient.

Yeah, it's really interesting because this isn't just in mental health…Generally speaking, in the initial stages, certainly when initiating medications or if there's concern about risk, I will always ensure an appointment is booked prior to them leaving the room…I think once they reach a period of stability, I tend to put the focus more on them booking their appointments because really by then they're better, so they should be able to proactively manage their health. You could argue in the midst of a depressive illness when they just do not have that motivation or indeed the cognitive abilities to master a diary, I think it's reasonable that we take the lead, but at some point, that has to be put back to the patients, but that has to be carefully explained so they understand.

GP17-M-35

Many of the GPs described being guided by the patient's preference, indicating the role of shared decision-making and the importance GPs place on this when agreeing management plans.

So, I usually ask the patient, sort of, quite frankly how they'd feel, whether they feel that they want me to book something in or whether they feel that they just would like to book it themselves. I'm usually guided by however they respond. So, I would have thought that if I've asked them the question and I've left it with them, you know, I've made it in keeping with whatever they felt at the time.

GP18-F-34

To summarise this section, GPs and people with lived experience were broadly in agreement that follow-up in the acute phase of depression requires a degree of proactive follow-up from the GP. Beyond this, where there is a reliance on patient-initiated follow-up, this is better determined after shared decision making between GPs and patients and with very clear communication about how and when to re-establish contact. Safety netting around what to do if things get worse is also an important part of making sure this is an acceptable approach for patients and GPs. Optimising practice

systems to allow for easy access back into the appointment system and for continuity where preferred is likely to help improve both patient and GP satisfaction.

8.3.3.3.2. Longer-term follow-up of depression

With respect to longer-term monitoring (continuation to maintenance phase), the main way in which this is operationalised in practice is through medication reviews, which would typically take place after a period of time on a repeat medication. GPs and patients both reported that these can be at risk of being superficial, without a real opportunity for patients to discuss their mental health or make informed choices around their medication going forward. Some GPs shared uncertainty, or at least a lack of confidence in their practice systems, that these would occur within the intended timeframe. This is exemplified by an illustrative data extract below, where the GP describes a typical approach of limiting the number of prescriptions that can be requested so that a prompt to arrange a medication review is initiated. The GP expressed some doubt over how consistently this review occurs in practice.

Regardless I tend to put a review in…or I tend to limit the amount of prescriptions rather that they can get, six months, just a bit of a review. Whether that happens or not is another question in practice [laughs]... But that's how I tend to try and get them to sort of come back and engage with us again at an appropriate point, trusting that at least at a year that they will come round to a medication review.

GP5-M-30

As well as expressing some doubt over how consistently practice systems would result in medication reviews taking place, GPs also reported that the medication reviews when they do occur can be unstructured and quite superficial.

238

I usually put it on for six months initially and then it'll flag to review. But again, that review at that point isn't very structured. It's a case of how you getting on with it. Do you want to continue? Yes. Excellent. Right, okay we'll carry on. Are you in a place where you're ready to come off it? And I usually wait for them to instigate the, 'I'd quite like to think about stopping this now'. But the system will flag when the review period is up for the repeats, just so that they do get a quick kind of, 'how are you'

In those sertraline reviews, for example, that come up annually, it's usually a, 'are you feeling alright? Yeah. Good. Okay.' And move on. I guess that's more of a timing issue than anything else, there's just a lot of workloads at the moment. If we were to do a full review of everyone who's on sertraline, we wouldn't have time to do anything else.

<div align="right">GP3-F-32</div>

People with lived experience of depression reflected on their experiences of medication reviews for antidepressants:

Obviously, every year you have to have your medical review and they just say, 'is everything okay?' So it's just easier to say, 'yeah', just get all your medication and that's…and they've never really discussed my mood or anything.

<div align="right">P2-F-40</div>

Because I worked away from home, it was always over the phone for me, so you basically got, you know, 'are you alright, do you feel like killing yourself', and it was like, 'well, no, because I've got kids and a wife, I'm not going to do that', and it was like, 'right, well, yeah, I'll put you another prescription in then', that was kind of it.

<div align="right">P18-M-34</div>

GPs reported that they often relied on the annual medication review as a monitoring appointment, usually triggered by a patient reaching the maximum number of authorised medications.

We'll ensure if they're on medicines, either one or maybe two reviews a year if they're all stable, because at the end of the day, if they're on medicines, they need a medication review, and that would entail a proper review of their mental health. GP11-M-40

GPs reported potential weaknesses in this system; in particular, there is a risk that medications can be authorised without a review taking place. Workload pressures, lack of capacity, inconsistent practice systems and inconsistent approaches by different clinicians seemed to be the biggest driver of these perceived weaknesses in the system. The data from GPs and people with lived experience of depression were aligned on this subject, with both participants expressing some concern that medication reviews either did not occur with the intended frequency and consistency and, where they do occur, they risk being superficial and offering people insufficient opportunity to discuss their depression. GPs explained some of the reasons for the uncertainty around whether medication reviews occur as intended:

It's very easy just to keep re-authorising someone's medication by looking at the notes to see that they're not coming in and they obviously must be well, therefore just give them another six months' worth and so on. But I think if it's done properly, it should be a case whereby you do explore things in depth and are happy that they are on the right dose and either continue it or change it over…I don't think the resources are there, I think because of the amount of work we're doing, some people will unfortunately take that shortcut, make assumptions about someone's wellbeing because they're not being seen because they're on the same dose for how many number of months, some people, what I've seen in my place of work that will just extend that without asking for an actual review to be taking place.

GP11-M-40

People with lived experience of depression corroborated these concerns and reflected how this is perceived from the patient perspective:

> *Every time I got a repeat prescription my review date just got later and later and I just thought, oh, well, I must be alright then. And then, like I say, my literally last review is online which I could have just put anything…there was no phone call to say, oh, are you okay, are you happy to still take these tablets, blah-de-blah? And then I got…I think it was either by email or text from the surgery saying, oh, we've got a…your review's due, could you please go, download this app or go into this app, onto this link, and complete the email questionnaire? And that's all I got.*
>
> P13-F-59

8.3.3.3.3. Differences in care and follow-up

GP accounts of follow-up revealed a difference in follow-up arrangements for patients depending on whether their depression is treated with antidepressant medication or psychological therapy. The findings under this part are wholly drawn from the GP data. Where patients are prescribed antidepressant medication, GPs accepted that it was their responsibility as prescribers to follow up patients appropriately as the main instigator of treatment. However, there was less of a sense of ownership and responsibility for people who were referred for psychological treatments (under NHS Talking Therapies, for example), as GPs thought that they were not the people administering the treatment (and therefore less responsible for monitoring the effects), but also that the provider of the psychological treatment would be providing follow-up and therefore the GP would be duplicating work. These views are illustrated further by two data extracts from GPs.

> *I suppose you maybe have a bit less regular input with some of those patients, so you maybe don't have the same opportunity, so if patients are feeling that they don't need medication and they're going*

to get some psychological input, then if I'm concerned about people I'd still maybe catch up with them in a few weeks' time, but a lot of the time you leave them a little bit more to their own devices to go take on that support. And I suppose with another professional involved, then you're not then maybe in touch with them as – rightly or wrongly, you're not maybe in touch with them as closely as you would be if you're titrating up medication or something like that.

GP10-F-44

With the patients on medication, I am treating them. I'm responsible for the effect of that medication, whether good or bad. Ultimately once I've assessed a patient with depression who decides against medical treatment and wants to do psychological therapies, in my mind I'm not offering those psychological therapies, so I can make the referral, but then it's a very much an open door for them to get in touch if they need. I see little merit in repeat consultations if they're just seeking psychological therapies, because I don't see that I'm then adding huge amount to their care in the slightest…yeah, I think as a rule, if they're seeking psychological therapies alone but not on psychotropic medications, I'm certainly adding less, so would not review them as closely as those on medications that I prescribed.

GP17-M-35

GPs also suggested that patients opting for psychological treatments rather than antidepressants may have less severe depression and be more motivated and able to take responsibility for their own follow-up:

I think actually the patients self-select…so if they're choosing the talking therapies, generally they're a bit more self-motivated, like big stereotypes, so generally don't want that much support because they know what they're doing. My counter then is that I find often that if they're not seeking medication, that they are generally not as severe, so they're sort of…yes, they may be depressed or anxious or whatever

but actually they're not as bad… this is just my perception, rightly or wrongly.

<div align="right">GP22-M-36</div>

8.3.3.4. The patient at the centre of care within general practice

This sub-theme was generated from the data which captured the importance of general practice at the centre of care and how longer-term care for depression and relapse prevention might be best incorporated into routine general practice. It fits most logically under "Relationships and communication" and builds on the earlier sub-themes but, as I discuss the implications of the findings, I will come back to a discussion of relapse prevention and discuss its place within primary care. This will lead us on to a discussion of barriers and facilitators to the implantation of relapse prevention in primary care as part of the following sub-theme ("Limited GP time and resources").

First, I will present and discuss the theme, through the form of illustrative data extracts, relating to the role of general practice (GPs, other primary care health professionals, non-clinical staff, and the practice itself) in people's care and life. Advantages from the patient perspective included proximity to home, GP surgeries being safe and familiar places, care being co-ordinated by a professional who knows them well, and avoiding perceived unnecessary involvement of additional services. The findings in this section have potentially wide-reaching implications for research into relapse prevention moving forwards, particularly in light of other national policy changes and local reorganisation.

The data extract below illustrates the benefits from the perspective of people with lived experience of depression of care being located within a local and familiar general practice setting. Not only is this more accessible but it is felt that the care is more joined-up and integrated when located within the same setting as the GP who has made the initial assessment.

A lot of people don't really want to go and travel out, or they don't have the confidence to make a call and explain something to someone that they don't know. I feel like if there was something in-house where they feel their GP has spoken to this professional and through proxy, sort of, knows them personally in a sense. There's just more personality to it and less coldness, I suppose.

P10-M-29

Other people with lived experience of depression elaborated on this further, describing the benefits of a familiar setting at a time when people are often feeling vulnerable anyway.

Because a lot of people have been to their GPs for a long time. When you live in an area for a long time, you get used to going to the doctors for things, it's familiar ground. Whereas, going to, say, a therapist office, or a place that you've not been before, it creates that extra element of unknown. Whereas, you view your GPs as more of a safe space because every time you've gone there, it's been because something's wrong with you, and you've come back, hopefully, with something that was wrong with you better, therefore, your brain goes, that's where I go to get better. So, it creates that link of you going to the doctors, there's a problem, doctors fix problems.

P20-M-25

To expand on this further, the following extract explains the perception by people with depression that care is more integrated and joined-up when located in one place under the supervision of one clinician, as well as reiterating the benefits of general practice as a familiar, comfortable space for patients.

I just think because when you're at your GP that…I don't know if we're expecting too much of GPs obviously but, then again, instead of having all these individual groups should it all be within the GP service? So, you can go…when you're not 100 per cent and you're

not confident you don't want to be going to all these unknown places, all these different places, whereas you know your GP is a familiar spot, it's somewhere that you're used to, it's somewhere that's nearby to where you live. Because obviously I lost my confidence in driving and getting on a bus, so if you need to travel to somewhere whereas I can just…if you can walk then, again, it's that comfort, it's your comfort zone. And it's somewhere that, no doubt, you will have visited. And also I just think it'll be helpful for the GP themselves if everything was under one roof because you could communicate, that word, communication, and have discussions with people and they could say, go to the GP and say, what's your view on this? And then the GP can give their advice to them. Surely it's better working as a team than individual people.

<div align="right">P13-F-58</div>

I explored with participants whether relapse prevention was something that would be better situated within primary care and, if so, what the barriers and facilitators to this. People with lived experience of depression thought there would be significant advantages to this, which build upon the extracts reported earlier.

Yeah, I think that would be a big advantage because then they could relate back to the doctor and maybe if they have picked up anything they could maybe refer them to the doctor in particular that was dealing with it. So, I think it has to be an in-house connection somewhere along the line. Otherwise it becomes totally detached… Especially in this day and age because as I say, you know, seeing a consultant now, the consultant seems to be detached from the doctor and then the doctor…You know, you're ringing the doctor up and he says, well I'll have a word with the consultant. The consultant says well haven't you seen your GP?

<div align="right">P3-M-67</div>

GPs agreed that relapse prevention was something that could and should be situated in primary care and discussed some of the opportunities through Primary Care Networks (PCNs)[22] as a way of making this feasible. However, the feeling from GPs was that this could not be delivered within the current resources in primary care and would require additional funding and resourcing.

> *I think in an ideal world it should be within primary care. The difficulty is…as primary care is currently structured, which I appreciate is going to be different practice on practice, but psychological therapies for reducing relapse as it stands, no, I don't think we've got the capacity to be doing that within primary care. I think it's absolutely feasible, so with, for example, the mental health workers new in post, I think having a funding stream for, look, you are commissioned to deliver four sessions of risk reduction of relapse prevention treatment*
>
> GP17-M-35

GPs discussed the need for a business case to provide justification for funding and delivering this in primary care. GP9 thought that being able to demonstrate an overall need for GP consultations in the longer term would be one such way of evidencing benefits of implementing this, as would demonstrating longer-term better outcomes for patients:

> *I think if it were…I think you've got to make a business case for it, I'm learning this, that if you want GPs to use a finite resource, they will do that if you show them the evidence that that prevents or helps patients in the long-term, and in the end, results in fewer consultations because they're better, healthier people.*
>
> GP9-M-45

---

[22] I refer the reader back to Section 1.5 for an explanation of Primary Care Networks.

The following sub-theme will build further on some of the ideas presented towards the end of this section. As part of a discussion of the current resource and time limitations within primary care, I will present the findings exploring further possible solutions to the problem of implementing relapse prevention in primary care.

8.3.3.5. GP consultations are a limited resource

Having established the case for embedding relapse prevention within primary care, but recognising the need for additional resource or funding, I will end by discussing the findings around limitations on GP time and resource and options that might support this. There was an overarching perception from both groups of participants that GP time and resource are limited, in a way that can impact on the quality of care provided by GPs and, in turn, patients' willingness to seek help from their GP. People with lived experience of depression were generally sympathetic to GPs regarding this but recognised limitations in funding as being responsible. Some people with lived experience of depression felt that the time-pressured environment of modern general practice has affected the attention and care offered by some GPs.

God knows how you'd come on now with your ten-minute slot at the doctors. I mean, when I went in there you spilt your heart out and told them how you were suffering and the likes. And they had a sympathetic ear where they were able to listen to you, but as I say now when you go in they're sort of watching the clock and you've got ten minutes and once it gets to nine and a half you're bundled out regardless. So whether you're cured or not.

P3-M-67

Similarly, other participants described recognising the pressures that GPs are under but also that this results in less satisfactory patient-doctor relationships.

247

So, you know, my feeling… And it's not blaming them. It's just how it is, really, isn't it? Everybody's under a lot of pressure, but nobody actually knows me. You know, I'm just somebody on their list of jobs for the day, really, aren't I? And that's an inevitability of how our system works at the moment.

<div align="right">P8-M-57</div>

One person with lived experience provided disconfirmatory evidence on this point and felt that GPs don't actually have the time or resource to deal with depression at all.

GPs haven't got time to explain about depression. I don't think they've got time to deal with depression really, when it boils down to it…They haven't got time for discussions. GPs don't have time to discuss things with anyone. They're so busy. I know that I've heard that this government or previous run governments since, say, 2010 have declared that they're putting X amount of money into mental health, and one good way of using that money would be to genuinely see that depression is different from any other illness. It's different from anxiety and it's different from any other mental health illness and it's different from any physical illness. So once they acknowledge that to themselves then they act differently, and when they act differently hopefully the people at the top would say right, we need a different service, and it can't be done through the GPs, they haven't got time. So it all boils down to time is money. Even the NHS has to be run on 'time is money' as if it's a business.

<div align="right">P19-F-60</div>

Regarding capacity within the current resource within primary care for relapse prevention, the following extract summarises the view of GPs interviewed.

So I'd say there's no capacity for GP. It just doesn't exist, let's be realistic, it just doesn't, does it…I can tell you from all the triage sessions I do, the capacity is not there.

<div align="right">GP4-F-37</div>

GP participants discussed the barriers and facilitators to implementing an embedded relapse prevention programme into primary care settings, and specifically a GP–led setting. Any provision made to enable additional recruitment would need to be followed by a commitment to fund IT and other essential estates, such as physical room space, to enable the embedding of more non-GP staff within the primary care setting.

I don't think we've got any resource to do it. I think in terms of if the resource was provided and you were like providing a room or whatever, then yeah, although rooms are really difficult at the moment for most people I think. Or you had a referral to an online thing and you just did that with them, yes, I think that would be perhaps fine, but us delivering it, no, probably not.

<div align="right">GP15-F-47</div>

A small number of GPs noted the perceived lack of capacity to provide psychological therapies for people with acute mental health presentations and how this would make it difficult to justify dedicating resource to relapse prevention.

It would be feasible, yes, it's tricky, because, I mean, one of the issues would be there's a primary shortage of CBT resources for people presenting as a primary, kind of, acute situation, where they feel their life is not, you know, it's not going well and they're not able to function on a day-to-day basis and they need acute and, what they would perceive, as pretty urgent support and therapy. And then to use some of that resource for extra additional sessions, that may or may not be beneficial to someone who is somewhat functional, at the point of which the session was offered. It would certainly need some evidence that

that was good use of resources and that it was actually, you know, beneficial and it would be difficult, given that resources are tight and that the professional, sort of, input is limited.

GP16-M-43

GPs stated finances and a lack of confidence in the relatively new mental health practitioner role[23] as being a barrier to embedding relapse prevention within primary care.

I think one of the barriers is always going to be a financial barrier in terms of, well, it's 2022 and as of yet we don't have a mental health worker that is seeing patients, that will change next month, but the vast majority of my time in general practice, that's seven years here, six of which as a GP, we have not had a mental health worker for any longer than 12 months. We as a practice wouldn't see that funding one ourselves as beneficial because it's not clear that it would take work away from the GPs, I think the feeling is we'd still be seeing the patients as well, and ultimately we don't want to be funding it when actually there should be funding in place already for good psychological therapies.

GP17-M-35

GPs also discussed the limitations on physical room space and estates to being able to accommodate additional services within the practice premises.

It's actually a critical issue for us here. When we built this place 20 years ago I thought we'll never get to 10,000 patients because we started at six and a half. And I thought I'll be retired by then. It's somebody else's problem but hey-ho here we are. Breached that number probably three years to go before I retire and now we're

---

[23] This role was introduced in 2019 via the PCN contract Additional Roles Reimbursement Scheme, to support the recruitment to non-GP roles in primary care (NHS England and NHS Improvement, 2019)

scratching our heads doing hot rooming or we actually provide three times as much EA as we need to, because I've got to shift people into EA hours just to get enough desk space…what PCNs have spectacularly failed to plan for and the same with CCGs is that fact that these staff…these ARRS staff need desks to work from. PCs to work from. Telephones. So, the entire infrastructure is lacking both in terms of estate and in terms of hardware.

GP20-M-53

As a clinical director of a PCN, I know one of our practices, the GPs have to work from home one day a week, they're forced to work from home, their manager can't find room for them in their own surgery, and that's because… And to be honest, those two practices, the smaller practices are served by the mental health worker and the big one isn't yet, and so we'll run into the same problems with space in the next two years of the PCN contract, is if we manage to recruit to all those roles, we won't have the room to put them in.

GP9-M-45

There were some positive reflections around the current role of mental health practitioners, and GPs felt there was scope to broaden the role to incorporate relapse prevention. The following data extract from GP9, a PCN Clinical Director, illustrates this view:

Yeah, so we've got a first contact or primary care mental health worker 50-50 role with the Mental Health Trust. There aren't enough of those roles though, so I'm in a PCN with 48,000 people and we've got one, we've just employed a second, but even two can't go over the whole population there. And we tend to use them at the start as first contact, rather than as looking at relapse prevention, and I could see that as a really, really good thing actually, perhaps a newer aspect of that role…We also have social prescribers and I would venture to guess that many, many people on their caseloads have mental health

problems too, but they're not trained medical people and I think they would worry if they were spotting signs of recurrent depression actively if they were part of the thing, I think they'd feel like that was slipping into making a diagnosis which is not something that they want to do.

<div align="right">GP9-M-45</div>

GPs saw definite possible advantages for patient care and integration of care by implementing relapse prevention within primary care rather than relying on external organisations and again felt that mental health nurses or practitioners, working within a practice or PCN footprint, as being the most feasible means of implementing this longer-term.

I think our mental health nurse would be brilliant at doing that and I mean at the moment…there definitely would be the time available there within her role to do some relapse prevention and maybe re-see people and go through our list. You could almost say, right who has had a code of depression in the past six months that hasn't had a review and she could call them and see them, that sort of thing, not necessarily do the medication review but do all the other factors. So absolutely, I think there is definitely a role for it in primary care, I don't think sometimes secondary care are quite aware of how much aware of how much we manage in primary care. I definitely think there would be lots of patients that would be very, very well managed in primary care, if we have enough resources to do it.

<div align="right">GP21-F-33</div>

GPs felt that this would have the advantage of allowing more joined-up care for patients, and highlighted deficiencies in the correspondence they receive from IAPT as unhelpful.

There's a theoretical opportunity because we recognise absolutely it is going to be far more beneficial our patients being treated by a mental health worker that works with us that we know who we can then talk to, because the care is very disjointed currently…having someone

that's employed on a more local footprint, whether that's practice level or PCN level, has the potential to be incredibly useful, because then we can have MDTs with our mental health patients, mental health staff and say, look, these are who we're worried about, these are the guys we think might be at risk of relapse, do you mind making some contact and seeing them, they can then report back to us. Because at the moment the way we get correspondence from IAPT it's from a practitioners' name we've never seen before who will send us a letter that unfortunately often just reads in a way that they wish to offload responsibility for that particular patient rather than true information sharing, which is a shame.

GP17-M-35

The findings in this theme point towards benefits of person-centred care, situated within general practice, for longer-term outcomes for people with depression. PCNs and MHPs are potential vehicles of delivery and implementation but additional resourcing must be considered.

8.4. Summary

In summary, GPs and people with lived experience generally agreed that they value active listening, empathy and continuity of care for people with depression. Both groups were accepting of limitations of time and resource but felt there were ways of improving and optimising care within the available resource. Patient-initiated follow up during the acute phase can be difficult and if practices are able to accommodate more proactive follow-up, where this is the patient's preference, care is likely to be more acceptable. Beyond the acute phase, patient-initiated follow up is likely to be used but it is important the conditions of this are communicated well to patients and that they are able to get back into the system if needed. Medication reviews are a good way of ensuring monitoring for people with depression and care should be taken to ensure these happen when needed and that they present a real opportunity for patients to discuss their mood. Relapse risk and prevention

are not routinely discussed in practice, although GPs and patients recognise the significance and importance of this issue. Patients should be offered the opportunity to discuss relapse with their GP or other healthcare professional.

There are two key implications arising from these findings. First, I have developed a framework (Figure 8.3) to guide the ongoing care and follow-up of people with depression in primary care. This framework captures the lessons drawn from the study findings and presents simple, achievable measures that can be taken to improve care in general practice using the current, available resourcing. It is intended to provide evidence-informed recommendations to guide care in a way that builds on the ideas and preferences of both GPs and people with lived experience of depression. It is not intended to place additional financial or workload demands on GPs.

Secondly, the findings presented here have demonstrated that people with lived experience of depression value their ongoing care being situated within general practice. Within the current GP contract and provision, there is unlikely to be sufficient capacity to allow relapse prevention in primary care, although patients and GPs feel this would be the best place for the delivery of such care. This would require additional resourcing, and tthere are current barriers but also potential future opportunities to improving provision of care in this area. I will provide a more wide-ranging discussion of how policy, local implementation and practice-level changes might facilitate a more formalised implementation of relapse prevention in primary care in Chapter Ten.
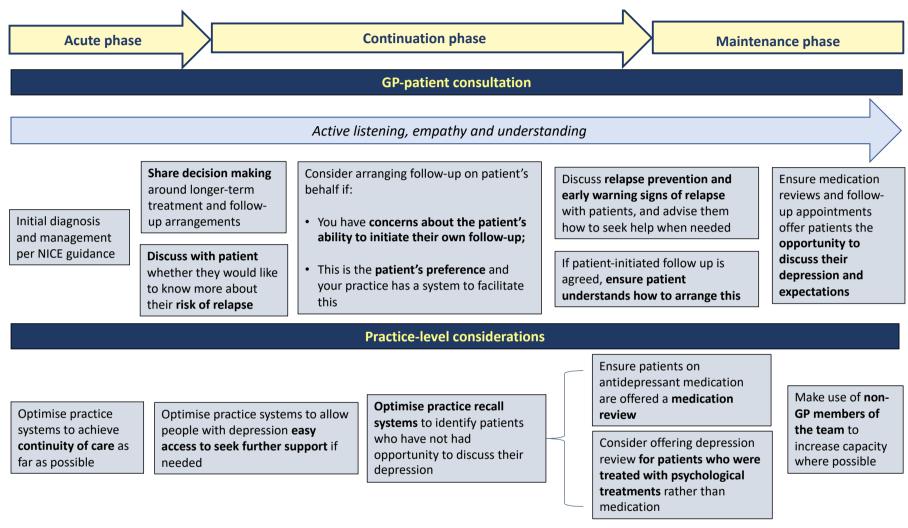
| Acute phase | Continuation phase | Maintenance phase |

GP-patient consultation

Active listening, empathy and understanding

Initial diagnosis and management per NICE guidance

Share decision making around longer-term treatment and follow-up arrangements

Discuss with patient whether they would like to know more about their **risk of relapse**

Consider arranging follow-up on patient's behalf if:

- You have **concerns about the patient's ability to initiate their own follow-up;**

- This is the **patient's preference** and your practice has a system to facilitate this

Discuss **relapse prevention and early warning signs of relapse** with patients, and advise them how to seek help when needed

If patient-initiated follow up is agreed, **ensure patient understands how to arrange this**

Ensure medication reviews and follow-up appointments offer patients the **opportunity to discuss their depression and expectations**

Practice-level considerations

Optimise practice systems to achieve **continuity of care** as far as possible

Optimise practice systems to allow people with depression **easy access to seek further support** if needed

Optimise practice recall systems to identify patients who have not had opportunity to discuss their depression

Ensure patients on antidepressant medication are offered a **medication review**

Consider offering depression review **for patients who were treated with psychological treatments** rather than medication

Make use of **non-GP members of the team** to increase capacity where possible

Figure 8.3: A Framework to guide the ongoing care of people with depression in general practice

Chapter Nine

Qualitative Findings 2: Perspectives on prediction models and risk communication in general practice consultations

This chapter presents the findings of the qualitative study that pertain to prediction models and communication of risk in general practice, with a specific focus on the prediction of risk of relapse of depression. The focus of this study has been prognostic models (models that predict future outcomes from the values of pre-defined predictor variables). However, the implementation and use of diagnostic prediction models (models that predict the presence or absence of a particular diagnosis, given the values of certain variables) also have implications for the thesis and were also discussed with GPs.

Perspectives on prediction models were explored generally and then with a specific focus on risk prediction for depression and depressive relapse. This formed part of the mixed methods approach taken throughout the thesis. Qualitative work with patients and providers alongside prognostic model development is recommended to guide development and implementation (Hoesseini A, van Leeuwen N and Sewnaik A, 2022). This qualitative work was intended to support the development of the relapse risk prediction tool reported in Chapter Six, and understand enablers and barriers to implementation in primary care.

9.1. Research Aims

The aims of the qualitative work overall were presented in Chapter Seven. The aims addressed by this chapter are:

- To understand GPs' current practice with respect to risk models and to explore the ways in which they are currently used and communicated to patients;
- To explore the potential implementation of a prognostic model to predict relapse risk in primary care from the perspectives of people with lived experience of depression and GPs, including what the enablers and barriers to this might be.

9.2. Prediction models in primary care: the GP perspective

As the purpose of this section was to understand the use and perceptions of risk prediction models by GPs, this section presents findings from the analysis of interviews with GPs only. It is the only section of the qualitative findings for which data were generated through interviews with only one group of participants. GPs reported using prediction models in primary care for the purposes of both diagnosis and prognosis. I will report in this section which models are used, why they are used and how they are used.

9.2.1. Which prediction models are used in general practice

Analysis suggests that a relatively small number of prediction models are used for physical health presentations. The prognostic prediction models reported to be commonly used by GPs include: the QRISK (Collins and Altman, 2009), CHA_2DS_2-Vasc (Zhu, Xiong and Hong, 2015), HAS-BLED (Zhu et al., 2015), FRAX (Schwartz et al., 2011), and ORBIT (Lip and Lane, 2015). Diagnostic prediction models reported as being used were: NAFLD fibrosis score (Angulo et al., 2007), FIB-4 (Lee et al., 2021), QCancer (Chiang et al., 2015), the Well's scores for pulmonary embolisms and deep vein thromboses (Silveira et al., 2015) and FeverPAIN for predicting the likelihood of a bacterial rather than viral sore throat (Flynn and Hooper, 2020) (see Table 9.1).

No diagnostic or prognostic prediction models for mental health presentations were reported as being routinely used in primary care by GP participants.

Table 9.1: Table outlining prediction models reported as being used in primary care

Name of model	Type of prediction model (diagnostic or prognostic)	Purpose of model
QRISK	Prognostic	Predicts 10-year risk of developing cardiovascular disease
CHA$_2$DS$_2$-Vasc	Prognostic	Prediction of stroke risk in people with atrial fibrillation
HAS-BLED	Prognostic	Predicts major bleeding risk in people with atrial fibrillation who are anti-coagulated
FRAX	Prognostic	WHO Fracture risk algorithm for people with osteoporosis
NAFLD fibrosis score	Diagnostic	Predicts presence of liver fibrosis in people with non-alcoholic fatty liver disease (NAFLD)
FIB-4	Diagnostic	Predicts presence of liver fibrosis in people with non-alcoholic fatty liver disease (NAFLD)
QCancer	Diagnostic	Predicts risk of having cancer
ORBIT	Prognostic	Predicts major bleeding risk in people with atrial fibrillation who are anti-coagulated
Well's scores	Diagnostic	Predicts presence of deep vein thrombosis and pulmonary embolism
FeverPAIN	Diagnostic	Predicts presence of bacterial (Streptococcal) pharyngitis in people with sore throat

9.2.2. Why prediction models are used in general practice

Two key themes were generated on the subject of why prediction models are used in general practice: some models are perceived to be useful by GPs (i.e., GPs are internally motivated to use them), and some models have to be used (GPs are externally motivated to use them). These two perspectives are not mutually exclusive, but if neither applies then this study suggests there is no reason for them to be adopted and used by GPs in practice. The analysis also showed some key enablers and barriers to models being routinely used in practice.

9.2.2.1. Some models are perceived to be useful

GPs perceived prediction models to be useful for four key reasons: they guide discussions around management with patients; they aid clinical decision-making; they are perceived as valid (trustworthy); or they protect GPs (medico-legally). Models were perceived to be useful when they guide discussions with patients, particularly when their proposed management plan is something the GP expects the patient may not be happy with. The use of FeverPAIN for sore throat was used as example by GPs to illustrate the way in which prediction models can tangibly support diagnosis, treatment plan and communication with patients to explain decisions. GPs described the benefit of being able to provide patients with an objective measure to explain the likelihood of viral infection rather than bacterial infection, to justify a clinical decision that can sometimes lead to difficult discussions with patients.

> *So, I think the FeverPAIN score is really helpful. I don't use that, necessarily, for my own clinical judgement but I use it for my patients, because I think when I'm not prescribing antibiotics it can be a difficult conversation. But actually, when I say 'there's a really good and clever reason and I can reassure you for that', then I think patients take it better than saying 'I'm a really clever doctor and I don't think you need antibiotics'.*

GP6-M-33

GPs reported that prediction models are more likely to be used when they are perceived to directly influence clinical management. This was particularly the case when they perceive that the model provides an additional piece of information that could not have been gained from history-taking or routine clinical assessment alone (i.e., the tool adds something above what clinical assessment can offer):

I think also if I can see a need for it. And that's why I suppose for the QRISK, deciding on whether the patient should have a statin, I need that evidence. Because a lot of the times, patient will have a high cholesterol but actually their risk can be one per cent. And I would have started them on a statin unnecessarily and vice versa. Whereas I suppose with the PHQ-9, I feel I can get all that information just by asking questions and talking to them. But there's no questions I can ask for a QRISK that will get me to the same answer.

GP2-F-38

GPs were generally more likely to perceive models as useful if they were seen as trustworthy and valid. Many GPs were interested in the evidence base for prediction models, particularly those they used commonly, but said they usually would not consult or critically appraise the primary evidence. Usually, the fact that prediction models were endorsed by the NHS or NICE was enough to convince GPs that the models were sufficiently supported by evidence to warrant using.

I think…how widely used they are, but they're embedded in your clinical system, therefore you think that they are what everyone does, they're recommended by…NICE, and so therefore, they must be kosher…the vast majority of GPs aren't going to go reading an academic journal about a new risk score and working out for themselves whether it's good or not, they're going to do the one that NICE tells them to do.

GP15-F-47

For tools that were regularly used, GPs were more interested in understanding and appraising the underlying evidence.

I think if they were on NICE guidelines that gives me a lot of confidence that they are rigorous, as they've been well researched…that gives me a lot of confidence that the research behind them is good quality. For things that I'm using all the time, like the QRisk, I like to have a better understanding of the population that it was studied on 'cause some might be based in America or, you know, other areas that might be less relevant and I'd have less confidence in those. So, things that are more relevant to my population, I think, it's good to be able to explain that to patients. If it's a large study, as well, that gives me more confidence, and if it's over a long time… I'd have less confidence with a retrospective study than a prospective study, and that might also, yeah, change how I'm thinking about, how much I place on that decision making to that tool. It would still form part of it, but I might not rely on it quite as much, if that makes sense.

GP1-F-35

Some GPs described a change in their practice over time and reported that they were more likely, with experience, to look into the evidence for models than they had been earlier in their career.

I think certainly when something new comes along now, I will look into it, whereas I think I've probably been very guilty, in the past, of just accepting this is the tool we use and just getting on with it. I think, as you become more used to critically appraising tools and you get more confident with them, I think there is sometimes the assumption, 'oh well if it's been publicised, somebody must have critically appraised it.' But actually, the more senior you get, you think, 'oh maybe their appraisal wasn't that great or maybe there are flaws with this and they've not been really adequately explored or they're not really adequately discussed, so that people know the limitations of them.'

GP21-F-33

As well as the guidelines and primary evidence, GPs cited role-modelling and use by secondary care colleagues, incorporation into formal clinical templates, and publicity in the medical press as being most influential in affecting the trust GPs have in models.

I suppose ones that you see specialists using and could also be useful in primary care, you'd feel a bit more confident because if your local consultants are using it, then you'd probably feel like it's a reasonably validated tool. And I guess we just moved to SystmOne… I'm presuming that most of the risk scores that have been put onto there and are therefore being used by most primary care professionals in the country are probably all fairly acceptable.

<div align="right">GP10-F-44</div>

Participants suggested that structured models and templates were likely to be used earlier in their career if such tools had been a feature of their GP training or formal medical education.

I think, for myself, like, the QRisk…it's just always been part of my training, and how I've worked. So, I think it depends, you know, if these new things are introduced later on in your career, you might be less adaptable to change because you've already got your way of doing it. Whereas, for myself, this is already part of my dialogue.

<div align="right">GP1-F-36</div>

GPs also perceived models to be useful if they are likely to save clinician time overall.

You would need to know that it works and that it saves time in the long run. And obviously that it works for patients as well, but I think for us, there's a huge pressure on our workload at the moment. And if you can tell us that this will save you time. Yes, please that would be great.

<div align="right">GP3-F-32</div>

Finally, GPs value prediction models that they perceive to be potentially medico-legally protective. By using a more objective measure of likelihood of the presence or absence or a particular condition, the outputs of models can provide justification for clinical decisions made.

In some situations you're not safe if you don't use it. So, you know, if you're determining a few things, if somebody has DVT you, sort of, have to demonstrate that you made a safe decision by using a Wells score... telling yourself that you've made a safe decision there because the risk score is not telling you that this is a high-risk situation and, also, from a point of view of documentation and medico-legal aspects.

GP18-F-34

Many GPs drew a distinction between a more intuitive approach, using clinical reasoning, and a more "scientific" approach, using a validated tool to provide back-up for clinical decisions and, again, offering some medico-legal protection.

I mean, I suppose the ones that are widely accepted just help you maybe make a decision that you're not sure about, and I suppose you've got some medico-legal back-up if you've gone one way or another and then there's a problem… probably it's just those patients that are in a grey area and you're just not quite sure, it's just useful to do it, take a step back and do it a little bit more scientifically than just doing it by feel really.

GP10-F-44

9.2.2.2. Some models have to be used

Following on from the internal motivators discussed in the previous section, GPs also discussed some of the external motivators that make them feel that a model has to be used. These included financial incentivisation [for

example, through the Quality and Outcomes Framework[24] (QOF)], Care Quality Commission (CQC) recommendations, or inclusion in practice policies and local referral pathways. GPs discussed the role of financial incentives such as QOF and discussed how these increase the utilisation of tools and models by GPs in the primary care consultation:

> *It may well be something like you have to incorporate into something like QOF, for example, some sort of tool which then means actually there's a financial incentive to do it and…the background to that is that then it's the clinical benefits to the patient as a result of that. I think to make someone do something which is very different nowadays, there has to be incentive. Just getting better patient care isn't necessarily it - I'm being a bit cynical there - but I think that's what it is at the end of the day…you try and incorporate all these additional tools and stuff into your consultations, but I think some people just need some sort of incentive.*
>
> GP11-M-40

Other GPs described how regularly reminding GPs of the evidence-based benefits and QOF implications may promote use by GPs in practice.

> *I think if there were, if it was really evidenced, cost-effective and worked for patients and QOF got a hold of it and told us that there was a financial incentive to do it, I think that would probably encourage practices to put more emphasis on encouraging their GPs to use it. And that tends to have more of an effect, doesn't it? So, the QOF points and things come up on your home screen, you're reminded about them regularly. If you're not reminded about stuff, it does tend to fly out of your brain.*
>
> GP3-F-32

---

[24] An NHSE financial incentive structure used in primary care, focussed around resourcing and rewarding good practice with respect to a number of pre-defined clinical indicators (Dixon et al., 2010)

Other GPs suggested that if a model was recommended by the CQC, they would feel an obligation to incorporate the model into practice. Being adopted as part of a practice or local policy was another factor that made GPs more likely to use a model.

> *I mean, sometimes it's a practice protocol and policy, sometimes it's QOF-related, it's just your local sort of protocols and guidelines and things as to what you follow as well as just good practice.*
>
> GP8-F-52

9.2.2.3. Barriers and enablers to using prediction models in general practice

GPs discussed the practical barriers and enablers to use in practice. Understanding these may also help us understand why some models are used while others are not. Barriers to using prediction models in practice were primarily similar to the barriers encountered earlier for relapse prevention in general practice. Time and resource were viewed as the major issues, with a prediction model being one of many competing considerations in a consultation.

> *I think a lot of this is down to time, isn't it, most appointments are still ten minutes, some practices might adopt the 15 minutes approach, but…when you manage someone with mental health issues, ten minutes, even 15 minutes isn't enough at all, and then to then open up a pop-up with lots of questions, it's going to take another five, ten minutes to do, so it's just it's time constraints more than anything else.*
>
> GP11-M-40

Time pressures within an appointment, remembering the models exist and being able to access the forms easily were also described as limiting factors in routinely employing a particular tool.

Time, so, it would be another thing that you'd have to do as part of that appointment. Awareness, so, remembering that it's there and that it would be something that would be useful, either for yourself or in terms of subsequent, kind of consultations to do…and to actually, you know, find the necessary forms

GP16-M-43

The background and context to this is the increasing complexity of the GP consultation generally, with many different demands on the relatively short amount of time available and multiple different issues requiring discussion. Some GPs described being overwhelmed with items to consider within a consultation and finding a place within this for an additional task is potentially unfeasible.

So, I rarely get a patient that comes who's only talking about one thing, and it's time, so I'm using their time and they're using my time, so they've often got four other things, so that patient who's coming to talk about depression also has a rash somewhere and is worried about mobility and is worried about this, that, and the other, and then we've got all the other QOF…their undiagnosed hypertension and do they have diabetes, et cetera.

GP9-M-45

Other challenges were identified in the way data and clinical information are coded within the GP medical record.

The problem as ever is it all comes down to coding and, you know, different people code in different ways unfortunately…different staff will think oh that's important for me to code that. So they'll code that, but they'll miss something else out and a lot of the intel you need to predict I suppose a risk of a recurrence of depression is kind of the soft value intelligence where there might not be a code for it. And that gets missed out or somebody types it in free text and it's lost forever.

GP20-M-53

Given the pressures outlined, GPs felt that the outputs and recommendations from models had to be realistic within the clinical context. Where demands arising as a result of models were overly onerous or unrealistic, GPs suggested that this was not a good use of time to use such a model.

> *It's just time, isn't it? So, it would depend how efficient it was. I suppose, you know, could it pick out information from the notes or is it something that we would need to input? Would it kind of update itself, as it were? I don't know, I think essentially, you know, we don't want to have to do extra work because there's time pressure when actually we don't really get much information from it? And I suppose, you know, what information it would deliver at the end of its assessment, as it were, do you know what I mean? You know, if it's going to say something unrealistic like you need to ring this patient every week, then no. But if it's kind of saying, right, this is the management plan that we would suggest and you think, yeah, that sounds realistic.*
>
> GP2-F-38

Finally, GPs reported that the abundance of models and pop-ups (small windows that appear in the foreground while using a clinical system) makes it challenging to prioritise and identify the most important models and tools to use.

> *I'll be honest with you, there's lots and lots and lots and lots of tools and that's the problem, we have too many tools for the same thing especially, then it just renders it meaningless in some respect. And pop-ups are the other things, they can be reminder…but too many pop-ups, it's just meaningless, you just start closing everything down.*
>
> GP11-M-40

Enablers to use were integration with the electronic clinical record through templates or other means; models that can easily be delegated to

268

other non-GP members of the primary care team or filled in by patients directly; and models that do not require additional work to utilise.

> *So it would need to be like a template that's accessible that people can find on their computer system, via Ardens or whatever, and then kind of recorded into the computer system, that's always preferable. It'd need to be really quite simple because some people are more IT-literate than others…I think it's going to be how long it takes to do, how easy it is to do, and where you find it, then having the resource to do something about it, because if you've got a great tool but actually there's nothing at the other end of it…then people aren't going to use it because they're just going to think, what's the point.*

<div align="right">GP15-F-47</div>

Some models are perceived to be useful	Some models have to be used	Barriers and enablers to use of models in the general practice consultation
• Avoid additional workload • Considered to be medicolegally protective • Guides decision-making • Produces useful output • Can be used to justify or explain decisions to patients • Aids discussion with patient • Help to motivate patients • Where benefit of use is visible to GP and/or patient • Promoted or publicised by medical press • Role modelling by specialist colleague	• Incentivised by QOF • Part of practice policy • Recommended by CQC • Inclusion in local careand referral pathways • Part of NICE or other NHS guideline • If the model is widely used	• **Barriers:** • Lack of time • Complexity of GP consultation • Unavailability of data • Incorrect coding of GP record • Not a priority in consultation (building rapport and taking history takes precedence) • **Enablers:** • Embedded within GP IT system • Automated • Patients can fill in themselves • Can be delegated • Easy to access/find and use

Figure 9.1: Summary of GP perspectives on use of prediction models in the general practice consultation

9.2.3. How prediction models are used in general practice

Finally, GPs discussed the practicalities of using models in practice, specifically dealing with missing predictor information and borderline outcomes. GPs all reported that where predictor information was missing, they would endeavour to collect the necessary information to enable predictions to be as accurate as possible. This was discussed most commonly with respect to the QRisk, but GPs felt this view applied to any model that GPs had deemed necessary and valuable to use in the first place. An exception to this was in instances where it wasn't felt safe to wait until the additional information was gathered (for example, where an acute prescription may be indicated), in which case GPs reported using clinical judgement rather than relying on incomplete predictor information.

> *I always try and get up-to-date information. So, you know, the classic thing is we haven't got their blood pressure or we haven't got their weight or their cholesterol's out of date. And I think if we're going to do it, we should do it properly really.   I also think, you know, sometimes I've done it with a cholesterol from three years ago and their risk is nine per cent. And, therefore, it's going to make a difference because obviously if it's gone up, then it will push them into that category. But if it was done, you know, and their risk score is two per cent and I kind of know that even if their cholesterol's gone up by say one or two, it's still not going to push them up.*
>
> GP2-F-38

A deviant case provided some dis-confirmatory evidence on this point and the GP reported being more comfortable with taking a pragmatic approach with prognostic models and, where predictor information is missing, reported using a "best guess".

> *Missing variables or predictors, I'm still going to use a model. I think the model, it would depend in which direction the model was pointing, so if I still wished to use the model, then take a best guess at the value*

270

of the missing variable and the output of the model, were the model complete. So, you'd take a punt at it and an educated, sort of, guess or a gut instinct, or what have you, and run with what you reckon…see what the patient wants to do and the guidelines are guidelines, if you run the numbers again, you're going to get slightly different sets of results anyway. If you do ten readings, each one is going to be a bit different, so don't get too hung up on the point one factor, you know, look at the whole numbers.

<div align="right">GP16-M-43</div>

Borderline results from prediction models (results that are very close to being either normal or abnormal) were reported as being used by GPs to initiate discussions with patients around shared management decisions. This point was illustrated in the context of using QRisk for cardiovascular risk prediction.

Yeah, so for me, if someone's borderline, I take into account what the patient thinks, and if they have strong feelings either way about starting medication. If you've got someone very motivated to make lifestyle changes, I would tend to let them try that, and then try and arrange a follow up appointment to review how that's going. If they don't feel that they can make any lifestyle changes, it's just not within their capabilities and you're trying to motivate them and it feels like it's not going well, I might be a bit more cautious and think, well actually this is only going to get worse as they get older, and things, if they can't do anything about the underlying problems, and some things you can't, you know, you can't change your age or gender, your family history and things, then I'd probably err towards sort of prescribing or at least if the patient was really reluctant or resistant to that, I would just make sure I could follow up that patient and to get that reassessed in a few years to when it might not be quite so borderline.

<div align="right">GP14-F-45</div>

In summary, where there are missing predictor values needed for prediction models, most GPs will make an effort to acquire these before using the model. Where outputs are borderline, prediction models are reported by GPs to be used in an advisory and pragmatic way to facilitate individualised shared decision-making in general practice consultations with patients.

9.3. Use of the output

9.3.1. Outputs of prognostic models

This section and the remainder of the chapter reports findings from analysis of data from interviews with GPs and people with lived experience of depression. Prognostic models can produce various kinds of outputs and risk can be quantified and presented as likelihood, probability, or a category/class membership (for example, high, medium or low risk category). They can provide instruction about what should happen next (clinical decision rules) or they can provide information to guide discussions and decision-making (decision aids). Participants discussed the kinds of outputs from models that facilitate discussions and preferences around statistics and visualisations. Some people with lived experience of depression felt that risk categories suffice and felt that people were not keen on hearing numerical information to back these up.

> *I think low, medium, and high is more than sufficient...a lot of people glaze over when you start talking about you've got 63 per cent chance of this and that sort of thing, it doesn't really mean much, I don't think. Like I say, I think saying somebody's got a high risk is better than saying you've got a 90 per cent risk… I think people have a tendency just to glaze over when they start hearing the numbers.*
>
> P7-M-48

Some of the GPs interviewed agreed that keeping outputs simple and using risk categories can avoid explanations and discussions with patients from becoming too confusing or complicated.

I think just something more simple, I think it's just like three categories really, low, medium, or high, and they understand that. We make it too many different levels, then it just gets a bit confusing and they're not really sure what to make of it.

<div align="right">GP11-M-40</div>

A variation on risk categories (e.g., high, medium and low risk) suggested by some GPs and people with lived experience of depression was a traffic light system or Red-Amber-Green (RAG) rating.

I was going to say traffic lights are always good…we need something really simple that we don't have to think about, so I think if you had like a traffic light, a red, amber, green, so green, low risk, don't have to do anything, red, need to do whatever, amber have a think about it, something like that…I think something really straightforward and simple would be better because we don't need anything else to have to think about and we're all really stressed, aren't we.

<div align="right">GP15-F-47</div>

Some people with lived experience of depression also suggested a RAG rating, although some raised the potential issue of using "red" ratings for people who are anxious and felt this might need some additional consideration.

You could have like a RAG rating, bloody RAG ratings, there's a RAG rating for everything, but they are like, they do have their place. But again, it's whether you want to create the connotation of you are red risk, red risk ohhh, which is probably not great for someone who's quite anxious.

<div align="right">P17-F-36</div>

Other GPs suggested that risk categories (e.g., high, medium and low risk of a particular outcome) are vague and explained the importance of being able to quantify risk categories if asked to by patients.

> *I think it depends on the patient, doesn't it, some people would want to know what high was, so what does that convert to, what percentage is it, so some people would want that translating into some hard data that they can then understand, whereas other people, low, medium and high is probably all the information that they want to know, and that might be good enough for them to make a decision accordingly.*
>
> GP10-F-44

Other interview participants described the benefits of having quantifiable, numerical data to back up risk categories for certain patients. They felt that having this option would be clearer for both patients and GPs and would enable an objective, informed discussion.

> *I think the challenge with that is…it's vague for me. What does high risk mean? What does low risk mean? What does medium risk mean? I don't really understand it. And maybe that's the problem. But I'm not sure patients would understand it either…For that patient that you say 25 per cent, for him that might be low risk. And for the person with 12 per cent it might be high risk. So, I think risk is very much…level of risk is dependent on personality and what a person's comfort with taking risk is…So unless there was something quite objective, or that quantified risk quite clearly, then it'd be challenging…In that way, then, actually if there was something quantifiable in terms of what does low, medium and high risk mean, I think it'd be helpful.*
>
> GP6-M-33

9.3.2. To use numbers or not?

GPs and people with lived experience of depression did not think that numerical information around risk was appropriate or necessary for all

patients. However, there was a clear view that some patients would want to be told such information in addition to less precise risk information, such as risk categories, as discussed in the previous section. It was suggested that patient preferences around such information are explored and that the level of detail provided is tailored to the individual patient. Proportions were generally viewed as preferable to percentages because they were seen as easier to understand and also more personal. Visualisations were viewed as useful by GPs and people with lived experience of depression.

> *Everybody's understanding of percentages and statistics is very varied and for some people it's easy and for some people, understandably, it's not. And it's trying to kind of pitch that. And often you can tell sometimes when you're explaining things that it's just not appropriate…I just try and bring it back to, well, if there were 10 of you, one of you could end up having a heart attack and you might be one of the nine that never have a heart attack and so you take that statin and it makes no difference, or you could be that one person that is going to have a heart attack and you take the statin and then because of taking the statin, you don't have the heart attack.*

GP2-F-38

The use of proportions when providing numerical information seemed to resonate more with GPs and people with lived experience of depression. One of the main reasons both groups of participants thought this was the case, aside from proportions generally seeming to be easier to understand than percentages, was that using proportions enabled people to see themselves as part of the explanation.

> *From a layperson's point of view, I would probably think proportions would work better, they can see themselves one in ten, so that's one person in ten, that's them, whilst ten per cent is the same, but that's just a bit more wordy, a bit more numbery, they may not fully understand it.*

GP11-M-40

People with lived experience of depression agreed with this concept and suggested that this made the explanation more tailored to the individual, reassuring and easier to visualise.

If you sort of put a number to it then, again, I suppose it's labelling it up really badly - 50 per cent might relapse. So, I think if you put it like you just said in a group of people, yeah, that would be better, for me anyway…probably because I think, and I know this sounds really silly, if you were to say you're probably this…in 40 per cent of people you'd think you're not on your own, that there's another 39 per cent stood behind you. There's 39 stood there with you.

P6-F-52

Finally, GPs and people with lived experience of depression reported finding visualisations (such as Cates plots[25]) useful.

In terms of risk, I don't find percentages particularly helpful for patients in isolation, so often we would use percentages, and I quite like the visual representations which are reflected in percentages, but they just think it much easier, 100 green faces, three of them are sad, two of them are dead, those kind of things I think are much easier to demonstrate risk than just talking through numbers.

GP17-M-35

It's obvious that people have a different way of learning, don't they…but, I must admit, I can visualise things better by actually seeing them, I mean percentages and things but sometimes, yeah, an actual diagram I know a lot of people would…that would make it a lot easier and simple for people to understand.

P13-F-58

---

[25] A diagram that uses colour-coded "smiley faces" to visually communicate the risks and benefits of a given treatment to healthcare professionals and patients (Cates, 2010).

Thus, GPs and people with lived experience of depression reported that numerical information is useful and should ideally be tailored to individual patients. Visualisations can be helpful for explaining risk to patients. People with lived experience of depression thought that information around risk should be delivered to patients by a trained professional who was able to explain, elaborate and answer questions rather than through an automated system.

> *If somebody is saying to you low risk or high risk or medium risk if they're face to face then straightaway you can say, well, what do you mean by that?   So you can get that backup and have that more information.*

<div align="right">P13-F-58</div>

GPs described how they explain the outputs of prognostic models to patients, a good illustrative example of this is included below:

> *So to explain the score, I normally say that if they, so, 'cause it's a percentage, I normally say, if there was 100 people exactly the same as you, same age, gender, blood pressure, blood results, all those things, this number out of 100 would have some sort of heart or stroke even in the next ten years. And I say, the higher that is the more likely it is that at some point that person's going to be you, out of that 100. And I say what number we tend to think of as being low risk, and what we think of as being high risk, so if the number's over 20 per cent, then, you know, then one in five, is actually quite a significant number of people who end up with heart disease, and that's why we want to do something about it, and try and reduce that risk down.*

<div align="right">GP14-F-45</div>

9.4. Prediction models for mental health problems in primary care

9.4.1. Predicting outcomes for mental health problems

So far in this chapter, I have explored barriers, enablers and perceptions of the use of prediction models in general practice. As explained, the majority of models used are for diagnostic or prognostic risk prediction for physical health conditions, with no prognostic model being routinely used for mental health presentations in primary care. This section explores the perspectives of GPs and people with lived experience of depression with respect to risk prediction for mental health problems in primary care, and then specifically of relapse risk prediction in depression. This section is the part of Chapter Nine most pertinent to the themes presented in Chapter Eight around relapse risk discussion, but with a focus specifically on risk prediction and how this might be implemented in practice.

Some people with lived experience thought that mental health presentations are more personal and individualised than physical health and therefore require a personalised approach to management, and are thus potentially more difficult to apply statistical models to:

Mental health is such a personal, personal thing, it's something that's so hard to put a statistic on or do averages and things on because everything's so personal. I don't know, I don't think it's even ever scientifically possible to put an exact number on anything to do with anything like this, so I don't think it's even worth trying.

P10-M-29

No, none of these markings or anything like that, statisticians love them, but we're talking about the mind here, we're not talking about an extra two inches on your putt or a climb or something. We're talking about reassurance. If a person feels relaxed and trusts you and the treatment that you are jointly going to go through with them… In the

environment I had, yeah, you know, you want the right equipment in the right place at the right time to work and that is to kill people. Yeah, you need to be analytical, very direct, there's no messing about. This is not, this is a soft science here.   P9-M-75

In contrast, other people with lived experience of depression found the comparison with physical health problems to be helpful and thought that applying similar models to those that already exist for physical health conditions might be comforting or reassuring to people with depression.

It almost makes it like, more digestible doesn't it? It makes it more similar to physical health conditions that people are more familiar with? Like quantifiable – almost like you can measure depression, you know? Like you can measure blood pressure or something like that. Like that might provide a source of comfort for people, rather than it being this intangible thing that's going on in your brain…It's actually, look, here's these figures about what goes on, not in a scary way, not like, one in three people get cancer, you know, not something like that, but it almost makes it less scary I think, because it's like, oh look, there's so many people that experience this, there's figures about it, and I can be one of those people, and that's normal.

P16-F-24

Some GPs had similar concerns around trying to predict outcomes in mental health conditions, compared with physical health.

So it's very difficult to explain risk, I mean, there's whole books written about it for clinicians and it's still difficult. And I think for mental health…it is even more difficult. Again, it's very patient-led, some people are ambivalent, others are keen on treatment, others are keen to avoid really, so I think in those situations, it's trying to identify what the patient wants, what kind of person they are, but also looking at their future risks, if you know that their risk is borderline now, you can often say, well, you know what, in two years, your risk will be this, so why don't we initiate

treatments now. Which again for mental health would be more difficult, because when you've got finite parameters such as age, cholesterol values, blood pressures, it's very easy for me to predict a future risk, mental health, again, much more abstract.

<div align="right">GP17-M-35</div>

It was felt by GPs that there would need to be a possibility for a greater degree of individualisation with prognostic models for mental health than for physical health.

I think depression is one of those things that everybody is affected by it so differently, that I just think that we always have to be careful with mental health tools that there is an element of individualisation and personalisation available. I think it's difficult with risk stratification tools, isn't it, because you're trying to put in…input data and have an answer. Mental health is always a bit more grey than black and white, than some other medical problems, isn't it.

<div align="right">GP21-F-33</div>

9.4.2. Predicting relapse of depression

This section brings us back to the themes discussed in Chapter Eight around discussion of relapse risk within the constraints of a primary care consultation. This section introduces the specific topic of prognostic models for relapse risk prediction and builds on the earlier sections of this chapter to explore whether there is a role for a relapse risk prediction model and how this might look in practice. Overall, GPs and people with lived experience thought that a relapse risk prediction model would be useful, as long as there were options to offer those people who were identified as high risk.

GPs discussed the feasibility of incorporating such a tool into a consultation and, in line with the discussion around barriers earlier, felt that as long as it was focussed and not overly long, it would be acceptable to ask GPs to use a prognostic model for relapse risk prediction.

You've got to be able to fit it into a ten-minute consultation, but at the same time, if it's going to save the patient a relapse and more time in the long-run, then I guess that's the thing to be thinking about. So, I think it would be fine as long as it wasn't too onerous and you could actually get through it. And usually by the time you've got to that point and if you're thinking about weaning or stopping, you know the patient fairly well, you've got that rapport, you know the background and things, so there's a lot of stuff you don't need to necessarily cover in that consultation. So, like I say, if there weren't too many questions, then I think it would be manageable.

GP8-F-52

People with lived experience of depression discussed the importance of having a plan in response to the risk prediction and that it would not be helpful to be told one is high risk without additional support being provided. According to participants, this need not necessarily be as substantial as a relapse prevention intervention but should at least include some relapse prevention planning and discussion of early warning signs of relapse.

For me, I would say definitely, yes, because I would say… To me, that would all be part of somebody saying, you know, we are looking at this holistically. We're not just, you know, bandaging you up and getting rid of you. We care about what might happen downstream…Where I would find it gets a little bit difficult is if somebody said you are high risk of relapse but then there were no actions in place to mitigate that.

P8-M-57

GPs discussed how a risk prediction model might help with using resources more effectively. Drawing on the themes from Chapter Eight with respect to the discussion around patient-initiated and proactive follow-up, GPs thought that a risk stratification tool might enable GPs to more effectively identify patients for whom different follow-up strategies are more appropriate.

I think it's always, you know, the more information that we have, the better, you can make more educated decisions. And I suppose, ultimately, we're wanting to make sure that patients are safe and, therefore, if we could flag up the ones that are more likely to relapse, and obviously by relapse they're more likely to, you know, self-harm, suicide, risky behaviour. But also, you know, from a workload point of view, if it meant that the patients that…aren't likely to relapse, maybe…maybe we could be more confident in saying, you know, we'll leave the doors open, you're in control of follow-up and we don't need to keep checking in and doing medication reviews every three to four months if actually they're stable.

GP2-F-38

GPs discussed how such a tool might be used in practice, such as having an objective measure or justification for arranging earlier or more regular follow-up reviews of patients.

I guess the advantage would be that you would have something that you could then justify saying, this person needs a regular appointment, say every three to six months to just check in on them and how they're doing and make sure that things aren't going south for them.

GP21-F-33

Some GPs worried about the implications of having conversations around relapse once a patient has recovered but felt that this would be dependent on how it was framed and explained. In line with the findings presented in Chapter Eight, the extent to which patients wanted to understand their risk could be explored. Discussions could be tailored to individuals and the information could be used to promote ongoing protective lifestyle changes and behaviours which may in turn reduce the risk of relapse.

I don't know how I'd discuss that with a patient, someone who's just recovered and feeling that they're coming out the other side, to say, actually you're at really high risk of relapse. That's quite a negative thing

to suggest to someone who's coming out, to be honest. So I think the conversation would have to be managed quite carefully, so that you can carry on encouraging people with the life changes that they've made, and maybe stating that actually this might help prevent a relapse and things, by carrying on with all these things.

GP14-F-45

People with lived experience of depression agreed with this approach and felt that sharing knowledge from individualised risk predictions would form part of educating patients fully on their condition and form part of setting realistic expectations about the course of depression.

I feel like that's the sort of knowledge I'd prefer to have than not have. Say if I've had a particularly rough…or if I'm a patient that's had a particularly bad stint of depression, and categorises the high relapse category, I'd want to know that information, just so that I'm aware it can happen and I can take that into account so that I don't get false hope to be shattered further, kind of thing. Because there's nothing worse than believing that something's going to cure all your problems and that it's just going to go away, and then it coming back immediately and you're not knowing that it could do that or why. But knowing that you're in a…you're potentially high relapse category, but if that does happen, it doesn't mean it's the end for you, there is more that can be done, there is more that can be helped, just that reassurance of, it might happen to you, it's highly likely to happen to you, but if/when it does, don't become discouraged because it can be fought, things can improve still.

P20-M-25

GPs suggested that, if a relapse risk prediction tool had been demonstrated to improve outcomes and save time in the consultation, then they would be prepared to gather additional information. This would include using the PHQ-9 and GAD-7 if necessary to contribute to an effective model.

I think possibly, I think it's always difficult because GPs are obviously very, very different in their approach to everything, let alone mental health. I'm aware that some GPs will still routinely use PHQ-9s, GAD-7s and things, within our practice not so much, but other places I've worked they've used them a lot more. I think there's good reasons to use these tools…I think if there was a good relapse prediction tool, if that requires PHQ-9s and GAD-7s, I think GPs would certainly experiment with it.

<div align="right">GP17-M-35</div>

Some GPs did however raise some potential issues around capturing the necessary information in routine practice. Alluding back to points outlined in the previous section, there was a perception from GPs that some of the information required for risk prediction in depression is "soft" and less "easily quantifiable".

There's always a place for anything that's going to help patient care, but I suppose it's how you capture a lot of that, to me it feels like quite sort of soft information. So, I mean, some of it is quite clear, so as you say, talking about patients who've had episodes before, so there's maybe a few things that are quite obvious and would be quite clear, but I guess a lot of the other things tend to be, I don't know, I suppose a conversation and then just a feeling you get for rather than something that's easily quantifiable.

<div align="right">GP10-F-44</div>

Other GPs also highlighted further implications of such a tool, including concerns around patient access to medical records and medico-legal considerations.

So, I guess, something that you could call up to calculate that score, when you're maybe reviewing people towards the end of their treatment, and to know what to do next, would be helpful, 'cause then you can base that around the discussion with the patient. I would be

concerned about that just automatically going on a health record,
particularly with more people have access to health records and things
and insurance companies and things, I'd want to have that conversation
first before assigning a number to that patient's permanent record. My
worry is, that 'cause patients have access to their records any score like
that, that's written in your records, has the potential to frighten a patient,
if they read it without being in a consultation or having it explained to
them. So I'd want it to be something that was discussed with patients,
rather than just automatically added to the patient's records, 'cause I
think that could be quite harmful.

<div align="right">GP14-F-45</div>

The medicolegal aspects referenced were raised in a way that was supportive of a risk prediction tool but also raised caution about some of the implications. First, risk stratification could be used to justify decisions and give GPs some evidence-based medicolegal justification for clinical decisions taken.

I don't want to keep coming back to medicolegal aspects because
our job is so much more than that, but if one of my patients has a stroke
on an anticoagulant and I have to attend coroners' court to explain their
treatments, it is very, very easy because I've got the data, I've got the
evidence base behind it. If the same was in mental health, it would be
lovely if we could say we had done this tool, this patient posed as
moderate risk, so this was introduced, it would be very useful.

<div align="right">GP17-M-35</div>

However, calculating individualised risk of relapse and failing to offer an appropriate plan in response was felt to have the potential to leave GPs feeling medicolegally vulnerable, which reinforces the need to any model developed to be implemented with advice around actions to be taken.

Unfortunately, with this increasingly litigious society, my mind goes
to, well, if I've got a relapse tool and either I've chosen not to use it or

even worse, I've used it, recognised a patient is at high risk of relapse, the patient then relapses and I'm seen to have recognised that but then done nothing, you wonder, well, goodness me, am I going to be in trouble. Because I suppose if you know a patient's at risk of relapse, what could you meaningfully do in terms of, most of us have appointments that are available for, what, eight weeks or so…I suppose if they're at higher risk of relapse, you might be more inclined to do a proper sort of mental health plan with them, perhaps giving them a copy of what we think their relapse triggers are, what the early signs are, what to do in that situation. So yeah, I think it could be useful, but there'd have to be some sort of caution with it I suspect.

<div align="right">GP17-M-35</div>

Finally, and again reinforcing findings from Chapter Seven, relapse was conceptually associated by GPs with consultations for antidepressant reviews and reducing and stopping medication, and many GPs made the point that they felt a relapse prediction tool would be useful at the point at which a GP and patient were making a shared decision around either reducing or continuing antidepressant medication.

If you've got something like that, then I think you could use that in the consultation to discuss risk, I think it would be really useful, so patients are sort of more aware, and whether you then think, oh, if you've got a risk of whatever, then you probably ought to be on medication, we ought not to be doing this weaning off process in the first place. No, I think that would be great, yeah.

<div align="right">GP8-F-52</div>

9.5. Summary

In summary, GPs generally reported that prognostic models and risk prediction might be potentially useful in primary care. They suggested that GPs are more likely to use models if they perceive them to be useful, and/or feel they have to use them. If neither of these criteria are met, prognostic

models are unlikely to be used in general practice. If they are met, then there are still practical barriers (e.g., time) to use that must be overcome and enablers (e.g., integration with IT and the use of routinely collected data), that should be considered in any implementation process.

The routinely-used models are all for physical health conditions and used for either diagnosis or prognosis. Risk prediction models are not yet routinely used in general practice for mental health presentation. Patients are keen to learn about their risk outcomes, as long as there are meaningful implications or outcomes from being informed about risk. Most are comfortable with being given numerical information and there appears to be value in producing numerical and other kinds of output (e.g., risk categories) and allowing the patient and GP together to decide on the most appropriate information to share to guide a discussion tailored to the individual patient. GPs felt that it was important, particularly in general practice, that prediction models are used as a guide and there must be scope for flexibility and tailoring of outcomes to individual GPs. This would fit more with the concept of a decision aid rather than decision rule.

There was more scepticism from both groups of participants about the use of prognostic models for mental health problems compared to physical health conditions. These outcomes and predictors were seen by GPs and people with lived experience of depression as more abstract and harder to measure than those used in physical health. However, some people with lived experience of depression reported that they would find an approach more in line with physical health to be reassuring. This chapter has contributed additional information that would guide the development and implementation of prognostic models for relapse risk prediction for depression in primary care going forwards.

Chapter Ten

Discussion

10.1. Overview of chapter

In this thesis, I have reported a mixed methods study. In this chapter, I will summarise the findings from each component in turn and discuss them in the context of the wider literature. I will then integrate the findings from the two components, before discussing the strengths and limitations of the study, the clinical and policy implications of the findings, and suggestions for further research. Finally, I will present my reflections on the study, the research process and the impact on my development as a researcher, before closing the thesis with my final conclusions.

10.2. Summary of findings and comparison with literature

10.2.1. Summary of findings

In order to answer the research question as to whether depression relapse can be predicted and prevented in primary care, I began this thesis by conducting a systematic review of prognostic models developed for predicting relapse or recurrence of depression. The review found a lack of pre-existing models suitable for use in primary care, due to a combination of high risk of bias in the development or validation studies, poor predictive performance of the developed models or inclusion of predictors that are not applicable to a primary care setting. I then attempted to develop a novel prognostic model, using data from a primary care setting, and including well-

established relapse predictors that would be routinely available to GPs in primary care. The model I developed had inadequate predictive performance and does not warrant further validation in its current form. I also undertook a secondary analysis including some less well-evidenced predictors. In the secondary analysis, the association between relationship status and relapse was statistically significant (on univariable and multivariable analysis), even after correcting for multiple significance testing (not being in a relationship was associated with an increased risk of relapse). The inclusion of relationship status as a predictor in the secondary analysis marginally improved the predictive performance of the multivariable prognostic model, although this improvement was probably not clinically significant.

Concurrent qualitative work, using semi-structured interviews, explored the perspectives of GPs and people with lived experience of depression around relapse risk and prevention in practice. People with lived experience of depression and GPs thought that social, personal and environmental factors were important determinants of depression course. Discussion of relapse risk was recognised to be important by both groups of participants but relapse is not routinely discussed in general practice. The terms 'relapse', 'remission' and 'recovery' themselves appear to have limited relevance to GPs or people with lived experience of depression; they do not appear to be routinely used, identified or documented.

The importance of the GP-patient relationship and communication was recognised, in particular with respect to empathy, listening and continuity of care. GPs and people with lived experience of depression reported consistent experiences in the acute phase of depression, in line with best practice. Experiences of ongoing care and support were more variable and there was a tension between GPs' view of patient-initiated follow-up being appropriate in contrast to the preference of people with lived experience of depression, on the whole, for proactive follow-up. There was a recognition that COVID-19 has adversely impacted on peoples' mental health. Restrictions arising as a consequence of the pandemic have necessitated changes to general practice operations and systems.

Building on the qualitative findings, I developed a framework to guide the ongoing care of people after an episode of depression and facilitate discussions around relapse and relapse prevention. The implementation of this framework is intended to be feasible without an added burden on GPs or their primary care teams, using the current resource available to practice. GPs and people with lived experience of depression suggested that relapse prevention could and should be effectively embedded within general practice but the key barriers to this were time and resources. GPs suggested that additional funding and resource would be required to enable implementation of relapse prevention in general practice consultations.

10.2.2. Comparison with literature

Here, I discuss the findings from the quantitative and qualitative components in the context of the wider literature. A discussion of the systematic review findings in the context of the wider literature was presented in Chapter Three, because the review directly informed the methodology and methods of the quantitative component of the study.

10.2.2.1 Quantitative study

10.2.2.1.1. Primary analysis

The primary quantitative analysis included predictors selected on the basis of good pre-existing evidence for their role as prognostic factors associated with relapse. As a reminder these were: number of previous episodes of depression; residual depressive symptoms; comorbid anxiety; and severity of depression at baseline. The predictive performance of the model was inadequate. As the systematic review (Chapter Three) found, risk prediction for relapse of depression is challenging. Even when individual predictors have strong associations with relapse, it does not follow that these can be combined to produce accurate individualised risk estimates (Riley et al., 2019b).

On univariable and multivariable analysis, the presence of residual depressive symptoms was associated with increased risk of relapse in a statistically significant way. This finding is in keeping with the prognostic factor literature, where residual symptoms are recognised to be strongly associated with risk of subsequent relapse (Buckman et al., 2018; Wojnarowski et al., 2019). Residual symptoms can be emotional (e.g., ongoing low mood, anxiety, as was measured using the PHQ-9 in this study), somatic (e.g., muscle aches, stomach aches, fatigue) or cognitive (e.g., impaired memory) (Israel, 2010). As well as increasing the short-term risk of relapse, residual symptoms are also known to be associated with a more chronic course of depression and poorer psychosocial functioning (Kennedy and Foy, 2005). For these reasons, residual symptoms (or "partial" or "incomplete remission") have long been a treatment target with antidepressant medication or psychological treatment (Paykel, 2008).

Severity of depression at baseline (last episode of depression) was also statistically significantly associated with relapse on univariable and multivariable analysis in this study. The pre-existing evidence for severity as prognostic factor for relapse was slightly more equivocal than for residual symptoms (Buckman et al., 2018), but strong enough to warrant inclusion in the primary (confirmatory) rather than secondary (exploratory) analysis. Some prior research suggests that residual symptoms are more likely in people with more severe initial depressive illness (Paykel et al., 1995) and so it might be that the presence of residual symptoms is a mediator of the relationship between severity at baseline and relapse.

Conversely, in this study, number of previous episodes and comorbid anxiety were not statistically significantly associated with relapse on univariable or multivariable analysis. Number of previous episodes of depression is recognised as a strong predictor of relapse (Wojnarowski et al., 2019; Buckman et al., 2018) and forms part of clinical guidance on identifying people who are at higher risk of relapse (NICE, 2022). The lack of association between this variable and relapse in the current study was unexpected and is not consistent with the consensus view. It is possible that

this finding occurred due to limitations in defining the predictor variable. Because IPD from different sources were combined and harmonised, this predictor was limited to a binary variable ("no previous episodes" and "one or more previous episodes"). This was due to the different ways in which this information had been captured in the primary studies. It may be that more granularity within this variable may have allowed us to predict relapse more accurately; reducing this categorical variable to a binary one may have meant a loss of information and power resulting in a non-statistically significant association with the outcome. It was also measured using self-report in all primary studies, rather than the more objective measures used for some of the other predictors, and so recall bias may have affected reliability of the measure. Finally, this was the primary analysis variable with the highest proportion of missing data (10.2%), and there was an uneven split between those with (n=824) and without (n=293) previous episodes; this may have impacted on the findings.

Similarly, comorbid anxiety has been recognised as a predictor of relapse; an umbrella review of risk factors for relapse of depression found that "a history of or current comorbid anxiety" (Buckman et al., 2018) was associated with an increased risk of relapse. In particular, higher levels of anxiety at baseline have been found to predict a shorter time to relapse after treatment, whereas post-treatment anxiety levels have not been found to be predictive (Forand and Derubeis, 2013). In this study, I used the GAD-7 at baseline (when depressed) as a measure of comorbid anxiety. As described in Chapter Five, for one of the studies (REEACT), GAD-7 was not available. The CIS-R anxiety subscale was used for this study and anxiety measures combined using standardised scores. It may be the case that an isolated measure of anxiety symptom severity at a point in time is a crude measure and less important than knowing an individual's history of comorbid anxiety. This information could be gathered in general practice and different ways of optimally capturing information about comorbid anxiety could be explored in future studies.

The presence or absence of RCT intervention was controlled for in the analysis. I used a random slope on this variable to account for the fact that different RCTs tested different interventions which may have been associated with the outcome in heterogenous ways. It was expected that these non-pharmacological acute-phase interventions were unlikely to differentially affect relapse rates between intervention and control groups. While the multilevel analysis confirmed there was significant heterogeneity between the different sources of IPD, the analysis appeared to confirm that the presence or absence of experimental acute-phase intervention within the RCTs did not affect the likelihood of relapse, when controlling for other factors.

This study provides additional evidence that residual symptoms of depression can be effectively measured and operationalised in a valid way using the PHQ-9 (score of 5-9). This has been done elsewhere with similarly highly statistically significant results (Ali et al., 2017), although it is worth reflecting that the PHQ-9 does not capture the whole range of residual symptoms (Kroenke, Spitzer and Williams, 2001). Similarly, measuring severity using the PHQ-9 would be potentially feasible and acceptable for GPs and, given the association, it is likely that this is a useful relapse predictor in general practice. Patient-reported symptom measures (both PHQ-9 and GAD-7) have been used for successful prediction of psychological treatment outcome (Bone et al., 2021) and risk of relapse (Lorimer et al., 2020) for depression and anxiety. However, the two studies mentioned used data gathered routinely through the IAPT service (now NHS Talking Therapies) rather than general practice. The utility of using patient-reported measures in principle for risk prediction in depression is supported by these studies, as well as the current study.

10.2.2.1.2. Secondary analysis

The secondary analysis identified relationship status as the only exploratory predictor with a statistically significant association with relapse. While this was a finding of the secondary rather than primary analysis, the

analysis was pre-planned and the association was statistically significant after adjusting for multiple significance testing by applying the Bonferroni correction. Marital status (being single) has been recognised as a risk factor for developing depression (Burcusa and Iacono, 2007). A recent study also identified marital status (being single or no longer married) as being associated with a worse prognosis (more depressive symptoms) at 3-4 months (but not beyond 3-4 months), after adjusting for disorder characteristics and other confounders (Buckman et al., 2021b). It is not, however, an established predictor of relapse or recurrence of depression (Burcusa and Iacono, 2007; Buckman et al., 2018). As reported in Chapter Three, some evidence from the systematic review of prognostic models pointed towards marital status ("not married") (Wang et al., 2014) and "not having a partner" (Johansson, Lundh and Bjärehed, 2015) as being worthy of investigation to further explore its role in relapse risk prediction. While weaknesses in the methodology of those studies meant that this was an exploratory rather than confirmatory analysis in my study, the further strong statistically significant association between relationship status and relapse in this study mean this is worth further confirmatory prognostic factor research going forwards. It is possible that relationship status affects relapse through providing social support although, as the qualitative findings demonstrated, the reliance on marital/relationship status as a proxy measure for social support is potentially more complex than this.

Age, gender, ethnicity, employment status and multimorbidity were not associated with relapse in a statistically significant way in this study, which is consistent with other findings from the literature (Wojnarowski et al., 2019; Buckman et al., 2020; Ali et al., 2017).

10.2.2.2. Qualitative study

Here, I will discuss some of the qualitative findings in the context of the literature. Many of the findings of the qualitative study had implications for the interpretation of the quantitative work and will be considered under mixed methods integration in the next section.

10.2.2.2.1. Social, environmental and personal factors and depression

Qualitative analysis demonstrated the perceived importance to people with lived experience of depression and GPs of social (e.g., social support, employment and socioeconomic status), environmental (e.g., childhood adversity) and personal factors (e.g., self-esteem, self-worth, coping styles and personality types) in determining depression course. The importance placed on these factors by participants contrasted with the preponderance of research focussing on clinical predictors of relapse, including depression-related factors or comorbid conditions.

Depression onset is known to be associated with a wide range of factors, which include: age (onset is often in second or third decades); gender (depression is around twice as common in females than males); history of other mental illness; history of substance misuse; family history of depression or suicide; chronic medical illness; unemployment; poor social support systems; recent stressful events involving loss; and intimate partner violence (Ferenchick, Ramanuj and Pincus, 2019). The association of many of these factors with the outcome of relapse (once depression has occurred and subsequently remitted) is less clear.

Stressful life events more commonly precede a first episode of depression and are less commonly associated with subsequent episodes (per the kindling hypothesis of relapse and recurrence, introduced in Chapter Two) (Monroe and Harkness, 2005). An IPDMA examining factors associated with depression prognosis found that reporting of major life events by patients prior to seeking help for depression did not impact on prognosis (improvement of depression at 3-4 months) after controlling for other clinical prognostic factors, demographics and social support (Buckman et al., 2022). The authors of this IPDMA were not able to conclude whether reporting stress life events was associated with a delayed time to remission. They also recommended that clinicians continue to ask about stressful life events and consider targeting specific problems (such as employment or financial

difficulties) as part of a holistic treatment approach. Social support was also associated with worse prognosis at 3-4 months (not relapse); again, this association was less strong when adjusting for other clinical prognostic factors (Buckman et al., 2021a).

The perceived importance of some of these factors for relapse, as expressed by participants in the study, is supported by the wider literature. In a high quality and comprehensive umbrella review by Buckman et al. (2018), childhood maltreatment and rumination were found to among the strongest risk factors for relapse of depression, along with residual symptoms and previous depressive episodes. Adverse childhood experiences are hypothesised to exert an effect on neurological structures responsible for processing affective information and stress responses (Nanni, Uher and Danese, 2012; Buckman et al., 2018), although more research is needed to explore the mechanism by which this association occurs. Rumination is conjectured to have a role in increasing the risk of relapse by reinforcing negative thinking patterns and dysphoric states through dwelling on negative past events and potential future consequences (Kearns et al., 2016; Michalak, Hölz and Teismann, 2011).

Neuroticism, a personality trait characterised by a tendency towards negative emotions and sensitivity to stress (Yoon, Maltby and Joormann, 2013), was among the next strongest risk factors for relapse in the review by Buckman et al. (2018). Neurocognitive factors (information processing and cognitive biases; reactions to stress or changes in mood; attentional or cognitive control) were also identified as "potential risk factors" in the same review (Buckman et al., 2018). Of note, however, a meta-analysis of studies examining predictors of relapse after CBT found that cognitive reactivity, a frequently-cited neurocognitive relapse predictor, was not associated with relapse on pooled analysis (Wojnarowski et al., 2019).

To summarise, the factors described by GPs and people with lived experience of depression as important focus more on the life situations and adverse events impacting on people with depression, as well as "internal"

factors around personality and self-concept. As discussed, the role of some of these factors in increasing the risk of relapse is less well-supported by the evidence, either because of a lack of high-quality evidence or because their role has not been investigated. Given the perceived importance of these factors by GPs and people with lived experience, it is perhaps worth refocussing and looking to explore these in primary care-based models in the future, rather than focussing exclusively on clinical and depression-related factors. The means of measuring and capturing such information in a valid way within primary care needs consideration and I will discuss this further when considering implications for future research.

People with lived experience of depression and GPs thought that the salience of environment and social factors as determinants of depression course had increased as a consequence of the COVID-19 pandemic. While the effect of COVID-19 on relapse of depression has not been explicitly studied, the perception that COVID-19 exacerbated the role of life stressors in driving depression course is certainly supported by the literature. The prevalence rates of depression, anxiety and psychological distress were higher in the general population across many countries during the pandemic than before (Lakhan, Agrawal and Sharma, 2020; Salari et al., 2020; Dozois, 2021; Bueno-Notivol et al., 2021; Pfefferbaum and North, 2020; Vindegaard and Benros, 2020). There is also evidence that those with lower socioeconomic resources and greater exposure to environmental stressors (such as job loss) were disproportionately likely to experience depression during the pandemic (Ettman et al., 2020). The role of perceived social support, as described by participants in the qualitative study, seemed to be particularly important in protecting against depression during the pandemic (Grey et al., 2020). As described by some of the GP participants, the pandemic also had an adverse impact on the wellbeing and mental health of GPs. This was due to changing working practices, increased personal and medical risk, and lack of support and pandemic preparedness, among other factors (Trivedi et al., 2021; Jefferson et al., 2022).

10.2.2.2.2. Management of people with depression in primary care

The focus of this study was the ongoing care and relapse of depression. However, I will briefly consider here the overall management of people with depression as explained to me by participants in the study. The majority of GPs and people with depression reported initial assessment, diagnosis and management that was in line with best practice and NICE guidance. The recently-published NG222 Guideline (NICE, 2022) for depression recommends that clinicians to conduct a comprehensive assessment that "does not rely simply on symptom count", assess level of functional impairment, assess risk of suicide and self-harm, and share decisions around management (NICE, 2022; Ferenchick, Ramanuj and Pincus, 2019). Almost all GPs interviewed reported doing these things in practice, although of course this study did not allow me to observe or verify what actually happens in practice. However, most people with lived experience also generally reported experiencing care in line with the NICE guideline, though there were some exceptions to this as reported in Chapter Eight. A small number of interview participants described experiences of not having felt listened to or having had the opportunity to share decisions around management. Overall, however, the qualitative findings suggest that the experience of GPs and people with lived experience is quite consistent with respect to management of the acute phase of depression[26].

Longer term, beyond the acute phase, there were differences in the accounts of management, suggesting that this became less consistent. The PHQ-9 was reportedly used by a small number of GPs interviewed, primarily for monitoring progress of depression and usually not for assessment of severity. This aligns with NG222, which suggests using validated measures

---

[26] It is worth noting that I have used NG222 as a standard in this section of the thesis, although it was published and replaced CG90 (NICE, 2009) on 29 June 2022, after some of the interviews had taken place.

for routine outcome monitoring (NICE, 2022)[27], and the pre-existing literature, which has shown that the main uses of PHQ-9 (and other validated symptom inventories in primary care) are assessing severity and monitoring the progress of treatment (Arroll et al., 2017). GPs interviewed in this study generally valued their clinical assessment and judgement over standardised and validated tools, which is also consistent with previous qualitative research (Dowrick et al., 2009). While there is some uncertainty as to the benefits of monitoring symptoms with the PHQ-9 (Shaw et al., 2013), there is some evidence that regular monitoring with PHQ-9 improves patient outcomes (Yeung et al., 2012) and that a failure to improve sufficiently on these measures does result in changes to subsequent treatment (Moore et al., 2012). It is notable that the PHQ-9 was not widely used to monitor depressive symptoms among the GPs interviewed.

NG222 makes updated recommendations around relapse prevention for people with depression who are increased risk, which include: discussing relapse, continuation or maintenance antidepression medication (with a medication review every six months), and CBT or MBCT with an explicit focus on relapse prevention (NICE, 2022). The findings from this study suggest that relapse is thought to be an important topic by both GPs and people with lived experience of depression, but that it is an under-discussed subject in routine primary care consultations. GPs lack confidence and guidance in assessing individualised risk for patients and, where it is discussed, it is often centred around seeking help if experiencing relapse rather than proactive efforts to prevent it. GPs and people with lived experience of depression were generally unaware of relapse prevention being offered in practice. The majority of people with lived experience of depression interviewed felt that there would be a benefit to having a discussion, both around relapse risk and prevention, and GPs recognised this would be a useful thing to incorporate into their consultations. Pre-existing evidence exploring the extent to which relapse risk and prevention

---

[27] The previous NICE Guideline (CG90) did suggest using the PHQ-9 to assess severity and guide management (NICE, 2009).

are discussed and provided in practice for comparison is lacking, but these findings are aligned with feedback from the PAG, where all members felt this was a subject that was both worrying and of concern to patients and also under-discussed in primary care consultations.

Antidepressant medication reviews were identified by GPs and people with lived experience of depression in this study as being an area where longer-term depression care could be improved. Both groups felt these were often superficial or did not occur, with some GPs reauthorising medications without a formal review. This finding is consistent with the literature (Leydon, Rodgers and Kendrick, 2007; Maund et al., 2019). Previous qualitative research has shown that patients view clinicians negatively when prescriptions for antidepressant medication are provided without an assessment of need for ongoing treatment (Coe et al., 2023). GPs have previously reported a lack of knowledge of guidance around when and how to stop medication and tended to continue medication, partly due to concerns around risk of relapse (Maund et al., 2019).

The qualitative findings also showed that GPs reported generally following up patients less closely when they opt for psychological treatments as opposed to medication. Relapse rates are not known to be significantly different in the early stages between these two groups and combination therapy significantly reduces the risk (Breedvelt et al., 2021b). One could argue, therefore, that the group with psychological treatments only are at a higher vulnerability of relapsing and yet, within primary care, may be waiting for a long time to receive interventions and may be left without follow-up from either provider. Furthermore, we know that waiting lists for psychological treatments as a result of the pandemic are increasing. Some GPs thought that this barrier to psychological therapies probably had increased the prescribing of antidepressants, making this a quicker and more appealing short-term strategy. While there was an initial decrease in referrals to IAPT at the onset of the COVID-19 pandemic, this was transient and waits for psychological treatments through the NHS are now longer than they were prior to COVID-19 (Larsson et al., 2022). This experience was reflected in

the qualitative data from GPs and people with lived experience of depression. This poses a challenge when considering a role for relapse prevention.

In summary, the management of new depression in primary care appears to be in line with best practice, while the longer-term follow-up and ongoing care is less consistent and less aligned with best practice.

10.2.2.2.3. GP-patient relationship, communication and continuity of care

GPs and people with lived experience of depression described the importance of relationships, communication and continuity of care. This is well supported by the pre-existing literature and, therefore, I will present only a short discussion of it here. The importance of demonstrating empathy and active listening identified in this study is consistent with previous literature highlighting the impact this can have on building trust and rapport within the therapeutic relationship (Arroll, Moir and Kendrick, 2017; Morgan et al., 2023; Johnston et al., 2007).

Similarly, continuity of care is recognised as an essential feature of the GP-patient relationship. The Royal College of General Practitioners' policy is to promote continuity of care in general practice and they have published guidance to help practice improve this (Royal College of General Practitioners, 2023; Hill and Freeman, 2011; Royal College of General Practitioners, 2016, 2019). In particular, and in light of the barriers and challenges described, this is prioritised for those patients who will benefit from it most (patients with multimorbidities; older people; people with mental health difficulties; and patients receiving palliative care) (Jeffers and Baker, 2016). A recent health and social care committee report similarly highlighted the importance of continuity in general practice for reducing acute pressures on the overall health system, improving outcomes for patients and for improving professional satisfaction for GPs (House of Commons Health and Social Care Committee, 2022).

The kind of continuity referred to by people with lived experience of depression and GPs in the qualitative study is known as relationship, or relational, continuity. This describes a form of continuity that is rarely formally recorded but is based on accumulated knowledge of patients' circumstances and mutual trust acquired through familiarity between GPs and their patients. The other forms of continuity are informational continuity (formally recorded or working knowledge of patients' values, preferences and context) and management continuity (of shared management plans and follow-up) (Guthrie et al., 2008). Informational and management continuity are supported by increasing standardisation of care through the use of clinical guidelines, care pathways and electronic health records (Guthrie et al., 2008).

There is evidence that relational continuity reduces hospital admissions (Barker, Steventon and Deeny, 2017) and mortality (Gray et al., 2018), increases patient satisfaction (Fan et al., 2005) and is valued by GPs (Nowak et al., 2021). Barriers to being able to provide relational continuity in practice are primarily the lack of GPs and funding into general practice (Jeffers and Baker, 2016). Relationships between GPs and patients have been found to be the main reason GPs draw meaning and value from their work, rather than technical aspects of clinical diagnosis and management (Fairhurst and May, 2006). While valuing continuity, some GPs report aspiring to deliver this through teams rather than being able to guarantee personal continuity; it is less clear whether this approach offers the same benefits as personal, relational continuity (Ridd, Shaw and Salisbury, 2006). There is less pre-existing research around continuity of care and its specific role in depression management. Studies that have explored this, however, do suggest benefits of continuity of care to people with depression (Udo et al., 2019; Uijen et al., 2014).

The findings of this study reinforce the importance of relational continuity, and have shown that it is an important consideration in the management of people with depression. The study also reinforces the

importance of an empathic approach from GPs and the value of the therapeutic consultation in the context of depression.

10.2.2.2.4. GP consultations are a limited resource

The findings from this study showed that GPs and people with lived experience of depression perceive GPs and primary care services generally to be under pressure and lacking in time and resources. There is reason to think that this is set to be become an increasingly significant problem. Increasing numbers of commentators make reference to dual workload and workforce crises in general practice (Gopal and Mulla, 2020; Jefferson and Holmes, 2022), which were exerting pressure before the onset of the COVID-19 pandemic, but have now been exacerbated and do not show signs of improving. As of November 2022, data from the British Medical Association (BMA) found that, compared to September 2015, there are 1973 fewer fully qualified GPs looking after, on average, an additional 2206 more patients per practice (put another way, 0.44 fully qualified GPs per 1000 patients compared with 0.52 in 2015). The numbers of GPs being recruited do not come close to the levels needed to reverse this trend and to meet the escalating demand; 31.3 million appointments were booked nationally in November 2022 alone (more than half the population of England) (British Medical Association, 2022). Any recommendations to improve care must be made in the context of these pressures on services.

10.2.2.2.5. Impact of COVID-19

The qualitative findings also illustrated the way COVID-19 is perceived to have impacted on general practice. GPs reported an increased focus on triage and remote consulting, whereas people with lived experience noted it being harder to receive an in-person consultation when needed. The implementation and future of remote consulting is an area of active research and not the primary focus on the thesis so I will discuss this only briefly here. The perception that primary care delivery has been transformed as a consequence of the COVID-19 pandemic is supported by evidence. Routine

care was postponed as resources were directed towards acute illness and triage, with a resulting backlog of routine care for chronic conditions (Wanat et al., 2021). Practices generally moved away from the pre-pandemic "open door policy" and remote care was delivered for many conditions without training in remote consulting and to the detriment of certain patient groups (for example, older people) (Wanat et al., 2021; Rawaf et al., 2020). There were also perceived advantages to this approach, however, including convenience to patients and protection against infectious disease (Verma and Kerrison, 2022). There is also evidence that patients were broadly satisfied with remote consulting, except for when discussing complex or sensitive conditions (Anderson et al., 2021). It is likely that the model of triage and remote consulting will persist in some form, but that it needs adjusting to decrease clinical risk and to accommodate patient preferences (Murphy et al., 2021).

10.3. Mixed methods integration

10.3.1. Integration of findings

I outlined the convergent approach to mixed methods and integrating findings taken in this study in Chapter Four. Having discussed the findings from each of the components in turn, I will integrate the findings in this section. I have used a process of triangulation to compare and contrast findings between the two workstreams and to produce a deeper interpretation and understanding of the problem than would have been possible using just one methodological approach (Creswell and Plano Clarke, 2011; O'Cathain, Murphy and Nicholl, 2010). In particular, the qualitative findings have implications for, and provide a lens through which to interpret, the results of the quantitative workstream.

Principally, the qualitative findings may offer some explanation as to why the prognostic model did not predict relapse with sufficient accuracy. Two conceptual challenges were raised in the findings from the qualitative

sub-study: first, the validity of a diagnosis of depression in primary care (and how this is differentiated from associated constructs, primarily emotional distress); and second, the validity and applicability in primary care of the constructs of relapse, remission and recovery. The complexities of general practice and undifferentiated case-load of patients may make it difficult to apply these terms in a straightforward and unambiguous way.

The qualitative findings also demonstrated the perceived importance of personal, social and environmental factors in determining the course of depression and outcomes. While some of these were included in the secondary analysis in the quantitative study (employment and relationship status), they were not included in the primary analysis. Relationship status in particular seems to have some statistical value as a prognostic factor, which is aligned with the qualitative findings. A further consideration is that the qualitative analysis demonstrates some of the ways in which we measure personal, social and environmental factors are not necessarily congruent with what we are trying to capture. GPs and people with lived experience of depression both perceived problems with the measurement and recording of certain predictors. For example, employment status can be used as a proxy for a social connectedness and security but, as some participants explained, particularly following a rise in remote working as a result of the pandemic, perhaps this is not as valid as it once was. Similarly, with respect to relationship or marital status, the point was made several times that this is dependent on the quality of the relationship and whether it is a source of support or a source of stress. Thought needs to be given around how best to capture more meaningful data around prognostic factors.

Finally, the use of mixed methods in this study has provided additional meaning and context to the problem of depression relapse in primary care. It was clear that, regardless of the current accuracy of statistical models, relapse is an area that is recognised as important but for which current care is lacking. The mixed methods integration has informed the recommendations for clinical practice, policy and future research that follow in section 10.5 of this chapter.

10.3.2. Comparison of integrated findings with literature

10.3.2.1. Depression and distress in primary care

Chapter Two outlined the diagnostic classifications and criteria (DSM-5 and ICD-11) used for diagnosing depression. Emotional distress, which can be an expected response to life stressors, as opposed to depression, is a less well-defined construct and is usually down to clinical judgement to discern. The literature sometimes refers to the distinction between disorder (depression) and non-disorder (distress) (Wakefield, 2007; Geraghty et al., 2015). In this discussion, I will distinguish between depressive disorder (for depression meeting standard diagnostic criteria) and emotional distress (for presentations that would not meet the threshold for a formal diagnosis).

A detailed discussion of the distinction between emotional/psychological distress and depression is beyond the scope of this thesis and has been discussed in detail in the literature. Previous qualitative research has explored this; one study presented a thematic analysis of interviews with 21 UK GPs (Geraghty et al., 2019). In line with the findings from this study, depression and emotional distress were thought to be on a continuum and difficult to distinguish. This study found that GPs thought emotional distress was more likely in the presence of a stressor and in the absence of biological symptoms, but that they would also use the terms "reactive depression" and "endogenous depression" to draw a similar distinction. This distinction is also made in the literature; depression with an obvious environmental cause has been termed reactive depression, whereas depression where the onset is not obviously provoked by an external cause is sometimes termed "endogenous" (Clarke et al., 2008). The DSM-5 now refer to the former as "adjustment disorder, with or without depressed mood" (American Psychiatric Association, 2013). What is being referenced in the findings of the qualitative study is that primary care providers identify and manage social and personal problems that, while not necessarily biomedical problems, do cause morbidity and may be labelled or thought of as depression (Gask et al., 2008). The avoidance of a depressive disorder

diagnosis for all people who have a proximate trigger in the history is not likely to be a helpful approach, particularly given that we know that people with a history of depression often have a lower threshold for developing a recurrence of their depressive disorder in response to a trigger due to, among other theories, the kindling hypothesis (Burcusa and Iacono, 2007).

A study of the Four-Dimensional Symptom Questionnaire (4DSQ) in primary care found that when this measure (which specifically focuses of distress dimensions, as well as core depressive symptomatology) was used, it decreased the number of people diagnosed with depressive disorder. This suggested that depression may be over-diagnosed in primary care. GPs often find it challenging to be confident in making a distinction between the two (Geraghty et al., 2015, 2019). In particular, distress can be easily misdiagnosed as depression when standardised diagnostic instruments are applied without holistic contextualisation of the patient's overall presentation. It may be that depression is sometimes diagnosed in favour emotional distress as GPs are more familiar with the treatment options for depression, whereas distress is perceived as less of a clear diagnosis with less clear management options (Mendive, 2009). However, research also suggests that patient satisfaction would be improved in emotional distress were more commonly recognised and discussed by GPs (Gross et al., 2007).

While this discussion outlines some of the complexities of valid diagnosies and constructs, it is useful for a range of reasons (research and the standardisation of clinical care, for example) to be able to identify clinically significant conditions in a consistent manner (Arroll, Moir and Kendrick, 2017). Previous literature has reflected the challenge in applying standardised diagnostic criteria and definitions of depression, which are often based on secondary care research, to a primary care population (Gask et al., 2008). In primary care, patients are usually undifferentiated (meaning that they present without prior assessment), have fluctuating symptoms and present with a range physical, psychological and social problems. Indeed, some researchers have suggested that depression may require a new classification system and taxonomy for primary care (Gask et al., 2008). For

our purposes, it is important to reflect on the challenges posed to the construct of depression underlying the study and to interpret the findings in light of this.

10.3.2.2. Conceptualisation of relapse, remission and recovery

The qualitative findings also raised a challenge to the constructs of relapse, remission and recovery, which underly the quantitative study. These terms (along with recurrence and response) are sometimes referred to as the 5 Rs, or change-points, and were first properly defined and operationalised by Frank et al. (1991). The findings revealed that these change-points are not consistently used in practice and many participants questioned their relevance and usefulness in general practice.

The 5 Rs presented by Frank et al. (1991) were conceptual definitions with associated operational criteria and were intended to be empirically validated. I described in Chapter Two the lack of an empirically-derived temporal cut-off for these different change points (Frank et al., 1991; Rush et al., 2006) and that the terms remain inconsistently operationalised in the literature (Beshai et al., 2011; Bockting et al., 2015). A recent systematic review explored the various efforts at empirically validating the change-points (De Zwart, Jeronimus and De Jonge, 2019). The review concluded that remission is best defined by decreasing the symptom severity cut-off (the review suggested a Hamilton Rating Scale for Depression 17-item version (HAMD-17) ≤ 4 instead of the established ≤ 7), but that a duration criterion is less important for defining remission. Specified duration thresholds of symptomatic improvement to distinguish remission from recovery (and therefore relapse from recurrence) are not meaningful. This distinction is also different in the most commonly used diagnostic manuals (DSM-5 and ICD-11). This suggests that distinguishing relapse from recurrence is difficult and possibly of limited clinical importance.

The terms are used extensively in the psychological literature, but there is no previous literature describing how these terms are used by GPs

and patients in a general practice setting. The findings from this study illustrate that, while these terms are understood by GPs, they are not generally used clinically to describe depression course and are not measured in a consistent way. In primary care, it appears that they are generally assessed in a more subjective way, according to what the patient describes as "sufficient improvement" or return to "usual" functioning. The original change-points and criteria (Frank et al., 1991) did not include reference to antidepressant medication and relapse as described in the literature is not dependent on the presence or absence of any particular treatment. It is therefore notable that GPs associated the concept of relapse and stopping antidepressants so closely.

The accepted goal of treatment for depression has long been remission, although there is less consensus on what the specific indicators of remission should be. Given that we lack biomarkers of relapse and remission, Keller (2003) made the point that remission of depression is identified by changes in the number and severity of symptoms, which are in turn based on descriptive phenomena rather than results of specific tests. Both groups of participants in my study also thought that features like reduced functional impairment and feeling like their "usual selves" were more important than clinically-driven measures of improvement. This finding was consistent with findings from a previous study exploring patients' views on how remission is most appropriately defined (Zimmerman et al., 2006). Patients in that study thought that the most important features necessary for remission to be identified were the presence of features of positive mental health (specifically "optimism, vigour and self-confidence"), a return to one's usual normal self, and a return to usual level of functioning. The group surveyed in the study were broadly representative in terms of age and gender, although notably were all psychiatric outpatients recruited from a private practice in the US and so the results are not necessarily transferrable to a primary care population. It seems an absence of depressive symptoms, measured using a diagnostic standard, is not the only criterion of significance to people with depression or GPs.

Research on remission has highlighted that improvement in symptoms and symptoms checklist score often overlook aspects of remission that are of more importance to patients, such as quality of life and daily function (Zimmerman et al., 2012, 2008). Daily function and quality of life are not part of the standard definition of remission and the task force advised that these should be measured as secondary outcomes and symptom ratings were preferred as a primary outcome (Rush et al., 2006). Health-related quality of life is an important outcome for people with depression and a recent study found that health-related quality of life returns to that of the general population when remission is diagnosed using conventional clinical cut-offs (Riihimäki et al., 2023). Research exploring the link between remission and full functional recovery is missing (McIntyre, Lee and Mansur, 2015), but is clearly prioritised by GPs and people with lived experience of depression.

Many of the participants interviewed thought that terms such as relapse and recovery, which imply an episodic, discrete disorder, did not resonate with their experience of long-term depressive symptoms. Given that more than half of people with depression will have a chronic or fluctuating course (Ramanuj, Ferenchick and Pincus, 2019), there has long been an argument in favour of applying the chronic disease model to depression care (Tylee and Walters, 2007). This view further raises the potential need for a reframing of these terms in the context of depression in primary care. A further issue is that the conceptualisation of the 5 Rs is focussed towards formal diagnosis of MDD, per the diagnostic standard manuals. However, as the previous section outlined, and my findings reinforced, MDD is not necessarily diagnosable in some people who are managed with emotional symptoms in primary care.

Technically, based on the original descriptions, the identification of any of the 5 Rs would be dependent on a diagnosis of a depressive disorder rather than emotional distress. However, as the findings from this study and previous work have demonstrated, this boundary is not clear-cut. It might be helpful, when taking work forward looking to prevent relapse in depression, to consider this broader and more inclusive approach. It might be that phrases

such as "staying well after an episode of emotional distress" or "reducing the risk of becoming unwell again" are more helpful than talking in terms of depressive relapse. Segal et al. (2003) suggested some additional terms (for example, partial response, partial remission, durable recovery and continuous measures of residual impairment) to account for some of the conceptual issues posed by the initial descriptions (Segal, Pearson and Thase, 2003). It does not seem likely that these would solve the issue of relevance and usefulness in primary care. This section has discussed further challenges to some of the constructs underlying the study which, again, should be borne in mind when interpreting the findings and determining a course for future enquiry.

10.3.2.3. Prognostic models in primary care and communication of risk

The qualitative findings presented in Chapter Nine were directly related to the prognostic model development work presented in Chapter Six. Qualitative work alongside prognostic model development is recommended with patients and providers to guide development and subsequent implementation (Hoesseini A, van Leeuwen N and Sewnaik A, 2022). The purpose was to explore the acceptability and feasibility of implementing a relapse risk prediction model in practice, and to understand preferences of people with lived experience of depression and GPs around communication of results.

There have been a small number of previous qualitative studies exploring clinician perspectives of prognostic models and none exploring patient perspectives. Only one qualitative study explored the views of primary care providers, in the context of cardiovascular risk prediction. Lack of consistency with a holistic approach to patient care, questions around validity, concerns about increased workload and whether such models add value to the clinical assessment were raised as concerns by the clinicians interviewed (Takamine et al., 2021).

One study explored the views of mental health clinicians and healthcare administrators on prognostic models in the context of suicide risk prediction (Yarborough et al., 2022). In that study, clinicians shared the concerns of participants in my study that they would not have the clinical resources to adequately meet the potentially increased demand, as well as having medicolegal concerns. Medicolegal concerns were the same as for my study; once a clinician is aware of risk information, they feel they would be personally liable for not acting, or not acting quickly enough. Interviewees in the study by Yarborough et al. (2022) also had a preference for integration of the risk model with routinely-used clinical systems. Interviewees also discussed the importance of transparency and explanatory value of models for suicide risk prediction which, interestingly, is not something that was mentioned as being important by interviewees in my study.

As in my study, facilitating discussion with patients around prognosis and guiding patient outcomes were viewed by emergency department clinicians as key benefits in the qualitative evaluation of a prediction model for patients with blunt chest wall trauma.(O'Neill et al., 2020). Similar findings were reported in a qualitative study exploring clinicians' perspectives around the use of prognostic models in end-of-life care settings – facilitating communication was seen as one of the key advantages (Hallen et al., 2015). Only one pre-existing study identified explored patient perspectives of prognostic models; this was in the context of communicating prognosis to people with head and neck cancer (Hoesseini et al., 2020). Patients in this study were broadly accepting of the use of prognostic models and many of the patients interviewed were keen to understand the quantitative results. It is worth reflecting that this latter study was situated in a very specific context and the results from the study do not translate easily to a general practice mental health setting.

The study presented in this thesis also explored the communication of prediction models and risk with patients. A scoping review looked at the evidence around communication of prognostic model results by healthcare providers to patients (Walsh et al., 2021). The review reported a lack of

312

evidence on the best ways to communicate prognostic model results. It did, however, highlight the importance of contextualising information when discussing with patients. This contextualisation specifically referred to defining timeframes over which predictions are applicable, using results in the context of shared decision making, and using data visualisation as an explanatory tool (particularly for those with lower health literacy). The findings from my study were aligned with the review by Walsh et al. (2021) and pointed towards a need for a range of explanatory options to be tailored to the individual patient and their preferences.

Previous literature has explored the communication of risk in practice. The findings from this qualitative study suggest that communication of risk benefits from being individualised to the clinician and patient in question. Risk categories can be useful and meaningful but patients often want to understand details of the numbers involved and having these available is likely to be helpful in guiding a successful discussion in practice. The findings suggested that, for some people, numerical information is desired and that proportions are slightly preferrable to percentages when communicating risk to patients, but that these are dependent on patient preferences and level of numeracy. This is supported by research on the subject of risk communication which has found that either percentages, frequencies or proportions are acceptable for conveying numerical risk to patients (Spiegelhalter, 2017). Research by Spiegelhalter (2017) also supports other findings from this study, including the use of visualisations in multiple formats to cater for different patient preferences (illuminated with words and numbers where possible), and the importance of assuming low numeracy and providing optional additional detail as desired (rather than providing more complex information to start).

A notable finding of this study was that, for some people with lived experience of depression and GPs, the idea of statistical risk prediction for depression and other mental health presentations did not feel as appropriate as for physical health presentations. Mental health presentations were seen by some as more "individual", depending on "softer" information and

therefore harder to predict outcomes from using purely statistical means. Meehl would challenge this view and has outlined the benefits generally of statistical (or actuarial) over clinical prediction (Meehl, 1954, 1986). This study has provided some evidence that it may be the complexity of the constructs we are measuring and our current measurement tools that are contributing to the difficulties in achieving highly accurate individualised risk predictions. The earlier sections of this chapter expanded on some of these conceptual challenges. It is likely that outcomes can be predicted more accurately with more precise data which truly captures what we want to capture in terms of both predictors and outcomes, and this should drive future research.

In summary, some prognostic models are generally viewed as useful by GPs and patients. GPs and people with lived experience of depression in this study thought that a model for predicting relapse of depression in primary care would be of value and feasible for implementation in primary care. Prognostic models must have face validity and must be perceived as valuable by the clinicians and patients using them in practice. It is likely that any prognostic model developed for relapse risk prediction in the future would need to be used in conjunction with a holistic clinical assessment to guide decisions (decision aid) rather than dictate them (decision rule). I will discuss ideas for taking this forward when discussing implications for future research, later in this chapter.

10.4. Strengths and Limitations

A discussion of the strengths and limitations of the systematic review was included in Chapter Three. The role of the PAG and PPIE running throughout this study (discussed at length elsewhere in this thesis) has also been a major strength of the study. It ensured that the approach taken was meaningful to people with lived experience and improved the design of the research and the validity of the findings and conclusions. Here, I will discuss

the strengths and limitations of the quantitative and qualitative studies, and the mixed methods approach taken.

10.4.1. Quantitative study

The quantitative study was conducted according to best practice recommendations for methods (PROGRESS framework) and reporting (TRIPOD-Cluster). I made the decision to use a "full model" approach, as advocated by Harrell (2015), to ensure that the model had face validity, explanatory power and was acceptable to GPs and people with depression. I used an up-to-date review of the extant literature to guide predictor selection and the sample size used was in excess, relative to the number of predictor parameters, of those used in previous prognostic model studies. By pre-selecting the predictors with a robust, pre-existing evidence base, one removes the risk of overfitting associated with more data-driven approaches for predictor selection. A downside of this approach is that the exploratory analyses that would be possible with a data-driven approach were less feasible within the primary analysis, although as the secondary analyses showed it is unlikely in this study that this would have resulted in a model with significantly increased predictive performance. Many prognostic models are unstable (i.e., likely to be different if developed again in a different sample of the same size from the same population) (Riley et al., 2019b). In this study, I took steps to mitigate for instability by, first, ensuring a sufficient sample size for the number of predictors and number of events; second, adjustment for optimism using bootstrapping and the application of a uniform shrinkage factor; and, finally, using IECV to assess generalisabilty and explore stability.

The number of previous episodes has strong *a priori* evidence for its role as a predictor of relapse (Buckman et al., 2018) and, in this study, it was non-significantly associated with relapse on both univariable and multivariable analysis. This finding is not consistent with the literature and may be explained by the fact that this was operationalised as a binary

315

variable (no previous episodes or one or more previous episodes). This categorisation was necessitated as a best attempt to harmonise the IPD from the original datasets and may have resulted in the variable losing explanatory power in the model development. Similarly, for pragmatic reasons, I made decisions around how best to harmonise data for the exploratory predictors. Moving forward, some of these (in particular, relationship status) have some evidence that would mean they should be investigated further for their prognostic value. For the purposes of the quantitative workstream, several decisions needed to be made both for pragmatic purposes related to how the original data were collected and harmonised, based on justifications from the literature and discussions between me and my supervisory team.

The ideal dataset for developing a prognostic model is a prospective, pre-designed cohort study. The advantage of such an approach is that investigators retain control over inclusion and exclusion criteria, definition and measurement of predictors and outcomes, ensure appropriate timings, reduce missing data and minimize other potential biases (for example. selecting bias and blinding). However, the costs (financial and time) of carrying out a prospective study would be substantial and secondary analysis of good quality data from RCT and other cohorts is an accepted alternative (Pajouheshnia et al., 2019). Problems arising from the use of IPD from RCTs were missing data and the risk that predictors and outcomes may not have been recorded optimally during data collection. A further common pitfall of RCTs is the narrow eligibility criteria often stipulated which can impact on the generalizability of any findings to the target population of interest (in our case, a primary care patient population). I was reassured that the eligibility criteria for the included studies were inclusive and pragmatic with relatively small numbers of participants with missing data. I do however recognise that RCT participants may differ from the general population in important ways and results should be interpreted with this in mind. In the planning stage of this project, I had considered other data sources, in particular the Clinical Practice Research Datalink (CPRD), a large electronic database of routinely collected follow-up data from primary care. Following discussions with CPRD experts at the University of York, it was evident that the coding of measures

of relapse and recurrence were not optimal for identifying patients who relapsed and that this would have limited my ability to develop a reliable and generalisable model.

Further limitations relate to measurement of start-point (remission) and end-point (relapse or not), which were measured using PHQ-9 score. The gold standard would have been to use diagnostic interviews, which may have been possible with a prospective cohort study, although the PHQ-9 is a validated and widely-used tool with good sensitivity and specificity (Kroenke, Spitzer and Williams, 2001). A further point to consider is that the start- and end-points were defined at the next time-point they were actually measured rather than necessarily capturing the precise "real-world" moment of remission/relapse. However, this reflects the situation in general practice, where diagnostic tools will be applied at patient consultation rather than in real time. Therefore, I think this was justifiable and actually mirrors the clinical picture accurately.

I modelled RCT intervention as a predictor variable rather than make the assumption that the effects of the different interventions and controls in the RCTs were homogenous. It is not likely that the interventions had a significant effect on relapse rates, even where they did improve acute depression symptoms. However, it is possible that one or more of the interventions (or controls) did exert an effect on relapse of which we are not aware. A further limitation was that the data I analysed did not allow for survival analysis, as the follow-up time-points were insufficiently similar and infrequent. Time to relapse is important and would increase our understanding; future prospective work should consider this when designing strategies for data collection.

There are some predictors not included due to lack of relevance and usefulness to GPs. For example, neuroticism (the personality trait), childhood maltreatment and rumination have been found to be associated with increased risk of relapse and recurrence (Buckman et al., 2018), as has duration of index episode of depression and age at onset of first episode of

depression (Berlanga et al., 1999). These are not routinely measured in practice and have not been coded for in our cohorts and were therefore not included as predictors. The cohort has been designed to be as undifferentiated as possible to represent a GP case-mix. Increased predictive performance would be more likely if we were to be very specific in defining this cohort, but this would have implications for its utility in the real-world primary care setting.

There was a risk of selection bias given the way studies were selected for inclusion in the IPD meta-analysis. The studies were not selected as a result of a comprehensive literature search and the aim was not to present a summary statistic drawn from all of the available evidence. To mitigate the risk of selection bias, I did take a systematic approach (outlined in the methods) to identify studies, which involved a search of the NIHR trials registry and the use of a recent IPD MA that had used a comprehensive literature search, as well as the acquisition of data from trials conducted in the University of York. A strength of this process was that I was able to obtain IPD for all of those studies identified through this pragmatic approach, which aimed to combine data from well-conducted studies to provide a sample size with sufficient power to allow the development of an accurate and stable model. The intention was not to provide an overview of the literature of depression treatment in primary care and therefore, while the process of identifying studies was systematic, it was not dependent on a full literature search and screening. A final limitation to note, with respect to the IECV, is that the total number of events in most individual clusters was lower than that generally required for external validation (Collins et al., 2015), and so it is difficult to evaluate model performance at the level of these studies with certainty. The IECV did, however, provide a helpful assessment of generalisability of the model.

Some of the limitations described in this section will guide the recommendations for future prognosis research outlined later in the chapter.

10.4.2. Qualitative study

A strength of the qualitative study was a good number of subjects covering a broad range of participants in both groups. The GP group included GPs from a range of backgrounds, settings and years of experience as GPs. The sample of people with lived experience had fewer people from ethnic minorities than I aimed to achieve, despite efforts to increase the number of people from ethnic minority backgrounds by purposively approaching practices with higher ethnic diversity in their patient populations through the NIHR CRN.

The inclusion of two groups of participants (GPs and people with lived experience of depression) meant that I was able to contrast and compare experiences and perspectives to allow for a deeper analysis and understanding of the issues. There were some sub-themes where data from only one group of participants was used. An example of this were the findings around differences in follow-up for people on antidepressant medication compared with those having psychological therapies. While the data were analysed concurrently and iteratively, this was an example of a sub-theme that was generated after data collection and therefore could not be corroborated through interviews with people with lived experience of depression.

Two lay participants in the qualitative study were recruited through study advertisement rather than through GP database search. This meant that these two participants did not necessarily have a formally-coded diagnosis of depression on their clinical record. The history was explored and verified by me as part of the qualitative interview, using the submitted semi-structured topic guide. The results should be interpreted with this in mind. I think the benefits of this approach (recruiting from a wider pool of participants and approaching those who may have engaged less with health services, including harder-to-reach participants) outweighs any risk to the integrity of the study or its ability to address the aims and objectives.

Another key strength of this study was the involvement of the PAG, who were involved in developing lay participant-facing recruitment materials, pilot interviews and development of the semi-structured topic guide, in interpreting the findings. I am confident that this has made the findings and implications of the research relevant and meaningful to people with depression and GPs. Credibility of the findings was ensured by sharing findings with clinical research colleagues, the PAG, and also fellow primary care researchers at the Society for Academic Primary Care North conference in November 2022.

Due to COVID-19, all of the interviews were conducted remotely using Microsoft Teams software or, in two cases, telephone. This was not the original plan, when writing the funding application. Evidence suggests that while there are some minor advantages to in-person interviews (in particular, participants generally say more words in in-person interviews), the difference is modest and remote interviews are an acceptable alternative to in-person interviews for qualitative health research (Cachia and Millward, 2011; Johnson, Scheitle and Ecklund, 2021; Krouwel, Jolly and Greenfield, 2019). An advantage of using this approach was that I was able to recruit from a broader area.

This is the first qualitative study to my knowledge to explore the perspectives and experiences of GPs and people with lived experience of depression around relapse discussion and prevention in practice. It is also the first study to explore the views of primary care patients and GPs around the use of prediction models for mental health problems.

10.4.3. Mixed methods approach

The convergent mixed methods approach used in the study allowed the interpretation of findings to benefit from the strengths of two distinct methodological approaches. The quantitative component (and systematic review) allowed me to explore statistical associations of predictors and the

outcome of relapse to understand the role of different variables and the extent to which these can be combined to provide individualised risk estimates. The qualitative component allowed me to better understand peoples' experiences and perspectives within the real-world context of the study problem. As discussed, the qualitative findings raised some conceptual challenges to some of the assumptions made when designing the quantitative methods and, as a result of this work, we are in a stronger position to answer the "what next?" question. Triangulating the qualitative findings with quantitative findings has contributed to ensuring the conclusions of the study are robust and valid. I am certain that my understanding and ideas around future work have been enriched by the integration of the two sets of findings.

10.5. Implications for clinical practice and health policy

10.5.1. Implications for clinical practice and education

This study has shown that predicting individualised risk of relapse for people with depression is challenging and we are not yet able to reliably do so. It is possible that accuracy of prediction might be improved with further work. Until we can more accurately predict outcomes for individual patients, there is some evidence-based guidance available to guide clinicians in practice. The updated guidance from NICE (NG222) offers recommendations for identifying individuals who may be at higher risk of relapse (NICE, 2022). These include people with: a history of recurrent depression; a history of incomplete response to treatment; unhelpful coping styles (e.g., rumination and avoidance); a history of severe depression; other chronic physical or mental health problems; and personal, social and environmental factors that contributed to their depression and are still present.

This study has found that people with depression would value more discussion of their risk of relapse as part of their routine depression care, including some discussion of early warning signs and how to seek help if

needed. GPs, on the whole, said that such a discussion does not form part of their routine practice. On the basis of this study, primary care clinicians are advised to explore with patients the extent to which they wish to have a discussion around relapse and what the implications of a discussion would be. Clinicians should be aware that this can be a worrying and upsetting subject of conversation for people who have just improved following depression. This study has shown that this discussion would be most appropriate after improvement following the acute phase, and should be sensitive and guided by the patient. People with lived experience of depression said they would most appreciate a discussion of relapse where there were constructive implications, such as a recommendation around and availability of relapse prevention treatments or advice around self-monitoring and management. Moving forwards, there would be benefit to ensuring that GPs were more aware of risk factors for relapse and how to access relapse prevention for patients where indicated.

NG222 goes on to make some updated recommendations around relapse prevention. For people who are at higher risk of relapse according to the above criteria, the guidance suggests continuing the same treatment that achieved remission; either maintenance antidepressant medication at the same dose; the same psychological therapy, but adapted for relapse prevention; or one or both components of combination therapy. The guidance suggests switching to group cognitive behavioural therapy (CBT) or mindfulness based cognitive therapy (MBCT) for those who wish to stop medication. A review is recommended every six months for those on medication and when finishing relapse prevention psychological therapy. This study has found that the actual provision of psychological therapy with a relapse prevention focus is less consistent that the NICE guidance might suppose. Given the pressures on services, particularly exacerbated by the pandemic, clinicians should be aware of this reality and communicate explicitly with patients to guide expectations and arrangements for follow-up within primary care. The findings from qualitative study in this thesis suggested that the provision of intensive relapse prevention within primary care is unrealistic. However, primary care clinicians could be more mindful of

incorporating simple relapse prevention techniques (such as discussion of early warning signs or brief relapse prevention planning) into their routine practice. Clinical education with qualified and training GPs could be facilitated by commissioners to increase clinician confidence and expertise in this area.

This study has highlighted other ways in which routine clinical general practice could be improved to improve the ongoing care of people with depression. It is clear that the importance of continuity is recognised and its important role in depression care is further reinforced by the findings from this study. Practices could consider adapting current workflow systems to enable a greater degree of continuity without necessarily placing additional burden on practice resources. Many of the GPs interviewed felt that continuity of care for depression helped to facilitate more efficient follow-up consultations, as well as improving the consulting experience for GPs and patients, and therefore could be encouraged on this basis. I direct clinicians to the framework presented as Figure 8.3, which captures some of the simple, achievable actions that could be taken, based on the findings from this study. It is envisaged that this could be used to guide quality improvement activities and improvement of care within general practice. Consideration will be given as to how to produce accessible materials (for example, audio-visual materials) for clinicians to disseminate these findings more effectively. Longer-term, curricula for both medical student and GP training should consider whether to include more focus on the continuation and maintenance phases of depression in addition to the current focus on acute diagnosis and treatment. Education and training for GPs and other primary care clinicians focussing on the importance of regular antidepressant medication reviews, with a focus on holistic care, is also advised.

10.5.2. Implications for health policy and commissioning

In the following section on policy and commissioning implications, where specific policy developments are discussed, I have focussed on the

NHS in England as this is the context in which the qualitative interviews were conducted. The overall discussion beyond this is intended to have wider implications applicable to other health systems.

The previous section outlined the implications for clinical practice and education arising from the findings from this study. Many of the recommendations in that section can be implemented within the existing infrastructure and contractual frameworks with sufficient clinician buy-in. Implementation of services to address the other implications of the study would require political buy-in and policy change and will be discussed here. In Chapter One (Section 1.5), I outlined the contractual arrangements underlying the commissioning of general practice in England. As a reminder, primary medical services for most general practices are defined by the nationally-negotiated GP contract. Additional services can be contracted, usually in the form of nationally- (DES) or locally-defined (LES) enhanced services contracts. In a discussion of implementing change within general practice, these are sometimes described as "contractual levers" (Smith et al., 2013).

The findings from the study suggested that relapse prevention is poorly provided in the NHS currently. Both groups of participants interviewed felt that relapse prevention would be best situated within general practice, overseen by the primary care team. The main barriers to this currently were: funding and other resourcing (specifically: time, IT provision, room space, and staff); potentially large eligible patient population leading to high demand for relapse prevention; and lack of clinical capacity in general practice. A potential facilitator raised by participants in the study was the potential for recruiting mental health practitioners through the Primary Care Network (PCN) contract or other recruitment streams outside the core NHS GP contract.

The PCN contract is a DES (a nationally-directed additional contract that practices can choose to deliver in addition to their core offer) wherein individual practices form a network with other practices and collaborate at

scale, allowing the sharing of resources. The PCN DES allows practices to recruit additional primary care staff through the additional roles reimbursement scheme (ARRS). This is the mechanism through which mental health practitioners and social prescribers are recruited and employed in primary care currently (NHS England and NHS Improvement, 2019). GPs and people with lived experience of depression thought that these professionals represented a good opportunity to embed relapse prevention within primary care. Other non-GP healthcare professionals (e.g., nurse practitioners and paramedics) are increasingly employed to support GPs in general practice (Murphy et al., 2021). While the role of these professionals in the care of people with depression is less well-understood, it is certainly possible that there is a role to play for these professionals as well and this could be explored further.

Mental health practitioners have been available to PCNs through the ARRS since April 2021 with a view to improving integration of primary care and mental health services (NHS Confederation, 2021). Currently, mental health practitioners mainly see people with acute mental health problems and support GPs in the diagnosis and initial management. Increased provision of mental health practitioners through the PCN DES could offer the opportunity to build on this foundation to provide ongoing follow-up and relapse prevention. This would require additional funding but there are other challenges associated with this. One challenge will be the ability to recruit sufficient mental health practitioners to PCNs and, if able to, then the health system will need to be confident this will not result in fewer mental health professionals employed within other parts of the system, for example secondary care mental health providers. There is also a concern that the ARRS widens pre-existing inequalities within primary care, with more clinicians generally being recruited into PCNs located in more affluent areas (Nussbaum et al., 2021). A situation where more deprived communities had poorer access to relapse prevention for depression than those with less deprivation would not be acceptable. This workforce inequity is likely to impact on health inequalities and ongoing research is seeking to understand and make recommendations to mitigate this risk (Nussbaum et al., 2021).

Other opportunities for increasing the workforce available for delivering relapse prevention within primary care might be through increased integration of health services. This was a key recommendation of the Fuller Stocktake (Fuller, 2022), following the introduction of the Health and Care Act 2022. Increased integration was also a key recommendation of the NHS Mental Health Implementation Plan 2019/20 – 2023/24 (National Health Service, 2019), although progress on this has been impeded by the pandemic. Fuller recommended the formation of "integrated neighbourhood teams" which include members of the primary care and mental health teams and may allow more joined-up and collaborative care of patients, with better inter-professional communication. This approach may allow a more prominent role for community partners and third-sector organisations in the longer-term support of people with depression and may help address some of the NHS workforce challenges. Improved integration between primary care and NHS Talking Therapies is also recommended to allow a more patient-centred and less rigid approach to the care of people with mental health problems (Martin et al., 2022).

Other resource limitations highlighted by GPs interviewed in this study included clinical rooms and premises (known collectively as estates). The Fuller Stocktake also outlined an estates plan for primary care (Fuller, 2022), and without sufficient implementation of the estates plan, practices will lack the clinical space for additional staff to see patients in person. The findings from this study suggest a need for primary care estates to be prioritised to ensure that patients can receive ongoing care within a primary care setting.

Another contractual lever commonly used to effect change in primary care are financial incentives. The Quality and Outcomes Framework (QOF) is the main structure for this in primary care and is associated with the core GP contract. The Investment and Impact Fund (IIF) is a newer financial incentive structure associated with the PCN DES. Either QOF or IIF could be used as a means by which to incentivise some of the activities identified in this study as being beneficial in the ongoing and longer-term care of people with

depression. They could also be used to financially incentivise relapse prevention. While such incentive structures have been shown to lead to improvements in some specific outcomes in the short term (e.g., blood pressure, cholesterol and diabetes-related measures), they have been found to be less successful in leading to proactive and preventative care (Langdown and Peckham, 2014). They generally do not lead to improved integration or coordination of care, holistic care, or better patient experience (Forbes et al., 2017). If this approach were taken, there would need to be a clear rationale for using financial incentive structures to implement any of the findings from this study, and evidence that such an approach would be likely to improve patient care.

10.6. Recommendations for future research

10.6.1. Relapse of depression in primary care

This study has identified that discussion of relapse and provision of relapse prevention are inconsistently applied in practice. It would be useful to attempt to objectively verify and quantify this. Routinely-collected data from patient clinical records could be explored to see if they allow researchers to explore this further.

This study also identified a gap between the way in which the 5 Rs are conceptualised and used widely in the psychological literature and their applicability and perceived usefulness in a primary care setting. Further work to empirically validate these terms, which should include standardising and validating primary care definitions of relapse and remission, might help to standardise future prognosis research in this area. This is likely to require further qualitative work, as well as perhaps a Delphi process with lay representatives, mental health professionals and primary care experts. Definitions that are meaningful to a primary care setting are likely to need to incorporate information around symptomatic improvement, but be less

dependent on more traditional diagnostic frameworks for MDD and include more focus on psychosocial function and quality of life.

A similar tension was evident with respect to diagnosing depression and distinguishing it from distress. My critical view is that any evidence-based tools or frameworks developed for use in primary care must reflect the problems and constructs most relevant and valid to GPs (and non-GP primary care health professionals who see and treat people with depression) and patients in primary care. It may be that it is not important to differentiate between the distress and depression or perhaps, as Gask (2008) argues, we need to develop a new primary care-based classification system for common mental disorders. However, disregarding well-used technical definitions risks limiting our ability to transfer research findings from other settings to a primary care setting. Accurate risk stratification in the longer term is likely to be dependent on ensuring GPs are precise in their identification and coding of presenting conditions. Further work is recommended to explore these issues. It is likely that, in the short- to medium-term, we will need to take a pragmatic approach of relying on diagnostic standards to guide the standardisation of clinical care and research, whilst recognising and being mindful of their limitations.

10.6.2. Prognosis research for relapse of depression in primary care

The PROGRESS Framework outlines four categories of prognosis research: overall prognosis; prognostic factor research; prognostic model research; and predictors of treatment effect (Riley et al., 2019b). This study has not focussed on overall prognosis research, which is concerned with overall outcomes for people with a particular condition in the context of routinely available care. I will discuss the other three categories in this section.

10.6.2.1. Prognostic factor research

There has been some robust prognostic factor research looking at predictors of relapse and an umbrella review (discussed earlier in this chapter) summarised the literature (Buckman et al., 2018). It is notable that few potential prognostic factors have been studied in a primary care context. Given that individualised prediction incorporating recognised prognostic factors remains suboptimal, it may be worth revisiting the need for more prognostic factor research.

The qualitative findings from this study have demonstrated the perceived importance by GPs and people with lived experience of social, personal and environmental factors in determining the course of depression and outcomes. Deficiencies and inconsistencies in the way such data are collected and coded in routine in primary care make this more difficult to use for risk prediction. The findings from GPs and people with lived experience of depression also highlighted concerns that some of the variables that we consider to be either predictive of or protective against relapse may actually have a more complicated relationship with outcomes than is often assumed. For example, being married and employed were both raised as variables that are often assumed to signify strong social support but may actually be harmful to mental health depending on the specific circumstances. Further research is needed to better understand the important predictors of relapse in primary care. It would be valuable to understand if information about these predictors can be captured and recorded in an acceptable and valid way by GPs and other primary care health professionals.

In particular, relationship status was associated with relapse in statistically significant way. While there was a small amount of pre-existing evidence for this, the current study lends some weight to the idea that this bears further investigation beyond this study. Childhood adversity was reported to be an important factor by GPs and people with lived experience of depression and there is evidence suggesting this is a strong prognostic factor for relapse (Buckman et al., 2018). It was not included in this study

because it is not routinely captured by GPs (or the researchers conducting the studies the IPD was taken from). However, given that we know that routinely asking people about childhood maltreatment is not harmful (Becker-Blease and Freyd, 2006; Nanni, Uher and Danese, 2012) and given the importance placed on this as a factor by participants in this study, it may be feasible to explore the use of childhood maltreatment and adversity as part of routine relapse risk assessment in primary care going forwards. Similarly, age of onset is a recognised relapse risk factor and ways of capturing and coding this reliably in primary care records could be explored. There have also been some efforts to develop clinically useful and valid brief instruments to measure rumination (Barbic, Durisko and Andrews, 2014; Marchetti et al., 2018), which could be explored in a primary care setting.

Many of the factors reported by participants as important (e.g., employment, relationship status) appeared to be proxy measures for perceived social support. This was seen by participants as a particularly important factor in protecting against relapse and a poorer depression course. The role of perceived social support has not been examined as a relapse predictor and is not measured in a uniform way in primary care. The Multidimensional Scale of Perceived Social Support (MSPSS) is a brief, self-report measure which uses a Likert-type scale to assess subjective social support (Zimet et al., 1990). It has been used to measure perceived social support in primary care (Grassi et al., 2000), and with good reliability in a recent community-based cross-sectional survey study (Grey et al., 2020). Social support has been measured successfully elsewhere using an adapted Social Support Scale (Buckman et al., 2021a). The acceptability and feasibility of these measures in primary care research and general practice could be assessed going forwards.

10.6.2.2. Prognostic model research

The aetiology of depression and depressive relapse is multifaceted, and multivariable models are likely to be a more helpful approach to predicting outcomes than relying on the presence or absence of single

prognostic factors. Prognostic models are already well established in clinical practice for a number of physical health problems, for example cardiovascular health and primary prevention of cardiovascular disease (Steyerberg et al., 2013). While there are many prognostic model development and validation studies reported in the literature, only a small proportion of these end up being implemented in a clinical setting (Steyerberg et al., 2013). The model developed in this study had sub-optimal predictive performance and could not be recommended for implementation in its current iteration. This study has, however, provided evidence that relapse risk prediction is felt to be an important goal by GPs and people with lived experience of depression. While I have made an argument in this chapter that our current priority ought to be ensuring adequate provision of relapse prevention in primary care for all patients for whom it is felt necessary, there is also benefit in continuing to focus on prognosis model research in this area.

A limitation of this study was that the data in the included studies were taken from samples collected for other purposes, for example RCTs and longitudinal cohort studies. While these are considered acceptable and feasible sources of data for prognostic model studies (Pajouheshnia et al., 2019), there may be advantages to prospectively gathering data (in a pre-designed prospective cohort study) with the explicit purpose of prognostic model development (Riley et al., 2019b). A benefit of this is that researchers can control the collection and ensure standardised measurement of predictor and outcome information, but such an approach is more costly and time-consuming than the secondary analysis of pre-existing data and would require a commitment to resource and fund such work. The International Taskforce for relapse prevention of depression (ITFRA) (www.itfra.org) have begun to address these issues by bringing together data from trials of existing relapse prevention interventions and aiming to harmonise predictor and outcome measurement to improve personalised medicine in this area. In this study, I modelled the outcome of relapse as a binary outcome. The reason for this was to enable me to incorporate criteria for reliable and significant change to define the outcome. The prediction of a clinically

significant outcome is more likely to be clinically useful than a prediction of a continuous score on a validating screening tool, such as the PHQ-9. However, it might be informative as part of future work to model the outcome on its continuous scale (i.e., build a linear model that estimates the outcome value, on PHQ-9 or alternative measure, rather than the probability of a binary outcome). If dichotomisation were felt to be clinically important, this could be done post-prediction.

In the IECV for this study to assess generalisability, I used the mean intercept and predictor-outcome associations. The results of the IECV demonstrated that the generalisability of the developed model is not guaranteed. The average intercept was obtained by pooling the individual intercept estimates for each study using random effects meta-analysis. An alternative approach that could be considered in future work is intercept selection, where similarities in the outcome frequency in a new population (if known), or similarities in the baseline characteristics (for example, mean age or proportion female) of a new population to the studies used for model derivation, is used to guide the intercept used when applying the model in a new population (Debray et al., 2013).

In my quantitative study, I presented the performance measures related to discrimination and calibration. A further important measure of model performance it its clinical usefulness. Clinical usefulness is generally dependent on, but not dictated by, a certain level of discrimination and calibration (Steyerberg, 2019). Given the lack of promising statistical performance metrics in this study, I did not include an assessment of clinical usefulness. However, prior to implementation of a model in practice, clinical usefulness needs to be considered. This can be done through decision curve analysis to assess the net benefit of using the model. Decision analysis incorporates the clinical consequences, in the form of trade-offs, of using a prediction model (Vickers, Van Calster and Steyerberg, 2016). If the predictive performance of this relapse prediction model can be improved in the future through recalibration or updating, then clinical usefulness should be explored. The benefits of intervening with, say, a relapse prevention

intervention in those with a predicted risk of relapse above a certain threshold would be potentially preventing a relapse which are costly in terms of time spent seeking medical help from GPs and also further psychological therapies (Clarke et al., 2022). Other costs to consider are potential prolonged prescribing of antidepressant medication, economic impact of time off work (Gauthier et al., 2019) and morbidity and decreased quality of life for the individual. Further considerations are the costs of offering a relapse prevention intervention (which has costs associated with its delivery) to somebody who would not have relapsed, but who was incorrectly identified at high risk by the model. The net benefit of using the model could be explored at a range of cut-points using decision curves.

A further approach to consider in the future prognostic model research is the use of dynamic models, which update their predictions as predictor information is updated. Dynamic models have been developed for predicting psychological treatment outcomes (Bone et al., 2021) and relapse of anxiety and depression after treatment with low-intensity CBT through IAPT (Lorimer et al., 2020). This approach would be valuable in a primary care setting, and may be useful in support self-monitoring of patients. Most of the included predictors in the studies identified in this study were clinical or demographic variables. It is possible that including a greater number of biomarkers or genetic information may help move towards such a precision medicine approach, as has been shown promising in a number of other areas, including diagnosing mood disorders (Le-Niculescu et al., 2021). Nevertheless, such an approach may not be clinically feasible, and an important consideration for researchers is the context and setting in which a prognostic model is intended to be used. Models intended for a primary care setting, for example, may need to focus on a different set of predictors than those intended for use within a specialist service. Primary care-based models would ideally need to include predictors that were available and routinely collected in primary care, such as demographics, socioeconomic information, comorbidities and depression history characteristics. A final consideration is that it was notable in the qualitative study that relapse was frequently discussed within the context of antidepressant prescribing, and particularly

when stopping or reducing antidepressants. The drive to standardise guidance around stopping antidepressant medication has increased in prominence recently, with several studies (Duffy et al., 2021; Bowers et al., 2020) looking at ways to achieve this. A relapse prediction tool at the point of considering reducing or stopping antidepressants, rather than the point of remission, may be useful to GPs and people with depression when deciding whether, and how, to do this.

Longer term, any developed model with promising performance on internal validation should undergo external validation (in a different dataset, to assess generalizability) and, ideally, independent validation (by a different research team, to reduce risk of bias). External validation could be done on either an unrelated retrospective dataset or, preferably, a prospective dataset collected specifically for this purpose. Finally, the impact of a model should be evaluated, and the gold standard way of doing this is through a RCT with clinically meaningful outcome measures (Hingorani et al., 2013), and ideally a qualitative component with key stakeholders to explore feasibility (Hoesseini A, van Leeuwen N and Sewnaik A, 2022). Cuijpers recently highlighted the importance of assessing the effect of mental health treatments on patient-defined outcomes (e.g., quality of life and functional outcomes) as well as those determined to be important by researchers or clinicians (Cuijpers, 2020). This is applicable to health technologies, like prognostic models, and exploring patient-defined outcomes should form a part of any future implementation assessment.

10.6.2.3. Predictors of treatment effect

The last type of prognosis research outlined in the PROGRESS Framework is identifying predictors of treatment effect. This can help drive personalised medicine (Zanardi et al., 2020) and there have been some examples where this approach has been used to improve outcomes. One study used socio-demographic and clinical characteristics to personalise psychotherapy, resulting in a small (probably not clinically significant) improvement in outcomes (Bauer-Staeb et al., 2023). In primary care,

matching depression treatment to severity prognosis also led to small improvements favouring the intervention arm (Fletcher et al., 2021). A personalised medicine approach is an important future research goal for relapse prevention of depression, and work is already underway aiming to move beyond stratification to provide more robust evidence for treatment moderators and prescriptive factors in relapse prevention (Breedvelt et al., 2020).

10.6.3. Relapse prevention interventions in primary care

The findings from this study have shown that we cannot currently predict an individual's risk of relapse with sufficient accuracy to guide the provision of relapse prevention. The findings also suggest that the majority of people with remitted depression do not currently receive relapse prevention in practice. Relapse and recurrence occur in a significant proportion of people with remitted depression and are a source of considerable morbidity. The economic burden of depression is higher in those who experience relapse or recurrence than in those who do not (Gauthier et al., 2019) and, while interventions to prevent relapse or recurrence of depression (including pharmacological and psychological approaches) can be resource-intensive, they are effective (Clarke et al., 2015; Geddes et al., 2003; Breedvelt et al., 2021b) and cost-effective (Klein et al., 2019).

There is an interesting parallel to be drawn with lessons learned from attempts to prevent suicide through suicide risk prediction (Large, 2018). Large (2018) outlined the different approaches to prevention: universal approaches, which target whole populations; selective approaches, which target higher-risk groups; and indicated approaches, directed at individuals (Cuijpers et al., 2008). The author concluded, in the case of suicide prevention and in the absence of sufficiently accurate risk prediction tools, that universal approaches were most appropriate. I argue that we should take a similar (universal) approach to relapse prevention of depression in primary care. In the absence of accurate statistical risk prediction, and given that clinical prediction tends to predict worse or no better than statistical

prediction (Meehl, 1954), relapse prevention should be considered for all people experiencing an episode of depression. The therapeutic approach should be determined in partnership with patients, in line with principles of shared decision making, and could be targeted at ongoing social, personal and environmental factors if appropriate.

In Chapter Two (Section 2.4.1), I presented a summary of the evidence for relapse prevention of depression in primary care. As we saw, despite the fact that the vast majority of people with depression are managed in primary care (Ramanuj, Ferenchick and Pincus, 2019), evidence for the effectiveness and cost-effectiveness of interventions to prevent relapse in primary care is lacking (NICE, 2022). The availability and supply of psychological treatments as recommended by NICE is inadequate at present and it is possible that these interventions do not constitute realistic treatment options in the real-world NHS (Mental Health Task Force, 2016). It is not well understood which interventions work best for which patients, and national guidelines differ in terms of risk stratification and treatment recommendations. Furthermore, there are multiple kinds of interventions available. Patient preferences and individual patient circumstances will always be important in guiding treatment decisions through discussion with healthcare professionals. An ability to be able to inform patients about the relative effectiveness of the range of different treatment options is an essential part of this decision-making process.

In the qualitative study in this thesis, GPs and people with lived experience of depression reported that they view relapse prevention to be the purview of primary care and that there are benefits to having such care situated within the GP surgery or primary care setting, rather than being referred elsewhere. Lessons from studies of relapse prevention for depression in primary care may help us to develop new brief interventions, tailored for a primary care setting, which are affordable, feasible and scalable.

The interventions for relapse prevention that have been trialled in primary care settings are heterogeneous and not directly comparable in a pair-wise meta-analysis. However, a network meta-analysis would allow for mixed treatment comparison using all of the evidence base of relapse prevention interventions in primary care to provide a quantitative synthesis of relative effectiveness. Formal risk of bias assessment of the RCTs, using a standardised tool such as the Cochrane risk-of-bias tool (Higgins et al., 2011) or the more recently-developed RoB 2 tool (Sterne et al., 2019), would be useful in helping to understand strengths and weaknesses in the RCTs. Finally, as well as a quantitative synthesis to enable us to understand the relative effectiveness, a realist synthesis is recommended prior to efforts to implement interventions (Rycroft-Malone et al., 2012). Realist syntheses allow us to better understand the underlying causal mechanisms and conditions under which interventions work or not. The Medical Research Council have published guidance on developing and evaluating complex interventions, which could be used to guide the development of novel primary care-based interventions (Craig et al., 2008). Implementation research would be required to ensure that any intervention developed can be made available to a greater number of patients in a scalable and feasible way.

A potentially useful modality in primary care is computerised relapse prevention. Internet-based and other computerised interventions have shown promise for acute depression care (Andersson and Cuijpers, 2009) and some [for example, computerised CBT (cCBT)] have been commissioned for this purpose in the UK NHS (Gilbody et al., 2015). Large-scale trials of cCBT have since shown that, in primary care, it is only more effective than usual care where it is guided and supported by somebody, rather than being a purely self-guided intervention (Gilbody et al., 2015, 2017). Providers and commissioners would need to be mindful of these findings, and also that online interventions are not suitable for all patients and allocation of interventions should be guided by patient preferences as far as possible.

10.7. Reflections on the study

10.7.1. Patient and Public Involvement and Engagement

The PPIE has been one of the more rewarding and enlightening aspects of the PhD and Fellowship as a whole. Feedback and ideas from the PAG have meaningfully shaped the research and plans for dissemination and communication to lay audiences. Here, I will describe and reflect on where the PPIE has impacted the research throughout the thesis.

The PAG were involved throughout the qualitative component. Throughout the thematic analysis, I hosted three workshops to develop and refine the themes. I also undertook two pilot interviews with PAG members to help refine the semi-structured topic guides used for the interviews prior to beginning data generation. One workshop was held at the beginning (September 2022) to discuss preliminary data and to shape the initial generation of themes, one was held in April 2023 to refine the themes and one in June 2023 to finalise themes.

These workshops were instrumental in shaping and refining the themes. PAG members described the importance of ensuring Theme 1 was explicit in differentiating between intrinsic (personal) and extrinsic (environmental and social) factors in determining depression course. Under Theme 2, the PAG advised that "importance of demonstrating empathy" should be "importance of empathy" as "demonstrating" was GP-centric and could potentially mean performative empathy was more important than sincere empathy. Similarly, the sub-theme "Patients at the centre of care within general practice" was originally titled "General practice at the centre of care". The PAG thought this was again GP-centred and was potentially disempowering to patients.

Interestingly, the PAG felt the framework (Figure 8.3) was a bit "obvious" and thought that GPs would be aligned with the guidance in the

framework already. However, as the qualitative findings demonstrated, practice with respect to follow-up of patients is not consistent and alignment with best practice prohibited by time and resource constraints.

When I have been asked about the use of PPIE in the research, my temptation has been to talk about the importance of involving people in research about them and hearing the patient perspective. This remains true, but the main point is that the use of PPIE was absolutely instrumental to my project being a success; it made the research better and the findings more relevant to the ultimate end-users of the results of the study.

10.7.2. Difference to me as researcher and clinician

Having had a small amount of research experience prior to starting this PhD, I felt quite confident embarking on this project. What I have learned, however, is that there is a big difference between undertaking discrete projects as an academic GP trainee and project-managing a large, externally-funded study. The first major difference was that I had devised the study from inception, with support from my supervisors. I had begun the study thinking in quite precisely-defined terms about the problem of relapse of depression and approaching it as an almost mechanistic way. It has only been after reflecting on the challenges of statistical prediction and integrating these with the findings from the qualitative workstream that I have realised that, particularly in primary care, these terms and problems are not quite as tightly defined as I may have expected them to be or as the literature might lead us to believe. Thinking about how terms are actually conceptualised and operationalised in practice seems to me to be key to ensuring that we, as a research community, can adequately explore and address clinical problems. Having the opportunity to engage critically with a programme of work, with input from my PAG, has informed and sharpened my thinking and I feel much more like a researcher having gone through that process.

There has also been a huge learning curve procedurally, in many different areas. First, this was the first application I had completed for a competitive, externally peer-reviewed research award (NIHR Doctoral Research Fellowship). It was also the first project I had undertaken that had involved handling several datasets and cleaning and analysing these using Stata. It also involved coordinating data-sharing agreements and navigating the University's data storage policies and arrangements. Secondly, I had to receive ethics approval and use the HRA IRAS system. Then, I worked with the Clinical Research Network, the research delivery arm of the NIHR, to recruit general practices and participants. This involved learning about the AcoRD guidelines and the attribution of research costs (reimbursement for interviews) and NHS support costs (database search and eligibility screening) (Department of Health, 2012). This is the kind of procedural knowledge that will be invaluable to me as I progress into more senior academic posts.

If I were to have done anything differently, on reflection it may have been useful to begin the qualitative workstream prior to the quantitative workstream to more effectively guide the definition of predictors. However, in practice, it was still possible for the qualitative work to impact on the delivery and interpretation of the quantitative work and findings and I would still have been limited by the data available from the studies. Furthermore, working with the PAG and multidisciplinary supervisory team on selecting the predictors and developing the definitions for data harmonisation gave the opportunity for validating these.

I think my academic work makes me a better clinician as well as researcher. It allows me to ask questions of my practice and apply knowledge and critical skills within the consulting room. Being involved in the generation of knowledge and having an understanding of how this becomes guidance and policy leads me to practice in a more informed and critical way. My hope is that the findings from the study can be helpful to GPs and improve care for people with depression. Undergoing the process of drafting the application and writing up the thesis has helped me to crystallise in my

mind what the key purpose and messages are from the work and I am pleased to have successfully completed the study.

10.7.3. Impact of COVID-19

I have avoided excessive discussion of the COVID-19 pandemic during this thesis, other than where it was directly relevant to the subject matter. However, given the impact of COVID-19 on the study, it is somewhat unavoidable. The NIHR Doctoral Research Fellowship was submitted in December 2017, and the interview was in June 2018. I then commenced the study in January 2019 planned over five years, and therefore the majority of this study has taken place against a backdrop of increased clinical pressures, constantly changing public health messaging and a shift to remote working. It has impacted on the training courses I had planned to undertake and conferences I had planned to attend, which were mainly online until very recently. The pandemic resulted in some difficulties obtaining data for the quantitative study, because of pressures on the data holders. As discussed, all the qualitative interviews were conducted remotely.

At all stages, I was supported in meeting the challenges posed by my supervisory team and Thesis Advisory Panel. Overcoming the various hurdles and barriers encountered along the way felt like a form of research training in itself, and I have learned that no problem is insurmountable and will often result in better research.

10.8. Conclusions

This thesis presented a programme of work aimed at improving the ability of primary care clinicians to predict relapse of depression and exploring views around relapse prediction and prevention. It opened with a problem: that a high proportion of people who are treated with depression in primary care will relapse after an initial improvement and there is limited guidance to help clinicians risk stratify and support such patients. The study

presented in this thesis has found that predicting relapse of depression in a primary care setting is challenging and we cannot yet do so with sufficient accuracy. Qualitative work with GPs and people with lived experience of depression suggests that relapse is an important problem but is not routinely discussed with patients and relapse prevention is not routinely made available. As such, we would currently be better focussing on facilitating discussion of relapse risk and prevention in primary care consultations, which GPs and people with lived experience of depression agree would be beneficial. By focussing on continuity of care, GP-patient relationships and general practice follow-up systems, primary care clinicians could help improve experiences and outcomes for people with depression. GPs think that relapse prevention could be part of primary care's role and patients would generally prefer to receive such interventions within their general practice. Further work should explore ways of developing and implementing feasible, scalable and acceptable relapse prevention interventions that could be offered to all patients. Barriers to the implementation and delivery of such interventions in primary care should be addressed by researchers and policy-makers.

References

Ali, S. et al. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy*, 94, pp.1–8. [Online]. Available at: doi:10.1016/j.brat.2017.04.006.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. 5th Edn. Washington, DC.

Anderson, J. et al. (2021). Patient Satisfaction With Remote Consultations in a Primary Care Setting. *Cureus*. [Online]. Available at: doi:10.7759/cureus.17814.

Andersson, G. and Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*, 38 (4), pp.196–205. [Online]. Available at: doi:10.1080/16506070903318960.

Andrade, C. (2021). Z Scores, Standard Scores, and Composite Test Scores Explained. *Indian Journal of Psychological Medicine*, 43 (6), pp.555–557. [Online]. Available at: doi:10.1177/02537176211046525.

Ang, B., Horowitz, M. and Moncrieff, J. (2022). Is the chemical imbalance an 'urban legend'? An exploration of the status of the serotonin theory of depression in the scientific literature. *SSM - Mental Health*, 2. [Online]. Available at: doi:10.1016/j.ssmmh.2022.100098.

Angstman, K. B. et al. (2017). Prediction of Primary Care Depression Outcomes at Six Months : Validation of DOC-6 ©. *J Am Board Fam Med*, pp.281–287. [Online]. Available at: doi:10.3122/jabfm.2017.03.160313.

Angulo, P. et al. (2007). The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology*, 45 (4), pp.846–854. [Online]. Available at: doi:10.1002/hep.21496.

Arroll, B., Moir, F. and Kendrick, T. (2017). Effective management of depression in primary care: a review of the literature. *BJGP Open*, 1 (2). [Online]. Available at: doi:10.3399/are.

Backs-Dermott, B. J., Dobson, K. S. and Jones, S. L. (2010). An evaluation of an integrated model of relapse in depression. *Journal of Affective Disorders*, 124 (1–2), pp.60–67. [Online]. Available at: doi:10.1016/j.jad.2009.11.015.

Badger, F. ; and Nolan, P. (2013). Concordance with antidepressant medication in primary care. *Nursing Standard* , 20 (52), pp.35–40.

Barbic, S. P., Durisko, Z. and Andrews, P. W. (2014). Measuring the bright side of being blue: A new tool for assessing analytical rumination in depression. *PLoS ONE*, 9 (11). [Online]. Available at: doi:10.1371/journal.pone.0112077.

Barker, I., Steventon, A. and Deeny, S. R. (2017). Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: Cross sectional study of routinely collected, person level data. *BMJ (Online)*, 356. [Online]. Available at: doi:10.1136/bmj.j84.

Bartley, M. and Ferrie, J. (2001). Glossary: unemployment, job insecurity, and health. *Journal of epidemiology and community health*, 55 (11), pp.776–781. [Online]. Available at: doi:10.1136/jech.55.11.776.

Bauer-Staeb, C. et al. (2023). Personalised psychotherapy in primary care: evaluation of data-driven treatment allocation to cognitive–behavioural therapy versus counselling for depression. *BJPsych Open*, 9 (2). [Online]. Available at: doi:10.1192/bjo.2022.628.

Baumann, L. C. and Ylinen, A. (2020). *Prevention: Primary, secondary, tertiary. In Encyclopedia of behavioral medicine*. Cham: Springer International Publishing.

Bazeley, P. (2018). *Integrating Analyses in Mixed Methods Research*. SAGE Publications Inc.

Becker-Blease, K. A. and Freyd, J. J. (2006). Research participants telling the truth about their lives: The ethics of asking and not asking about abuse. *American Psychologist*, 61 (3), pp.218–226. [Online]. Available at: doi:10.1037/0003-066X.61.3.218.

Bergin, M., Wells, J. S. and Owen, S. (2008). Critical realism: a philosophical framework for the study of gender and mental health. *Nursing Philosophy*, 9, pp.169–179.

Berlanga, C. et al. (1999). Personality and clinical predictors of recurrence of depression. *Psychiatric Services*, 50 (3), pp.376–380. [Online]. Available at: doi:10.1176/ps.50.3.376.

Beshai, S. et al. (2011). Relapse and recurrence prevention in depression: Current research and future prospects. *Clinical Psychology Review*, 31 (8), pp.1349–1360. [Online]. Available at: doi:10.1016/j.cpr.2011.09.003.

Biesheuvel-Leliefeld, K. E. M. et al. (2017). Effectiveness of Supported Self-Help in Recurrent Depression: A Randomized Controlled Trial in Primary Care. *Psychotherapy and Psychosomatics*, 86 (4), pp.220–230. [Online]. Available at: doi:10.1159/000472260.

Biesheuvel-Leliefeld, K. E. M. et al. (2018). A supported self-help for recurrent depression in primary care; An economic evaluation alongside a multi-center randomised controlled trial. *PLoS ONE*, 13 (12). [Online]. Available at: doi:10.1371/journal.pone.0208570.

Binder, H., Sauerbrei, W. and Royston, P. (2013). Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: A simulation study with continuous response. *Statistics in Medicine*, 32 (13), pp.2262–2277. [Online]. Available at: doi:10.1002/sim.5639.

Bland, J. and Altman, D. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310 (170). [Online]. Available at: https://doi.org/10.1136/bmj.310.6973.170 [Accessed 12 June 2023].

Bockting, C. L. et al. (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review*, 41, pp.16–26. [Online]. Available at: doi:10.1016/j.cpr.2015.02.003.

Bockting, C. L. H. et al. (2018). Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicentre, randomised controlled trial. *The*

Lancet Psychiatry, 5 (5), pp.401–410. [Online]. Available at: doi:10.1016/S2215-0366(18)30100-7.

Bone, C. et al. (2021). Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. *The Lancet Digital Health*, 3 (4), pp.e231–e240. [Online]. Available at: doi:10.1016/S2589-7500(21)00018-2.

Bosanquet, K. et al. (2017). Collaborative care for screen-positive elders with major depression (CASPER plus): A multicentred randomized controlled trial of clinical effectiveness and cost-effectiveness. *Health Technology Assessment*, 21 (67), pp.1–251. [Online]. Available at: doi:10.3310/hta21670.

Bosman, R. C. et al. (2016). Long-term antidepressant use: a qualitative study on perspectives of patients and GPs in primary care. *British Journal of General Practice*, 66 (651), pp.e708–e719. [Online]. Available at: doi:10.3399/bjgp16x686641.

Bouwmeester, W. et al. (2013). Prediction models for clustered data: Comparison of a random intercept and standard regression model. *BMC Medical Research Methodology*, 13 (1). [Online]. Available at: doi:10.1186/1471-2288-13-19.

Bowers, H. M. et al. (2020). Supporting antidepressant discontinuation: The development and optimisation of a digital intervention for patients in UK primary care using a theory, evidence and person-based approach. *BMJ Open*, 10 (3). [Online]. Available at: doi:10.1136/bmjopen-2019-032312.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3 (2), pp.77–101. [Online]. Available at: doi:10.1191/1478088706qp063oa.

Braun, V. and Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.

Braun, V. and Clarke, V. (2016). (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology*, 19 (6), Routledge., pp.739–743. [Online]. Available at: doi:10.1080/13645579.2016.1195588.

Braun, V. and Clarke, V. (2019a). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11 (4), Routledge., pp.589–597. [Online]. Available at: doi:10.1080/2159676X.2019.1628806.

Braun, V. and Clarke, V. (2019b). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 00 (00), pp.1–16. [Online]. Available at: doi:10.1080/2159676X.2019.1704846.

Braun, V. and Clarke, V. (2021a). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21 (1), pp.37–47. [Online]. Available at: doi:10.1002/capr.12360.

Braun, V. and Clarke, V. (2021b). Conceptual and Design Thinking for Thematic Analysis. *Qualitative Psychology*, 9 (1), pp.3–26. [Online]. Available at: doi:10.1037/qup0000196.

Braun, V. and Clarke, V. (2022). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a *knowing* researcher. *International Journal of Transgender Health*, pp.1–6. [Online]. Available at: doi:10.1080/26895269.2022.2129597.

Breedvelt, J. J. F. et al. (2020). Individual participant data (IPD) meta-analysis of psychological relapse prevention interventions versus control for patients in remission from depression: A protocol. *BMJ Open*, 10 (2), pp.1–8. [Online]. Available at: doi:10.1136/bmjopen-2019-034158.

Breedvelt, J. J. F. et al. (2021a). Continuation of Antidepressants vs Sequential Psychological Interventions to Prevent Relapse in Depression: An Individual Participant Data Meta-analysis. *JAMA Psychiatry*, 78 (8), pp.868–875. [Online]. Available at: doi:10.1001/jamapsychiatry.2021.0823.

Breedvelt, J. J. F. et al. (2021b). Psychological interventions as an alternative and add-on to antidepressant medication to prevent depressive relapse: Systematic review and meta-Analysis. *British Journal of Psychiatry*, 219 (4), Cambridge University Press., pp.538–545. [Online]. Available at: doi:10.1192/bjp.2020.198.

Brickley, B. et al. (2020). A new model of patient-centred care for general practitioners: results of an integrative review. *Family practice*, 37 (2), pp.154–172. [Online]. Available at: doi:10.1093/fampra/cmz063.

British Medical Association. (2022). *Pressures in general practice data analysis*. [Online]. Available at: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/pressures-in-general-practice-data-analysis [Accessed 2 February 2023].

van Bronswijk, S. C. et al. (2019). A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder. *Depression and Anxiety*, 36 (3), pp.252–261. [Online]. Available at: doi:10.1002/da.22868.

Buckman, J. E. J. et al. (2018). Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clinical Psychology Review*, 64 (July 2017), pp.13–38. [Online]. Available at: doi:10.1016/j.cpr.2018.07.005.

Buckman, J. E. J. et al. (2020). 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 Research*, 4, pp.1–20. [Online]. Available at: doi:10.12688/wellcomeopenres.15225.3.

Buckman, J. E. J. et al. (2021a). Is social support pre-treatment associated with prognosis for adults with depression in primary care? *Acta Psychiatrica Scandinavica*, 143 (5), Blackwell Publishing Ltd., pp.392–405. [Online]. Available at: doi:10.1111/acps.13285.

Buckman, J. E. J. et al. (2021b). Role of age, gender and marital status in prognosis for adults with depression: An individual patient data meta-analysis. *Epidemiology and psychiatric sciences*, 30, p.e42. [Online]. Available at: doi:10.1017/S2045796021000342.

Buckman, J. E. J. et al. (2022). Life events and treatment prognosis for depression: A systematic review and individual patient data meta-analysis. *Journal of Affective Disorders*, 299, Elsevier B.V., pp.298–308. [Online]. Available at: doi:10.1016/j.jad.2021.12.030.

Bueno-Notivol, J. et al. (2021). Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. *International Journal of Clinical and Health Psychology*, 21 (1). [Online]. Available at: doi:10.1016/j.ijchp.2020.07.007.

Burcusa, S. L. and Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical Psychology Review*, 27 (8), pp.959–985. [Online]. Available at: doi:10.1016/j.cpr.2007.02.005.

Cachia, M. and Millward, L. (2011). The telephone medium and semi-structured interviews: a complementary fit. *Qualitative Research in Organizations and Management*, 6 (3), pp.265–277.

Campbell, R. (2009). *Campbell's psychiatric dictionary*. 9th Edition. New York: Oxford University Press.

Cates, C. (2010). *Cates plot*. [Online]. Available at: https://www.nntonline.net/ [Accessed 2 July 2023].

Chew-Graham, C. A., May, C. R. and Perry, M. S. (2002). Qualitative research and the problem of judgement: lessons from interviewing fellow professionals. *Family Practice*, 19 (3).

Chiang, P. C. et al. (2015). Implementing a qcancer risk tool into general practice consultations: An exploratory study using simulated consultations with Australian general practitioners. *British Journal of Cancer*, 112, pp.S77–S83. [Online]. Available at: doi:10.1038/bjc.2015.46.

Clark, D. M. et al. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47 (11), pp.910–920. [Online]. Available at: doi:10.1016/j.brat.2009.07.010.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, 23 (4), pp.318–327. [Online]. Available at: doi:10.3109/09540261.2011.606803.

Clarke, C. S. et al. (2022). Cost-Utility Analysis of Discontinuing Antidepressants in England Primary Care Patients Compared with Long-Term Maintenance: The ANTLER Study. *Applied Health Economics and Health Policy*, 20 (2), pp.269–282. [Online]. Available at: doi:10.1007/s40258-021-00693-x.

Clarke, D. M. et al. (2008). What do general practitioners think depression is? A taxonomy of distress and depression for general practice. *The Medical journal of Australia*, 188 (12 Suppl). [Online]. Available at: doi:10.5694/j.1326-5377.2008.tb01872.x.

Clarke, K. et al. (2015). Can non-pharmacological interventions prevent relapse in adults who have recovered from depression? A systematic review and meta-analysis of randomised controlled trials. *Clinical Psychology Review*, 39, pp.58–70. [Online]. Available at: doi:10.1016/j.cpr.2015.04.002.

Coe, A. et al. (2023). Self-reported reasons for reducing or stopping antidepressant medications in primary care: Thematic analysis of the diamond longitudinal study. *Primary Health Care Research and Development*, 24 (9). [Online]. Available at: doi:10.1017/S1463423623000038.

Cohen, D. and Recalt, A. (2019). Discontinuing Psychotropic Drugs from Participants in Randomized Controlled Trials: A Systematic Review. *Psychother Psychosom*, 88 (2), pp.96–104.

Cohen, Z. et al. (2021). The Development and Internal Evaluation of a Predictive Model to Identify for Whom Mindfulness-Based Cognitive Therapy Offers Superior Relapse Prevention for Recurrent Depression Versus Maintenance Antidepressant Medication. *Biological Psychiatry*, 89 (9 Supp).

Collins, G. et al. (2014). External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Medical Research Methodology*, 14 (1), pp.1–11.

Collins, G. S. et al. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *British Journal of Surgery*, 102 (3), John Wiley and Sons Ltd., pp.148–158. [Online]. Available at: doi:10.1002/bjs.9736.

Collins, G. S. et al. (2021). Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*, 11 (7). [Online]. Available at: doi:10.1136/bmjopen-2020-048008.

Collins, G. S. and Altman, D. G. (2009). An independent external validation and evaluation of QRISK cardiovascular risk prediction: A prospective open cohort study. *BMJ (Online)*, 339 (7713), pp.144–147. [Online]. Available at: doi:10.1136/bmj.b2584.

Coupland, C. et al. (2011). Antidepressant use and risk of adverse outcomes in older people: Population based cohort study. *BMJ (Online)*, 343 (7819), pp.1–15. [Online]. Available at: doi:10.1136/bmj.d4551.

Coventry, P. et al. (2015). Integrated primary care for patients with mental and physical multimorbidity: Cluster randomised controlled trial of collaborative care for patients with depression comorbid with diabetes or cardiovascular disease. *BMJ (Online)*, 350, pp.1–11. [Online]. Available at: doi:10.1136/bmj.h638.

Craig, P. et al. (2008). Medical Research Council Guidance: Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337:a1655.

Creswell, J. and Plano Clarke, V. (2011). *Designing and Conducting Mixed Methods Research*. Second Edition. Thousand Oaks, CA: Sage.

Creswell, J. W. et al. (2011). *Best Practices for Mixed Methods Research in the Health Sciences*. [Online]. Available at: http://obssr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf [Accessed 3 July 2021].

Creswell, J. W., Fetters, M. D. and Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *Annals of Family Medicine*, 2 (1), pp.7–12. [Online]. Available at: doi:10.1370/afm.104.

Cruickshank, J. (2012). Positioning positivism, critical realism and social constructionism in the health sciences: A philosophical orientation. *Nursing Inquiry*, 19 (1), pp.71–82. [Online]. Available at: doi:10.1111/j.1440-1800.2011.00558.x.

Cuijpers, P. et al. (2008). Preventing the Onset of Depressive Disorders: A Meta-Analytic Review of Psychological Interventions. *Am J Psychiatry*, 165, pp.1272–1280.

Cuijpers, P. et al. (2013a). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, 3 (4), pp.1–8. [Online]. Available at: doi:10.1136/bmjopen-2012-002542.

Cuijpers, P. et al. (2013b). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons. *World Psychiatry*, 12 (2), pp.137–148. [Online]. Available at: doi:10.1002/wps.20038.

Cuijpers, P. et al. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, 19 (1), pp.92–107.

Cuijpers, P. (2020). Expert Review of Neurotherapeutics Measuring success in the treatment of depression : what is most important to patients ? Measuring success in the treatment of depression : what is most important to. *Expert Review of Neurotherapeutics*, 00 (00), pp.1–3. [Online]. Available at: doi:10.1080/14737175.2020.1712807.

Davies, J. and Read, J. (2019). A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: Are guidelines evidence-based? *Addictive Behaviors*, 97 (August 2018), pp.111–121. [Online]. Available at: doi:10.1016/j.addbeh.2018.08.027.

Debray, T. P. A. et al. (2013). A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Statistics in Medicine*, 32 (18), pp.3158–3180. [Online]. Available at: doi:10.1002/sim.5732.

Debray, T. P. A. et al. (2017). A guide to systematic review and meta-analysis of prediction model performance. *BMJ (Online)*, 356. [Online]. Available at: doi:10.1136/bmj.i6460.

Debray, T. P. A. et al. (2019). A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research*, 28 (9), pp.2768–2786. [Online]. Available at: doi:10.1177/0962280218785504.

Debray, T. P. A. et al. (2023). Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration. *BMJ*, p.e071058. [Online]. Available at: doi:10.1136/bmj-2022-071058.

Department of Health. (2012). *Attributing the costs of health and social care Research & Development (AcoRD)*. [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/351182/AcoRD_Guidance_for_publication_May_2012.pdf [Accessed 10 October 2021].

Department of Health. (2017). *A framework for mental health research*. [Online]. Available at: https://www.gov.uk/government/publications/a-framework-for-mental-health-research.

Department of Health and Social Care. (2023). *The Hewitt Review: An independent review of integrated care systems*. [Online]. Available at: https://www.gov.uk/government/publications/the-hewitt-review-an-independent-review-of-integrated-care-systems [Accessed 10 April 2023].

Dewidar, O. et al. (2021). PRIME-IPD SERIES Part 1. The PRIME-IPD tool promoted verification and standardization of study datasets retrieved for IPD meta-analysis. *Journal of Clinical Epidemiology*, 136, pp.227–234. [Online]. Available at: doi:10.1016/j.jclinepi.2021.05.007.

Dixon, A. et al. (2010). *The Quality and Outcomes Framework (QOF): does it reduce health inequalities?*

Dowrick, C. et al. (2009). Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: Qualitative study. *BMJ*, 338, p.b663. [Online]. Available at: doi:10.1136/bmj.b663.

Dozois, D. J. A. (2021). Anxiety and depression in Canada during the COVID-19 pandemic: A national survey. *Canadian Psychology*, 62 (1), pp.136–142. [Online]. Available at: doi:10.1037/cap0000251.

Drinkwater, C., Wildman, J. and Moffatt, S. (2019). Social prescribing. *BMJ (Online)*, 364. [Online]. Available at: doi:10.1136/bmj.l1285.

Duffy, L. et al. (2021). Antidepressant medication to prevent depression relapse in primary care: The ANTLER RCT. *Health Technology Assessment*, 25 (69), pp.i–62. [Online]. Available at: doi:10.3310/hta25690.

Ekers, D. et al. (2013). Nurse-delivered collaborative care for depression and long-term physical conditions: A systematic review and meta-analysis. *Journal of Affective Disorders*, 149 (1–3), pp.14–22. [Online]. Available at: doi:10.1016/j.jad.2013.02.032.

Ettman, C. K. et al. (2020). Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. *JAMA Network Open*, 3 (9). [Online]. Available at: doi:10.1001/jamanetworkopen.2020.19686.

Fairhurst, K. and May, C. (2006). What general practitioners find satisfying in their work: Implications for health care system reform. *Annals of Family Medicine*, 4 (6), pp.500–505. [Online]. Available at: doi:10.1370/afm.565.

Falconieri, N. et al. (2020). Developing risk models for multicenter data using standard logistic regression produced suboptimal predictions: A simulation study. *Biometrical Journal*, 62 (4), pp.932–944. [Online]. Available at: doi:10.1002/bimj.201900075.

Fan, V. S. et al. (2005). Continuity of care and other determinants of patient satisfaction with primary care. *Journal of General Internal Medicine*, 20 (3), pp.226–233. [Online]. Available at: doi:10.1111/j.1525-1497.2005.40135.x.

Ferenchick, E. K., Ramanuj, P. and Pincus, H. A. (2019). Depression in primary care: part 1-screening and diagnosis. *BMJ (Clinical research ed.)*, 365, p.l794. [Online]. Available at: doi:10.1136/bmj.l794.

Fetters, M. (2020). *The Mixed Methods Research Workbook. Sage.*

Fletcher, S. et al. (2021). Matching depression management to severity prognosis in primary care: results of the Target-D randomised controlled trial. *British Journal of General Practice*, 71 (703), pp.e85–e94. [Online]. Available at: doi:10.3399/bjgp.2020.0783.

Flynn, M. and Hooper, G. (2020). Antimicrobial stewardship though FeverPAIN score: Successes and challenges in secondary care. *Clinical Infection in Practice*, 7–8. [Online]. Available at: doi:10.1016/j.clinpr.2020.100024.

Forand, N. R. and Derubeis, R. J. (2013). Pretreatment anxiety predicts patterns of change in cognitive behavioral therapy and medications for depression. *Journal of Consulting and Clinical Psychology*, 81 (5), pp.774–782. [Online]. Available at: doi:10.1037/a0032985.

Forbes, L. J. L. et al. (2017). The role of the Quality and Outcomes Framework in the care of long-term conditions: A systematic review. *British Journal of General Practice*, 67 (664), pp.e775–e784. [Online]. Available at: doi:10.3399/bjgp17X693077.

Foroutan, F. et al. (2020). GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology*, 121, pp.62–70. [Online]. Available at: doi:10.1016/j.jclinepi.2019.12.023.

Fram, S. M. (2013). The Constant Comparative Analysis Method Outside of Grounded Theory. *The Qualitative Report*, 18 (1). [Online]. Available at: http://www.nova.edu/ssss/QR/QR18/fram1.pdf.

Frank, E. et al. (1991). Conceptualization and Rationale for Consensus Definitions of Terms in Major Depressive Disorder: Remission, Recovery, Relapse, and Recurrence. *JAMA Psychiatry*, 48 (9), pp.851–855. [Online]. Available at: doi:10.1001/archpsyc.1991.01810330075011.

Fuller, C. (2022). *Next steps for integrating primary care: Fuller Stocktake report*.

Fusar-Poli, P. et al. (2018). The Science of Prognosis in Psychiatry: A Review. *JAMA Psychiatry*, 75 (12), pp.1280–1288. [Online]. Available at: doi:10.1001/jamapsychiatry.2018.2530.

Gartlehner, G. et al. (2015). *Nonpharmacological Versus Pharmacological Treatments for Adult Patients With Major Depressive Disorder. Comparative Effectiveness Review No. 161*. Rockville, MD. [Online]. Available at: www.effectivehealthcare.ahrq.gov.

Gask, L. et al. (2008). Capturing complexity: The case for a new classification system for mental disorders in primary care. *European Psychiatry*, 23 (7), pp.469–476. [Online]. Available at: doi:10.1016/j.eurpsy.2008.06.006.

Gauthier, G. et al. (2019). Economic burden of relapse/recurrence in patients with major depressive disorder. *Journal of Drug Assessment*, 8 (1), pp.97–103. [Online]. Available at: doi:10.1080/21556660.2019.1612410.

Geddes, J. et al. (2003). Relapse prevention with antidepressant drug treatment in depressive disorders: A systematic review. *Lancet*, 361 (9358), pp.653–661. [Online]. Available at: doi:10.1016/S0140-6736(03)12599-8.

Geraghty, A. W. A. et al. (2015). Distinguishing between emotional distress and psychiatric disorder in primary care attenders: A cross sectional study of the four-dimensional symptom questionnaire (4DSQ). *Journal of Affective Disorders*, 184, pp.198–204. [Online]. Available at: doi:10.1016/j.jad.2015.05.064.

Geraghty, A. W. A. et al. (2019). 'I mean what is depression?' A qualitative exploration of UK general practitioners' perceptions of distinctions between emotional distress and depressive disorder. *BMJ Open*, 9 (12). [Online]. Available at: doi:10.1136/bmjopen-2019-032644.

Gilbody, S. et al. (2015). Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): Large scale pragmatic randomised controlled trial. *The BMJ*, 351, pp.1–13. [Online]. Available at: doi:10.1136/bmj.h5627.

Gilbody, S. et al. (2017). Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. *British Journal of Psychiatry*, 210 (5), pp.362–367. [Online]. Available at: doi:10.1192/bjp.bp.116.192435.

Gili, M. et al. (2015). Interventions for preventing relapse or recurrence of depression in primary health care settings: A systematic review. *Preventive Medicine*, 76 (S), pp.S16–S21. [Online]. Available at: doi:10.1016/j.ypmed.2014.07.035.

Girard, A. et al. (2017). Primary care nursing activities with patients affected by physical chronic disease and common mental disorders: a qualitative

descriptive study. *Journal of Clinical Nursing*, 26 (9–10), pp.1385–1394. [Online]. Available at: doi:10.1111/jocn.13695.

Global Burden of Disease 2017. (2018). Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990– 2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392 (10159), pp.1789–1858.

Glue, P. et al. (2010). Meta-analysis of relapse prevention antidepressant trials in depressive disorders. *Australian and New Zealand Journal of Psychiatry*, 44 (8), pp.697–705. [Online]. Available at: doi:10.3109/00048671003705441.

Goldman, L. S. et al. (1999). Awareness, diagnosis, and treatment of depression. *Journal of General Internal Medicine*, 14 (9), pp.569–580. [Online]. Available at: doi:10.1046/j.1525-1497.1999.03478.x.

Gopal, D. P. and Mulla, E. (2020). The GP crisis: a tale of two peaks. *British Journal of General Practice*, 70 (690), p.24.

Gough, B. and Madill, A. (2012). Subjectivity in psychological science: From problem to prospect. *Psychological Methods*, 17 (3), pp.374–384. [Online]. Available at: doi:10.1037/a0029313.

De Graaf, L. E. et al. (2011). One-year follow-up results of unsupported online computerized cognitive behavioural therapy for depression in primary care: A randomized trial. *Journal of Behavior Therapy and Experimental Psychiatry*, 42 (1), pp.89–95. [Online]. Available at: doi:10.1016/j.jbtep.2010.07.003.

Grassi, L. et al. (2000). Social Support and Psychological Distress in Primary Care Attenders. *and Psychosomatics*, 69 (2), pp.95–100. [Online]. Available at: doi:10.2307/48510572.

Gray, D. J. P. et al. (2018). Continuity of care with doctors - A matter of life and death? A systematic review of continuity of care and mortality. *BMJ Open*, 8 (6). [Online]. Available at: doi:10.1136/bmjopen-2017-021161.

Green, J. and Thorogood, N. (2018). *Qualitative Methods for Health Research*. Fourth Edition. SAGE Publications Ltd.

Grey, I. et al. (2020). The role of perceived social support on depression and sleep during the COVID-19 pandemic. *Psychiatry Research*, 293. [Online]. Available at: doi:10.1016/j.psychres.2020.113452.

Groenwold, R. H. H. et al. (2016). Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *Journal of Clinical Epidemiology*, 78, pp.90–100. [Online]. Available at: doi:10.1016/j.jclinepi.2016.03.017.

Gross, R. et al. (2007). Primary care physicians' discussion of emotional distress and patient satisfaction. *International Journal of Psychiatry in Medicine*, 37 (3), pp.331–345. [Online]. Available at: doi:10.2190/PM.37.3.i.

Gunn, J. et al. (2006). A systematic review of complex system interventions designed to increase recovery from depression in primary care. *BMC Health Services Research*, 6. [Online]. Available at: doi:10.1186/1472-6963-6-88.

Gunn, J. et al. (2013). A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *Journal of Affective Disorders*, 148 (2–3), pp.338–346. [Online]. Available at: doi:10.1016/j.jad.2012.12.021.

Guthrie, B. et al. (2008). Continuity of care matters. *BMJ*, 337 (a867).

Haggerty, J. L. et al. (2013). The strength of primary care systems. *BMJ (Online)*, 346 (7914). [Online]. Available at: doi:10.1136/bmj.f3777.

Hallen, S. A. M. et al. (2015). Physicians' perceptions of the value of prognostic models: The benefits and risks of prognostic confidence. *Health Expectations*, 18 (6), pp.2266–2277. [Online]. Available at: doi:10.1111/hex.12196.

Hansen, R. et al. (2008). Meta-analysis of major depressive disorder relapse and recurrence with second-generation antidepressants. *Psychiatric Services*, 59 (10), pp.1121–1130. [Online]. Available at: doi:10.1176/ps.2008.59.10.1121.

Hardeveld, F. et al. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, 122 (3), pp.184–191. [Online]. Available at: doi:10.1111/j.1600-0447.2009.01519.x.

Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. eBook. Springer International Publishing Switzerland. [Online]. Available at: doi:10.1177/096228020401300512.

Healy, D. (1997). *The antidepressant era.* Harvard University Press.

Heisig, J. P. and Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. In: *European Sociological Review*. 35 (2). 1 April 2019. Oxford University Press. pp.258–279. [Online]. Available at: doi:10.1093/esr/jcy053.

Hemingway, H. et al. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*, 346 (e5595). [Online]. Available at: doi:10.1136/bmj.e5595.

Hengartner, M. P. (2020). How effective are antidepressants for depression over the long term? A critical review of relapse prevention trials and the issue of withdrawal confounding. *Therapeutic Advances in Psychopharmacology*, 10, p.204512532092169. [Online]. Available at: doi:10.1177/2045125320921694.

Henwood, K. and Pidgeon, N. (1992). Qualitative research and psychological theorizing. *Br J Psychol*, 83 (1), pp.97–111.

Higgins, J. P. T. et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Online)*, 343 (7829), pp.1–9. [Online]. Available at: doi:10.1136/bmj.d5928.

Hill, A. and Freeman, G. (2011). *Promoting Continuity of Care in General Practice*. [Online]. Available at: https://www.rcgp.org.uk/getmedia/bda9fb5b-e52b-454d-af1f-e27f3df4100f/RCGP-Continuity-of-Care.pdf [Accessed 2 May 2023].

Hingorani, A. D. et al. (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ (Online)*, 346 (February), pp.1–9. [Online]. Available at: doi:10.1136/bmj.e5793.

Hoesseini, A. et al. (2020). Head and neck cancer patients' preferences for individualized prognostic information: A focus group study. *BMC Cancer*, 20 (1). [Online]. Available at: doi:10.1186/s12885-020-6554-8.

Hoesseini A, van Leeuwen N and Sewnaik A. (2022). Key Aspects of Prognostic Model Development and Interpretation From a Clinical Perspective. *JAMA Otolaryngol Head Neck Surg*, 148 (2), pp.180–186.

House of Commons Health and Social Care Committee. (2022). *The future of general practice: Fourth Report of Session 2022-23*. [Online]. Available at: doi:https://committees.parliament.uk/publications/30383/documents/176291/default/.

Howell, C. A. et al. (2008). Preventing relapse of depression in primary care: a pilot study of the 'Keeping the blues away' program. *The Medical journal of Australia*, 188 (12 Suppl), pp.10–13.

Hudon, C. et al. (2012). Patient-centered care in chronic disease management: A thematic analysis of the literature in family medicine. *Patient Education and Counseling*, 88 (2), pp.170–176. [Online]. Available at: doi:10.1016/j.pec.2012.01.009.

Hughes, M. and Duffy, C. (2018). Public involvement in health and social sciences research: A concept analysis. *Health Expectations*, 21 (6), pp.1183–1190. [Online]. Available at: doi:10.1111/hex.12825.

Hughes, R. et al. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*. [Online]. Available at: doi:10.1093/ije/dyz032.

Huguet, A. et al. (2013). Judging the quality of evidence in reviews of prognostic factor research : adapting the GRADE framework. *Systematic Reviews*, 2 (71).

Insel, T. R. (2009). Translating Scientific Opportunity Into Public Health Impact. *Archives of General Psychiatry*, 66 (2), p.128. [Online]. Available at: doi:10.1001/archgenpsychiatry.2008.540.

NIHR. (2021). *Briefing notes for researchers - public involvement in NHS, health and social care research*. [Online]. Available at: https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371 [Accessed 3 August 2021].

Iorio, A. et al. (2015). Use of GRADE for assessment of evidence about prognosis : rating confidence in estimates of event rates in broad categories of patients. *BMJ*, pp.1–8. [Online]. Available at: doi:10.1136/bmj.h870.

Israel, J. A. (2010). The impact of residual symptoms in major depression. *Pharmaceuticals*, 3 (8), pp.2426–2440. [Online]. Available at: doi:10.3390/ph3082426.

Jacobson, N. S. and Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology*, 59 (1), pp.12–19. [Online]. Available at: doi:10.1037/0022-006X.59.1.12.

Jeffers, H. and Baker, M. (2016). Continuity of care: Still important in modern-day general practice. *British Journal of General Practice*, 66 (649), Royal College of General Practitioners., pp.396–397. [Online]. Available at: doi:10.3399/bjgp16X686185.

Jefferson, L. et al. (2022). GP wellbeing during the COVID-19 pandemic: a systematic review. *British Journal of General Practice*, 72 (718), Royal College of General Practitioners., pp.E325–E333. [Online]. Available at: doi:10.3399/BJGP.2021.0680.

Jefferson, L. and Holmes, M. (2022). GP workforce crisis: what can we do now? *British Journal of General Practice*, 72 (718), Royal College of General Practitioners., pp.206–207. [Online]. Available at: doi:10.3399/bjgp22X719225.

Jimenez, G. et al. (2021). Revisiting the four core functions (4Cs) of primary care: Operational definitions and complexities. *Primary Health Care Research and Development*, 22. [Online]. Available at: doi:10.1017/S1463423621000669.

Johansson, O., Lundh, L. G. and Bjärehed, J. (2015). 12-Month Outcome and Predictors of Recurrence in Psychiatric Treatment of Depression: A Retrospective Study. *Psychiatric Quarterly*, 86 (3), pp.407–417. [Online]. Available at: doi:10.1007/s11126-015-9341-y.

Johnson, D. R., Scheitle, C. P. and Ecklund, E. H. (2021). Beyond the In-Person Interview? How Interview Quality Varies Across In-person, Telephone, and Skype Interviews. *Social Science Computer Review*, 39 (6), pp.1142–1158. [Online]. Available at: doi:10.1177/0894439319893612.

Johnston, O. et al. (2007). Qualitative study of depression management in primary care: GP and patient goals, and the value of listening. *British Journal of General Practice*, 57 (544), pp.e1–e14.

Jones, D. et al. (2019). Safety netting for primary care: Evidence from a literature review. *British Journal of General Practice*, 69 (678), pp.E70–E79. [Online]. Available at: doi:10.3399/bjgp18X700193.

Judd, L. L., Schettler, P. J. and Rush, A. J. (2016). A brief clinical tool to estimate individual patients' risk of depressive relapse following remission: Proof of concept. *American Journal of Psychiatry*, 173 (11), pp.1140–1146. [Online]. Available at: doi:10.1176/appi.ajp.2016.15111462.

Kamenov, K. et al. (2016). The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: A meta-analysis. *Psychological Medicine*, 47 (3), pp.414–425. [Online]. Available at: doi:10.1017/S0033291716002774.

Kato, M. et al. (2020). Discontinuation of antidepressants after remission with antidepressant medication in major depressive disorder: a systematic review and meta-analysis. *Molecular Psychiatry*. [Online]. Available at: doi:10.1038/s41380-020-0843-0.

Katon, W. et al. (2001). A randomized trial of relapse prevention of depression in primary care. *Archives of General Psychiatry*, 58 (3), pp.241–247. [Online]. Available at: doi:10.1001/archpsyc.58.3.241.

Kaymaz, N. et al. (2008). Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta- analysis of placebo-controlled randomized trials. *Journal of Clinical Psychiatry*, 69 (9), pp.1423–1436.

Kearns, N. P. et al. (2016). Does rumination mediate the relationship between mindfulness and depressive relapse? *Psychology and Psychotherapy: Theory, Research and Practice*, 89 (1), pp.33–49. [Online]. Available at: doi:10.1111/papt.12064.

Kelly, M. (2009). The role of theory in qualitative health research. *Family Practice*, 27 (3), pp.285–290. [Online]. Available at: doi:10.1093/fampra/cmp077.

Van Kempen, G. M. J. et al. (2010). Practice guideline for adult major depressive disorder [15]. *American Journal of Psychiatry*, 151 (4), pp.625–626. [Online]. Available at: doi:10.1176/ajp.151.4.625-b.

Kendler, K. S., Thornton, L. M. and Gardner, C. O. (2000). Stressful Life Events and Previous Episodes in the Etiology of Major Depression in

Women: An Evaluation of the "Kindling" Hypothesis. *Am J Psychiatry*, 157 (8), pp.1243–1251. [Online]. Available at: doi:10.1176/appi.ajp.157.8.1243.

Kendrick, T. and Collinson, S. (2022). Antidepressants and the serotonin hypothesis of depression. *The BMJ*, BMJ Publishing Group. [Online]. Available at: doi:10.1136/bmj.o1993.

Kennedy, N. and Foy, K. (2005). The Impact of Residual Symptoms on Outcome of Major Depression. *Current Psychiatry Reports*, 7, pp.441–446.

Kessler, R. C. et al. (1996). Comorbidity of DSM–III–R Major Depressive Disorder in the General Population: Results from the US National Comorbidity Survey. *British Journal of Psychiatry*, 168 (S30), pp.17–30. [Online]. Available at: doi:10.1192/S0007125000298371.

Kessler, R. C. (2018). The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Current Opinion in Psychiatry*, 31 (1), pp.32–39. [Online]. Available at: doi:10.1097/YCO.0000000000000377.

King et al. (2010). Development and Validation of an International Risk Prediction Algorithm for Episodes of Major Depression in General Practice Attendees. *Arch Gen Psychiatry*, 65 (12), pp.1368–1377.

Klein, N. S. et al. (2018). Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression. *Journal of Psychiatric Research*, 104, pp.1–7. [Online]. Available at: doi:10.1016/J.JPSYCHIRES.2018.06.006 [Accessed 3 December 2019].

Klein, N. S. et al. (2019). Cost-effectiveness, cost-utility and the budget impact of antidepressants versus preventive cognitive therapy with or without tapering of antidepressants. *BJPsych Open*, 5 (1), pp.1–9. [Online]. Available at: doi:10.1192/bjo.2018.81.

Kok, G. D. et al. (2013). Double Trouble: Does Co-Morbid Chronic Somatic Illness Increase Risk for Recurrence in Depression? A Systematic Review. *PLoS ONE*, 8 (3). [Online]. Available at: doi:10.1371/journal.pone.0057510.

Kroenke, K., Spitzer, R. L. and Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16 (9), pp.606–613. [Online]. Available at: doi:10.1046/j.1525-1497.2001.016009606.x.

Krouwel, M., Jolly, K. and Greenfield, S. (2019). Comparing Skype (video calling) and in-person qualitative interview modes in a study of people with irritable bowel syndrome-an exploratory comparative analysis. *BMC Medical Research Methodology*, 19 (1). [Online]. Available at: doi:10.1186/s12874-019-0867-9.

Kuyken, W. et al. (2008). Mindfulness-Based Cognitive Therapy to Prevent Relapse in Recurrent Depression. *Journal of Consulting and Clinical Psychology*, 76 (6), pp.966–978. [Online]. Available at: doi:10.1037/a0013786.

Kuyken, W. et al. (2015). Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): A randomised controlled trial. *The Lancet*, 386 (9988), pp.63–73. [Online]. Available at: doi:10.1016/S0140-6736(14)62222-4.

Kuyken, W. et al. (2016). Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse an individual patient data meta-analysis from randomized trials. *JAMA Psychiatry*, 73 (6), pp.565–574. [Online]. Available at: doi:10.1001/jamapsychiatry.2016.0076.

Lakhan, R., Agrawal, A. and Sharma, M. (2020). Prevalence of Depression, Anxiety, and Stress during COVID-19 Pandemic. *Journal of Neurosciences in Rural Practice*, 11 (4), pp.519–525. [Online]. Available at: doi:10.1055/s-0040-1716442.

Lam, R. W. et al. (2009). Canadian Network for Mood and Anxiety Treatments (CANMAT) Clinical guidelines for the management of major depressive disorder in adults. III. Pharmacotherapy. *Journal of Affective Disorders*, 117 (SUPPL. 1), pp.S26–S43. [Online]. Available at: doi:10.1016/j.jad.2009.06.041.

Langan, D. et al. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, (July 2018), pp.83–98. [Online]. Available at: doi:10.1002/jrsm.1316.

Langdown, C. and Peckham, S. (2014). The use of financial incentives to help improve health outcomes: Is the quality and outcomes framework fit for purpose? A systematic review. *Journal of Public Health (United Kingdom)*, 36 (2), pp.251–258. [Online]. Available at: doi:10.1093/pubmed/fdt077.

Large, M. (2018). The role of prediction in suicide prevention. *Dialogues Clin Neurosci*, 20. [Online]. Available at: www.dialogues-cns.org.

Larsson, P. R. M. et al. (2022). An observational study on IAPT waiting times before, during and after the COVID-19 pandemic using descriptive time-series data. *Mental Health Review Journal*, 27 (4), pp.455–471.

Lee, J. et al. (2021). Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review. *Liver International*, 41 (2), pp.261–270. [Online]. Available at: doi:10.1111/liv.14669.

van Leeuwen, E. et al. (2020). Discontinuation of long-term antidepressant use for depressive and anxiety disorders in adults. *Cochrane Database of Systematic Reviews*, 2020 (2), John Wiley and Sons Ltd. [Online]. Available at: doi:10.1002/14651858.CD013495.

van Leeuwen, E. et al. (2022). Should I, can I, dare I? Patients' view on stopping long-term antidepressant use, a qualitative study. *Acta Clinica Belgica*, 77 (6), pp.962–969. [Online]. Available at: doi:10.1080/17843286.2021.2024384.

Le-Niculescu, H. et al. (2021). Precision medicine for mood disorders: objective assessment, risk prediction, pharmacogenomics, and repurposed drugs. *Molecular Psychiatry*. [Online]. Available at: doi:10.1038/s41380-021-01061-w.

Levitt, H. M. et al. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology*, 4 (1), pp.2–22. [Online]. Available at: doi:10.1037/qup0000082.

Lewis, G. et al. (1992). Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine*, 22 (2), pp.465–486. [Online]. Available at: doi:10.1017/S0033291700030415.

Lewis, G. et al. (2021). Maintenance or Discontinuation of Antidepressants in Primary Care. *New England Journal of Medicine*, 385 (14), pp.1257–1267. [Online]. Available at: doi:10.1056/nejmoa2106356.

Leydon, G. M., Rodgers, L. and Kendrick, T. (2007). A qualitative study of patient views on discontinuing long-term selective serotonin reuptake inhibitors. *Family Practice*, 24 (6), pp.570–575. [Online]. Available at: doi:10.1093/fampra/cmm069.

Lin, E. H. B. et al. (1998). Relapse of Depression in Primary Care: Rate and Clinical Predictors. *Archives of Family Medicine*, 7 (5), pp.443–449. [Online]. Available at: doi:10.1001/archfami.7.5.443.

Lincoln, Y. and Guba, E. (1985). *Naturalistic inquiry*. NewburyPark, CA: Sage.

Lip, G. Y. H. and Lane, D. A. (2015). Assessing bleeding risk in atrial fibrillation with the HAS-BLED and ORBIT scores: Clinical application requires focus on the reversible bleeding risk factors. *European Heart Journal*, 36 (46), Oxford University Press., pp.3265–3267. [Online]. Available at: doi:10.1093/eurheartj/ehv415.

van Loo, H. M. et al. (2015). Multiple risk factors predict recurrence of major depressive disorder in women. *Journal of Affective Disorders*, 180, pp.52–61. [Online]. Available at: doi:10.1016/J.JAD.2015.03.045 [Accessed 3 December 2019].

van Loo, H. M. et al. (2018). Sex similarities and differences in risk factors for recurrence of major depression. *Psychological Medicine*, 48 (10), pp.1685–1693. [Online]. Available at: doi:10.1017/S0033291717003178.

van Loo, H. M. et al. (2020). Data mining algorithm predicts a range of adverse outcomes in major depression. *Journal of Affective Disorders*, 276 (March), pp.945–953. [Online]. Available at: doi:10.1016/j.jad.2020.07.098.

Lorimer, B. et al. (2020). Dynamic prediction and identification of cases at risk of relapse following completion of low-intensity cognitive behavioural therapy. *Psychotherapy Research*. [Online]. Available at: doi:10.1080/10503307.2020.1733127.

Ma, S. H. and Teasdale, J. D. (2004). Mindfulness-Based Cognitive Therapy for Depression: Replication and Exploration of Differential Relapse Prevention Effects. *Journal of Consulting and Clinical Psychology*, 72 (1), pp.31–40. [Online]. Available at: doi:10.1037/0022-006X.72.1.31.

Maanen, J. Van. (1979). Reclaiming Qualitative Methods for Organizational Research : A Preface. *Administrative Science Quarterly*, 24 (4), pp.520–526.

Malterud, K., Siersma, V. D. and Guassora, A. D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*, 26 (13), pp.1753–1760. [Online]. Available at: doi:10.1177/1049732315617444.

Marchetti, I. et al. (2018). The brief state rumination inventory (BSRI): Validation and psychometric evaluation. *Cognitive Therapy and Research*, 42 (4), pp.447–460. [Online]. Available at: doi:10.1007/s10608-018-9901-1.

Martin, C. et al. (2022). Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: How can it better meet the range of primary care mental health needs? *British Journal of Clinical Psychology*, 61 (1), pp.157–174. [Online]. Available at: doi:10.1111/bjc.12314.

Maund, E. et al. (2019). Barriers and facilitators to discontinuing antidepressant use: A systematic review and thematic synthesis. *Journal of*

Affective Disorders, 245 (September 2018), pp.38–62. [Online]. Available at: doi:10.1016/j.jad.2018.10.107.

Mcconnell-Henry, T. et al. (2009). Researching with people you know: Issues in interviewing. *Contemporary Nurse*, 34 (1), pp.2–9.

McIntyre, R. S., Lee, Y. and Mansur, R. B. (2015). Treating to target in major depressive disorder: Response to remission to functional recovery. *CNS Spectrums*, 20 (S1), Cambridge University Press., pp.20–30. [Online]. Available at: doi:10.1017/S1092852915000826.

McManus, S. et al. (2014). Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. Leeds: *Leeds: NHS Digital.*

McMillan, D., Gilbody, S. and Richards, D. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders*, 127 (1–3), pp.122–129. [Online]. Available at: doi:10.1016/j.jad.2010.04.030.

McMullin, C. (2021). Transcription and Qualitative Methods: Implications for Third Sector Research. *Voluntas*. [Online]. Available at: doi:10.1007/s11266-021-00400-3.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and Effects of My Disturbing Little Book. *Journal of Personality Assessment*, 50 (3), pp.370–375.

Mendive, J. (2009). Emotional distress: an alternative primary care perspective. *Mental Health in Family Medicine*, 6, pp.125–127. [Online]. Available at: www.who.int/classifications/apps/.

Mental Health Task Force. (2016). *Five Year Forward View for Mental Health*. [Online]. Available at: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf [Accessed 3 January 2019].

Michalak, J., Hölz, A. and Teismann, T. (2011). Rumination as a predictor of relapse in mindfulness-based cognitive therapy for depression. *Psychology and Psychotherapy: Theory, Research and Practice*, 84 (2), pp.230–236. [Online]. Available at: doi:10.1348/147608310X520166.

Mocking, R. J. T. et al. (2021). Metabolic features of recurrent major depressive disorder in remission, and the risk of future recurrence. *Translational Psychiatry*, 11 (1). [Online]. Available at: doi:10.1038/s41398-020-01182-w.

Moncrieff, J. et al. (2022). The serotonin theory of depression: a systematic umbrella review of the evidence. *Molecular Psychiatry*, Springer Nature. [Online]. Available at: doi:10.1038/s41380-022-01661-0.

Monroe, S. M. and Harkness, K. L. (2005). Life stress, the 'kindling' hypothesis, and the recurrence of depression: Considerations from a life stress perspective. *Psychological Review*, 112 (2), pp.417–445. [Online]. Available at: doi:10.1037/0033-295X.112.2.417.

Monroe, S. M. and Harkness, K. L. (2011). Recurrence in Major Depression: A Conceptual Analysis. *Psychological Review*, 118 (4), pp.655–674. [Online]. Available at: doi:10.1037/a0025190.

Moons, K. G. M. et al. (2012). Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*, 98 (9), pp.691–698. [Online]. Available at: doi:10.1136/heartjnl-2011-301247.

Moons, K. G. M. et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, 162 (1), pp.W1–W73. [Online]. Available at: doi:10.7326/M14-0698.

Moons, K. G. M. et al. (2019). PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Annals of Internal Medicine*, 170 (1), pp.W1–W33. [Online]. Available at: doi:10.7326/M18-1377.

Moore, M. et al. (2009). Explaining the rise in antidepressant prescribing: A descriptive study using the general practice research database. *BMJ (Online)*, 339 (7727), p.956. [Online]. Available at: doi:10.1136/bmj.b3999.

Moore, M. et al. (2012). Depression management in primary care: An observational study of management changes related to PHQ-9 score for depression monitoring. *British Journal of General Practice*, 62 (599). [Online]. Available at: doi:10.3399/bjgp12X649151.

Morgan, I. et al. (2023). The first consultation for low mood in general practice: what do patients find helpful? *Family Practice*. [Online]. Available at: doi:10.1093/fampra/cmad016.

Morgan, M. et al. (2009). The TrueBlue study: Is practice nurse-led collaborative care effective in the management of depression for patients with heart disease or diabetes? *BMC Family Practice*, 10 (1), p.46. [Online]. Available at: doi:10.1186/1471-2296-10-46.

Moriarty, A. et al. (2020). Predicting and preventing relapse of depression in primary care. *British Journal of General Practice*, 70 (693), p.166. [Online]. Available at: doi:10.3399/bjgp20X708977.

Moriarty, A. et al. (2021a). *Predicting depression relapse: is this possible and useful?* [Online]. Available at: https://www.evidentlycochrane.net/predicting-depression-relapse/ [Accessed 8 August 2021].

Moriarty, A. S. et al. (2019). Prognostic models for predicting relapse or recurrence of depression. *Cochrane Database of Systematic Reviews*, 2019 (12). [Online]. Available at: doi:10.1002/14651858.CD013491.

Moriarty, A. S. et al. (2021b). Interventions for preventing relapse or recurrence of major depressive disorder in adults in a primary care setting: a network meta-analysis. *Cochrane Database of Systematic Reviews*, 2021 (11). [Online]. Available at: doi:10.1002/14651858.CD014832.

Moriarty, A. S. et al. (2021c). Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. *Cochrane Database of Systematic Reviews*, 2021 (5). [Online]. Available at: doi:10.1002/14651858.CD013491.pub2.

Moriarty, A. S. et al. (2021d). The development and validation of a prognostic model to PREDICT Relapse of depression in adult patients in primary care: protocol for the PREDICTR study. *Diagnostic and Prognostic Research*, 5 (1). [Online]. Available at: doi:10.1186/s41512-021-00101-x.

Moriarty, A. S. et al. (2022). Predicting relapse or recurrence of depression: Systematic review of prognostic models. *British Journal of Psychiatry*, Cambridge University Press. [Online]. Available at: doi:10.1192/bjp.2021.218.

Moussavi, S. et al. (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*, 370, pp.851–858. [Online]. Available at: doi:10.3390/en10010025.

Mughal, F. et al. (2021). Experiences of general practice care for self-harm: A qualitative study of young people's perspectives. *British Journal of General Practice*, 71 (711), pp.E744–E752. [Online]. Available at: doi:10.3399/BJGP.2021.0091.

Muntingh, A. D. T. et al. (2019). Patient preferences for a guided self-help programme to prevent relapse in anxiety or depression: A discrete choice experiment. *PLoS ONE*, 14 (7). [Online]. Available at: doi:10.1371/journal.pone.0219588.

Murphy, E. et al. (1998). Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess.*, 2 (16), pp.1–274.

Murphy, M. et al. (2021). Implementation of remote consulting in UK primary care following the COVID-19 pandemic: A mixed-methods longitudinal study. *British Journal of General Practice*, 71 (704), pp.E166–E177. [Online]. Available at: doi:10.3399/BJGP.2020.0948.

Murphy, R., Ekers, D. and Webster, L. (2014). An update to depression case management by practice nurses in primary care: A service evaluation. *Journal of Psychiatric and Mental Health Nursing*, 21 (9), pp.827–833. [Online]. Available at: doi:10.1111/jpm.12153.

Musgrave, A. (2011). Popper and Hypothetico-deductivism. *Handbook of the History of Logic*, 10, pp.205–234.

Nanni, V., Uher, R. and Danese, A. (2012). Childhood Maltreatment Predicts Unfavorable Course of illness and treatment outcome in Depression: A Meta-Analysis. *Am J Psychiatry*, 169, pp.141–151.

National Health Service. (2019). *NHS Mental Health Implementation Plan 2019/20 – 2023/24*. [Online]. Available at: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/07/nhs-mental-health-implementation-plan-2019-20-2023-24.pdf [Accessed 11 April 2023].

NHS Confederation. (2021). *Supporting population health management: workforce Recruiting mental health practitioners through the Additional Roles Reimbursement Scheme (ARRS)*. [Online]. Available at: https://www.nhsconfed.org/publications/recruiting-mental-health-practitioners-through-additional-roles-reimbursement-scheme [Accessed 10 April 2023].

NHS England and NHS Improvement. (2019). *Network Contract Directed Enhanced Service: Additional Roles Reimbursement Scheme Guidance*. [Online]. Available at: https://www.england.nhs.uk/wp-content/uploads/2019/12/network-contract-des-additional-roles-reimbursement-scheme-guidance-december2019.pdf [Accessed 28 May 2023].

NICE. (2009). Depression in adults: recognition and management. *NICE Guideline (CG 90)*, (July). [Online]. Available at: http://www.nice.org.uk/guidance/cg90.

NICE. (2016). *Multimorbidity: clinical assessment and management*. (September 2016).

NICE. (2022). *Depression in adults: treatment and management NICE guideline (NG222)*. [Online]. Available at: www.nice.org.uk/guidance/ng222.

Nowak, D. et al. (2021). Why does continuity of care with family doctors matter? Review and qualitative synthesis of patient and physician

perspectives. *Canadian Family Physician*, 67, pp.679–688. [Online]. Available at: www.cfp.ca.

Nowell, L. S. et al. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16 (1). [Online]. Available at: doi:10.1177/1609406917733847.

Nussbaum, C. et al. (2021). Inequalities in the distribution of the general practice workforce in England: A practice-level longitudinal analysis. *BJGP Open*, 5 (5), pp.1–10. [Online]. Available at: doi:10.3399/BJGPO.2021.0066.

O'Cathain, A., Murphy, E. and Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *BMJ (Online)*, 341 (7783), pp.1147–1150. [Online]. Available at: doi:10.1136/bmj.c4587.

Oh, S. et al. (2020). Exploring women's experiences of identifying, negotiating and managing perinatal anxiety: A qualitative study. *BMJ Open*, 10 (12). [Online]. Available at: doi:10.1136/bmjopen-2020-040731.

O'Neill, C. et al. (2020). Prognostic prediction tools and clinician communication: A qualitative study of the effect of the STUMBL tool on clinical practice. *BMC Emergency Medicine*, 20 (1). [Online]. Available at: doi:10.1186/s12873-020-00331-0.

Ormel, J. et al. (2020). The antidepressant standoff: Why it continues and how to resolve it. *Psychological Medicine*, 50 (2), pp.177–186. [Online]. Available at: doi:10.1017/S0033291719003295.

Page, M. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372 (N71). [Online]. Available at: doi: 10.1136/bmj.n71.

Pajouheshnia, R. et al. (2017). Accounting for treatment use when validating a prognostic model: A simulation study. *BMC Medical Research Methodology*, 17 (1), pp.1–12. [Online]. Available at: doi:10.1186/s12874-017-0375-8.

Pajouheshnia, R. et al. (2019). When and how to use data from randomised trials to develop or validate prognostic models. *Bmj*, p.l2154. [Online]. Available at: doi:10.1136/bmj.l2154.

Paykel, E. et al. (1995). Residual symptoms after partial remission: An important outcome in depression. *Psychological Medicine*, 25 (6), pp.1171–1180.

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience*, 10 (4), pp.431–437. [Online]. Available at: doi:10.31887/dcns.2008.10.4/espaykel.

Pfefferbaum, B. and North, C. (2020). Mental Health and the Covid-19 Pandemic. *New England Journal of Medicine*, 383 (6), pp.510–512. [Online]. Available at: doi:10.1056/nejmp2013466.

Pintor, L. et al. (2009). Prediction of relapse in melancholic depressive patients in a 2-year follow-up study with corticotropin releasing factor test. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 33 (3), pp.463–469. [Online]. Available at: doi:10.1016/J.PNPBP.2009.01.008.

van Der Ploeg, T., Austin, P. C. and Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology*, 14 (1). [Online]. Available at: doi:10.1186/1471-2288-14-137.

Popper, K. (1959). *The Logic of Scientific Discovery*. New York: BasicBooks.

Post, M. (1992). Transduction of Psychosocial Stress into the Neurobiology of Recurrent Affective Disorder. *Depression*, (August), pp.999–1010.

Rait, G. et al. (2009). Recent trends in the incidence of recorded depression in primary care. *British Journal of Psychiatry*, 195 (6), pp.520–524. [Online]. Available at: doi:10.1192/bjp.bp.108.058636.

Ramanuj, P., Ferenchick, E. K. and Pincus, H. A. (2019). Depression in primary care: part 2 -management. *BMJ (Clinical research ed.)*, 365, p.l835. [Online]. Available at: doi:10.1136/bmj.l835.

Rawaf, S. et al. (2020). Lessons on the COVID-19 pandemic, for and by primary care professionals worldwide. *European Journal of General Practice*, 26 (1), pp.129–133. [Online]. Available at: doi:10.1080/13814788.2020.1820479.

Read, J. et al. (2020). A survey of UK general practitioners about depression, antidepressants and withdrawal: implementing the 2019 Public Health England report. *Therapeutic Advances in Psychopharmacology*, 10, pp.1–14. [Online]. Available at: doi:10.1177/2045125320950124.

Recalt, A. and Cohen, D. (2019). Withdrawal confounding in randomized controlled trials of antipsychotic, antidepressant, and stimulant drugs. *Psychotherapy and Psychosomatics*, 88 (2), pp.105–113.

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review*, 31 (7), pp.1117–1125. [Online]. Available at: doi:10.1016/j.cpr.2011.07.004.

Richards, D. et al. (2013). Clinical effectiveness of collaborative care for depression in UK primary care (CADET): Cluster randomised controlled trial. *BMJ*, 347 (f4913). [Online]. Available at: doi:10.1136/bmj.f4913.

Richards, D. A. et al. (2016). Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. *The Lancet*, 388 (10047), pp.871–880. [Online]. Available at: doi:10.1016/S0140-6736(16)31140-0.

Ridd, M., Shaw, A. and Salisbury, C. (2006). 'Two sides of the coin' - The value of personal continuity to GPs: A qualitative interview study. *Family Practice*, 23 (4), pp.461–468. [Online]. Available at: doi:10.1093/fampra/cml010.

Riihimäki, K. et al. (2023). Health-related quality of life-based definition of remission from depression among primary care patients. *Frontiers in Psychiatry*, 14. [Online]. Available at: doi:10.3389/fpsyt.2023.926410.

Riley, R. D. et al. (2018). Minimum sample size for developing a multivariable prediction model: Part II - binary and time-to-event outcomes. *Statistics in Medicine*, (February), pp.1–21. [Online]. Available at: doi:10.1002/sim.7992.

Riley, R. D. et al. (2019a). Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. *Statistics in Medicine*, 38 (7), pp.1262–1275. [Online]. Available at: doi:10.1002/sim.7993.

Riley, R. D. et al. (2019b). *Prognosis Research in Healthcare: Concepts, Methods, and Impact.* First edit. Oxford University Press.

Riley, R. D. et al. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, p.m441. [Online]. Available at: doi:10.1136/bmj.m441.

Riley, R. D. et al. (2021). Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small.

Journal of Clinical Epidemiology, 132, pp.88–96. [Online]. Available at: doi:10.1016/j.jclinepi.2020.12.005.

Riso, L. P. et al. (1997). A prospective test of criteria for response, remission, relapse, recovery, and recurrence in depressed patients treated with cognitive behavior therapy. *Journal of Affective Disorders*, 43 (2), pp.131–142. [Online]. Available at: doi:10.1016/S0165-0327(96)01420-6 [Accessed 7 December 2019].

Royal College of General Practitioners. (2016). *Continuity of care in modern day general practice*. [Online]. Available at: https://www.rcgp.org.uk/getmedia/11f26527-5d11-47f2-a593-1a894c2fff1b/Continuity-of-care-in-modern-day-general-practice1.pdf [Accessed 2 May 2023].

Royal College of General Practitioners. (2019). *RCGP Guidelines for Continuity of Care-Helping clinicians and practices maximise relationship continuity*. [Online]. Available at: http://www.gp-patient.co.uk.

Royal College of General Practitioners. (2023). *Fit for the Future: GP Pressures 2023*. [Online]. Available at: https://www.rcgp.org.uk/representing-you/future-of-general-practice [Accessed 2 May 2023].

Royston, P. and Altman, D. G. (1994). Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43 (3), pp.429–467. [Online]. Available at: https://www.jstor.org/stable/2986270.

Royston, P., Parmar, M. K. B. and Sylvester, R. (2004). Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Statistics in Medicine*, 23 (6), pp.907–926. [Online]. Available at: doi:10.1002/sim.1691.

Rubenstein, L. V. et al. (2007). Predicting Outcomes of Primary Care Patients With Major Depression: Development of a Depression Prognosis Index. *Psychiatric Services*, 58 (8), pp.1049–1056. [Online]. Available at: doi:10.1176/ps.2007.58.8.1049.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, Inc.

Ruhe, H. G. et al. (2019). Emotional biases and recurrence in major depressive disorder. Results of 2.5 years follow-up of drug-free cohort vulnerable for recurrence. *Frontiers in Psychiatry*, 10 (MAR), pp.1–18. [Online]. Available at: doi:10.3389/fpsyt.2019.00145.

Rush, A. J. et al. (2006). Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*, 31 (9), pp.1841–1853. [Online]. Available at: doi:10.1038/sj.npp.1301131.

Rycroft-Malone, J. et al. (2012). Realist synthesis: illustrating the method for implementation research. *Implementation Science*, 7 (1). [Online]. Available at: doi:10.1186/1748-5908-7-33.

Salari, N. et al. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health*, 16 (1), BioMed Central. [Online]. Available at: doi:10.1186/s12992-020-00589-w.

Salisbury, C. et al. (2016). Effectiveness of an integrated telehealth service for patients with depression: A pragmatic randomised controlled trial of a complex intervention. *The Lancet Psychiatry*, 3 (6), pp.515–525. [Online]. Available at: doi:10.1016/S2215-0366(16)00083-3.

Sandelowski, M. (1995). Focus on Qualitative Methods Sample Size in Qualitative Research. *Research in Nursing & Health*, 18, pp.179–183.

Schwartz, A. V et al. (2011). Association of BMD and FRAX Score With Risk of Fracture in Older Adults With Type 2 Diabetes. *JAMA*, 305 (21), pp.2184–2192. [Online]. Available at: https://jamanetwork.com/.

Sedgwick, P. (2013). Snowball sampling. *BMJ (Online)*, 347. [Online]. Available at: doi:10.1136/bmj.f7511.

Segal, Z. V., Pearson, J. L. and Thase, M. E. (2003). Challenges in preventing relapse in major depression: Report of a National Institute of Mental Health Workshop on state of the science of relapse prevention in major depression. *Journal of Affective Disorders*, 77, pp.97–108. [Online]. Available at: doi:10.1016/S0165-0327(02)00112-X.

Sekhon, M., Cartwright, M. and Francis, J. J. (2017). Acceptability of healthcare interventions: An overview of reviews and development of a theoretical framework. *BMC Health Services Research*, 17 (1), pp.1–13. [Online]. Available at: doi:10.1186/s12913-017-2031-8.

Shaw, E. J. et al. (2013). Assessing depression severity using the uk quality and outcomes framework depression indicators: A systematic review. *British Journal of General Practice*, 63 (610). [Online]. Available at: doi:10.3399/bjgp13X667169.

Silveira, P. C. et al. (2015). Performance of wells score for deep vein thrombosis in the inpatient setting. *JAMA Internal Medicine*, 175 (7), pp.1112–1117. [Online]. Available at: doi:10.1001/jamainternmed.2015.1687.

Van Smeden, M. et al. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, 16 (1), pp.1–12. [Online]. Available at: doi:10.1186/s12874-016-0267-3.

Smith, B. (2018). Generalizability in qualitative research: misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health*, 10 (1), pp.137–149. [Online]. Available at: doi:10.1080/2159676X.2017.1393221.

Smith, C. T. et al. (2015). Good practice principles for sharing individual participant data from publicly funded clinical trials. *Trials*, 16 (S2). [Online]. Available at: doi:10.1186/1745-6215-16-s2-o1.

Smith, J. et al. (2013). *Securing the future of general practice: new models of primary care*. [Online]. Available at: www.nuffieldtrust.org.uk/publications/securing-future-general-practice.

Snell, K. I. E. et al. (2018). Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? *Statistical Methods in Medical Research*, 27 (11), pp.3505–3522. [Online]. Available at: doi:10.1177/0962280217705678.

Sofaer, S. (1999). Qualitative methods: what are they and why use them? *Health services research*, 34 (5), p.1101. [Online]. Available at: doi:10.1557/proc-756-ee5.9.

Solomon, D. A. et al. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry*, 157 (2), pp.229–233. [Online]. Available at: doi:10.1176/appi.ajp.157.2.229.

Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application*, 4, Annual Reviews Inc., pp.31–60. [Online]. Available at: doi:10.1146/annurev-statistics-010814-020148.

Spitzer, R. L. et al. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166 (10), pp.1092–1097. [Online]. Available at: doi:10.1001/archinte.166.10.1092.

Starfield, B., Shi, L. and Macinko, J. (2005). Contribution of primary care to health systems and health. *Milbank Q*, 83 (3), pp.457–502.

Steinert, C. et al. (2014). The prospective long-term course of adult depression in general practice and the community. A systematic literature review. *Journal of Affective Disorders*, 152–154, pp.65–75. [Online]. Available at: doi:10.1016/J.JAD.2013.10.017 [Accessed 8 December 2019].

Sterne, J. A. C. et al. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *The BMJ*, 366. [Online]. Available at: doi:10.1136/bmj.l4898.

Stewart, L. A. et al. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement. *JAMA*, 313 (16), pp.1657–1665. [Online]. Available at: doi:10.1001/jama.2015.3656.

Steyerberg, E. et al. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*, 10(2): e10. [Online]. Available at: doi:10.1371/jour.

Steyerberg, E. W. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Second Edi. Springer Nature Switzerland AG.

Steyerberg, E. W. and Harrell, F. E. (2016). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology*, 69, Elsevier USA., pp.245–247. [Online]. Available at: doi:10.1016/j.jclinepi.2015.04.005.

Tabachnick, B. G. and L. S. Fidell. (1996). *Using Multivariate Statistics.* Harper Collins College Publishers: New York.

Takamine, L. et al. (2021). Understanding providers' attitudes and key concerns toward incorporating CVD risk prediction into clinical practice: a qualitative study. *BMC Health Services Research*, 21 (1). [Online]. Available at: doi:10.1186/s12913-021-06540-y.

Tariq, S. and Woodman, J. (2013). Using mixed methods in health research. *JRSM Short Reports*, 4 (6), p.204253331347919. [Online]. Available at: doi:10.1177/2042533313479197.

Tate, A. et al. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLoS One*, 14 (4), p.e0230389.

Thase, M. et al. (1992). Relapse after cognitive behavior therapy of depression: potential implications for longer courses of treatment. *American Journal of Psychiatry*, 149 (8), pp.1046–1052.

The King's Fund. (2020). *GP funding and contracts explained*. [Online]. Available at: https://www.kingsfund.org.uk/publications/gp-funding-and-contracts-explained#:~:text=The%20funding%20a%20general%20practice,where%20they%20have%20been%20agreed. [Accessed 10 December 2021].

Thomas, D. R. (2017). Feedback from research participants: are member checks useful in qualitative research? *Qualitative Research in Psychology*,

14 (1), pp.23–41. [Online]. Available at: doi:10.1080/14780887.2016.1219435.

Tickell, A. et al. (2020). Recovery from recurrent depression with mindfulness-based cognitive therapy and antidepressants: A qualitative study with illustrative case studies. *BMJ Open*, 10 (2). [Online]. Available at: doi:10.1136/bmjopen-2019-033892.

Tiffin, P. A. and Paton, L. W. (2018). Rise of the machines? Machine learning approaches and mental health: opportunities and challenges. *The British Journal of Psychiatry*, 213 (3), pp.509–510. [Online]. Available at: doi:10.1192/bjp.2018.105.

Tracy, S. J. (2010). Qualitative quality: Eight a"big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, 16 (10), pp.837–851. [Online]. Available at: doi:10.1177/1077800410383121.

Trainor, L. R. and Bundon, A. (2021). Developing the craft: reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 13 (5), pp.705–726. [Online]. Available at: doi:10.1080/2159676X.2020.1840423.

Trivedi, M. et al. (2016). Biobehavioral prediction of relapse in major depression: A prospective, multicenter, observational study. In: *Neuropsychopharmacology;41 (Supplement 1)():S517-S518*. 2016.

Trivedi, N. et al. (2021). Recovery, restoration, and risk: a cross-sectional survey of the impact of COVID-19 on GPs in the first UK city to lock down. *BJGP Open*, 5 (1), pp.1–9. [Online]. Available at: doi:10.3399/BJGPO.2020.0151.

Tylee, A. and Walters, P. (2007). We need a chronic disease management model for depression in primary care. *British Journal of General Practice*, 57 (538), pp.348–349.

Udo, C. et al. (2019). An interview study of the care manager function—Opening the door to continuity of care for patients with depression in primary care. *Nursing Open*, 6 (3), pp.974–982. [Online]. Available at: doi:10.1002/nop2.277.

Uijen, A. A. et al. (2014). Experienced continuity of care in patients at risk for depression in primary care. *European Journal of General Practice*, 20 (3), pp.161–166. [Online]. Available at: doi:10.3109/13814788.2013.828201.

Verma, P. and Kerrison, R. (2022). Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. *BJGP Open*, 6 (2). [Online]. Available at: doi:10.3399/BJGPO.2021.0192.

Vickers, A. J., Van Calster, B. and Steyerberg, E. W. (2016). Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ (Online)*, 352, pp.3–7. [Online]. Available at: doi:10.1136/bmj.i6.

Vindegaard, N. and Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain, Behavior, and Immunity*, 89, Academic Press Inc., pp.531–542. [Online]. Available at: doi:10.1016/j.bbi.2020.05.048.

Vittengl, J. et al. (2007). Reducing Relapse and Recurrence in Unipolar Depression: A Comparative Meta-Analysis of Cognitive–Behavioral Therapy's Effects. *J Consult Clin Psychol*, 75 (3), pp.475–488. [Online]. Available at: doi:10.1037/0022-006X.75.3.475.Reducing.

Wakefield, J. C. (2007). The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6, pp.149–156.

Walsh, C. G. et al. (2021). Risky business: a scoping review for communicating results of predictive models between providers and patients. *JAMA Open*, 4 (4). [Online]. Available at: doi:10.1093/jamiaopen/ooab092.

Wanat, M. et al. (2021). Transformation of primary care during the COVID-19 pandemic: Experiences of healthcare professionals in eight European countries. *British Journal of General Practice*, 71 (709). [Online]. Available at: doi:10.3399/BJGP.2020.1112.

Wang, J. L. et al. (2014). Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. *Depression and Anxiety*, 31 (5), pp.451–457. [Online]. Available at: doi:10.1002/da.22215.

Webster, L. A. D., Ekers, D. and Chew-Graham, C. A. (2016). Feasibility of training practice nurses to deliver a psychosocial intervention within a collaborative care framework for people with depression and long-term conditions. *BMC Nursing*, 15 (1), pp.1–11. [Online]. Available at: doi:10.1186/s12912-016-0190-2.

van Weel-Baumgarten, E. M. et al. (2000). Long-term follow-up of depression among patients in the community and in family practice settings. *Journal of Family Practice*, 49 (12), pp.1113–1120.

White, I. R., Royston, P. and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30 (4), pp.377–399. [Online]. Available at: doi:10.1002/sim.4067.

World Health Organisation. (2017). Weekly Epidemiological Report. *World Health Organiszation*, 30 (30). [Online]. Available at: doi:10.1371/jour.

Wilkinson, S. (1988). The role of reflexivity in feminist psychology. *Women's Studies Int. Forum*, II (5), pp.493–502.

Williams, J. M. G. and Kuyken, W. (2012). Mindfulness-based cognitive therapy: A promising new approach to preventing depressive relapse. *British Journal of Psychiatry*, 200 (5), pp.359–360. [Online]. Available at: doi:10.1192/bjp.bp.111.104745.

Willig, C. (2008). *Introducing qualitative research in psychology: adventures in theory and method*. Second Edition. Open University Press.

Wojnarowski, C. et al. (2019). Predictors of depression relapse and recurrence after cognitive behavioural therapy: a systematic review and meta-analysis. *Behavioural and Cognitive Psychotherapy*, 47 (5), pp.514–529. [Online]. Available at: doi:10.1017/S1352465819000080 [Accessed 22 March 2020].

Wolff, R. F. et al. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, 170 (1), p.51. [Online]. Available at: doi:10.7326/M18-1376.

World Health Organization. (2015). *People-Centred and Integrated Health Services: An Overview of the Evidence: Interim Report*. Geneva, Switzerland.

World Health Organization. (2019). *International Classification of Diseases (ICD-11)*. Eleventh revision.

Wynants, L. et al. (2015). A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction

models in clustered data. *Journal of Clinical Epidemiology*, 68 (12), pp.1406–1414. [Online]. Available at: doi:10.1016/j.jclinepi.2015.02.002.

Yarborough, B. J. H. et al. (2022). Clinical implementation of suicide risk prediction models in healthcare: a qualitative study. *BMC Psychiatry*, 22 (1). [Online]. Available at: doi:10.1186/s12888-022-04400-5.

Yardley, L. and Bishop, F. L. (2015). Using mixed methods in health research: Benefits and challenges. *British Journal of Health Psychology*, 20, pp.1–4. [Online]. Available at: doi:10.1111/bjhp.12126.

Yeung, A. S. et al. (2012). Clinical Outcomes in Measurement-based Treatment (COMET): A trial of depression monitoring and feedback to primary care physicians. *Depression and Anxiety*, 29 (10), pp.865–873. [Online]. Available at: doi:10.1002/da.21983.

Yoon, K. L., Maltby, J. and Joormann, J. (2013). A pathway from neuroticism to depression: examining the role of emotion regulation. *Anxiety, Stress and Coping*, 26 (5), pp.558–572. [Online]. Available at: doi:10.1080/10615806.2012.734810.

Zanardi, R. et al. (2020). Precision psychiatry in clinical practice. *International Journal of Psychiatry in Clinical Practice*, 25 (1), pp.19–27. [Online]. Available at: doi:10.1080/13651501.2020.1809680.

Zhu, W. et al. (2015). The HAS-BLED score for predicting major bleeding risk in anticoagulated patients with atrial fibrillation: A systematic review and meta-analysis. *Clinical Cardiology*, 38 (9), John Wiley and Sons Inc., pp.555–561. [Online]. Available at: doi:10.1002/clc.22435.

Zhu, W. G., Xiong, Q. M. and Hong, K. (2015). Meta-analysis of CHADS2 versus CHA2DS2-VASc for predicting stroke and thromboembolism in atrial fibrillation patients independent of anticoagulation. *Texas Heart Institute Journal*, 42 (1), Texas Heart Institute., pp.6–15. [Online]. Available at: doi:10.14503/THIJ-14-4353.

Zimet, G. D. et al. (1990). Psychometric Characteristics of the Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 55 (3–4), pp.610–617. [Online]. Available at: doi:10.1080/00223891.1990.9674095.

Zimmerman, M. et al. (2006). How Should Remission From Depression Be Defined? The Depressed Patient's Perspective. *Am J Psychiatry*, 163. [Online]. Available at: http://ajp.psychiatryonline.org.

Zimmerman, M. et al. (2008). Remission in depressed outpatients: More than just symptom resolution? *Journal of Psychiatric Research*, 42 (10), pp.797–801. [Online]. Available at: doi:10.1016/j.jpsychires.2007.09.004.

Zimmerman, M. et al. (2012). Why do some depressed outpatients who are not in remission according to the Hamilton depression rating scale nonetheless consider themselves to be in remission? *Depression and Anxiety*, 29 (10), pp.891–895. [Online]. Available at: doi:10.1002/da.21987.

De Zwart, P. L., Jeronimus, B. F. and De Jonge, P. (2019). Empirical evidence for definitions of episode, remission, recovery, relapse and recurrence in depression: A systematic review. *Epidemiology and Psychiatric Sciences*, 28 (5), pp.544–562. [Online]. Available at: doi:10.1017/S2045796018000227.

Appendices

NHS

National Institute for Health Research

NIHR Trainees Coordinating Centre
Leeds Innovation Centre
103 Clarendon Road
Leeds LS2 9DF

Tel: 0113 346 6260

www.nihr.ac.uk

Email: TCCawards@nihr.ac.uk

Dr Andrew Moriarty
University of York
Department of Health Sciences
ARCC Building
Heslington
York
United Kingdom (England)
YO10 5DD

24 July 2018

Dear Dr Moriarty,

NIHR Doctoral Research Fellowship - Stage 2
Award Ref: DRF-2018-11-ST2-044
Award Title: Depression in primary care: Development and validation of a prognostic model to predict relapse and evaluation of relapse prevention interventions

I am pleased to inform you that the NIHR Doctoral Research Fellowship - Stage 2 Panel has recommended your application for funding, and the Department of Health, in their capacity as the National Institute for Health Research (NIHR), has confirmed their intention to award funding. This is based upon acceptance of the terms and conditions set out in the Standard Research Contract (link below), and pending agreement to any conditions set by the Panel, details of which will have been communicated to you where relevant.

The Standard Research Contract, between Host Organisations and the Secretary of State for Health can be found on the NIHR website at:

https://www.nihr.ac.uk/funding-and-support/funding-for-training-and-career-development/trainingprogrammes/useful-information-for-applicants/.

Next Steps

The NIHR is committed to the rapid initiation of research following the decision to fund, in order to ensure that the benefit to patients is realised as soon as possible. Therefore, we expect you and your host organisation to be working towards gaining the necessary contractual agreements and governance approvals required to start the award by 01 October 2018.

The NIHR acknowledges the risk to organisations around committing resource to research before a contract is in place; however, it is rare to not reach contractual terms unless the circumstances of the awardee change. The NIHR, therefore, encourages organisations to commit staff to setting up projects at as early an opportunity as possible, in order to expedite the formal commencement of research.

It is acknowledged that there can be unforeseen delays in starting up a research project, but in order to help reduce these it is your responsibility, with the support of your host organisation, to work closely with your organisation's R&D department or equivalent as well as other colleagues / departments involved in the administration and management of the research, and to start these discussions at the earliest opportunity.

To ensure that the award starts within the agreed timeframe, with all the required agreements and approvals in place, you as the awardee, and where necessary any other appropriate shared or support staff, need to be in post to enable the award to start on 01 October 2018. These staff costs will ultimately be covered through the research funding award, but your organisation is encouraged to meet them, where possible, from Research Capability Funding (RCF) prior to the research contract being agreed.

To support the often-iterative process towards agreement of the contract, we have set out the guiding timeframes for the submission of responses or information for each step towards the agreement of the Standard Research Contract as well as the anticipated start date.

- Confirmation of acceptance of funding – 2 weeks from the date of this letter
- Responses to Finance and IP queries – 2 weeks following the issue of queries
- Submission of draft collaboration agreements and/or subcontracts (where applicable) – 1 month prior to the contract start date
- Contract signature – 2 weeks prior to the contract start date
- Contracted start date – 01 October 2018

On receipt of information as set out above, the NIHR through the Trainees Coordinating Centre is committed to responding to your submission of information within two weeks or we will update you on progress.

Please take the time to carefully read the enclosures to this letter which detail the feedback on your application, your contact, Rebecca Savage, within the Trainees Coordinating Centre who will be working with you on the contract, the processes to be undertaken during the next steps, as well as additional information relating to your award.

Yours sincerely

L Cotterill

                               Dr Lisa Cotterill

Director

Enclosures:

- Feedback – a summary of your reviews is provided separately.
- Trainee guidance which can be accessed here: https://www.nihr.ac.uk/funding-andsupport/documents/Current%20Trainee-Guidance-PersonalAwardsJuly15-original.pdf - Information for Themed Call applications. (Annex A).
- Additional information to support the contracting process. (Annex B).
- Ensuring publication of NIHR funded research (Annex C).
- CRN (Annex D).
- Your contact at NIHR (Annex E).
- Accepting your award offer (Annex F).

Editorials
Predicting and preventing relapse of depression in primary care

INTRODUCTION

Depression is now the leading cause of disability worldwide.[1] The majority of people with depression are managed in primary care.[2] There has been a shift in the understanding of depression as a discrete or episodic illness to being considered a long-term relapsing-remitting condition with possibly incomplete recovery between episodes for some patients. The literature draws a distinction between relapse (the re-emergence of depressive symptoms following some level of remission, but preceding full recovery) and recurrence (the onset of a new episode of depression following recovery), recurrence rates being lower than relapse rates.[3] This dichotomy may be more important to researchers and clinicians than it is to patients, who are likely to be less concerned with terminology and more concerned by the risk of 'becoming unwell again' and what can be done to reduce this risk.

After treatment of the first episode of depression, approximately half of all patients will relapse, and this risk increases for every subsequent episode (70% and 90% after a second and third episode respectively).[4] A recent study of a cohort of patients who had received psychosocial treatment through the Improving Access to Psychological Therapies (IAPT) service in England showed that, of those who relapse, the majority (79%) do so within the first 6 months.[5]

There is also evidence to suggest that the severity of depression and resistance to treatment increases with each successive episode,[6] so there are potential benefits of providing on-going care following remission, perhaps after the first episode, to prevent relapse and improve overall disease trajectory. This editorial examines the current evidence around relapse prevention in primary care before discussing the case for improved risk-stratification of patients and the implications that this would have for clinical practice.

CAN RELAPSE BE PREVENTED?

There are few studies looking at relapse prevention strategies specifically in a primary care setting;[7] the vast majority of studies looking at relapse have been undertaken in secondary care. During the development of the most recent update to the *Depression Guideline*, the National Institute for Health and Care Excellence (NICE) recommends that work be done to identify individuals at increased risk of relapse and provide relapse prevention strategies for these individuals.[8]

Current relapse prevention interventions recommended by NICE are a minimum of 2 years treatment with antidepressant medication for patients who have had two or more episodes of depression; high-intensity, mindfulness-based cognitive therapy (MBCT) for patients who have had three episodes or more of depression; and high-intensity individual cognitive behavioural therapy (CBT) for patients who have relapsed despite antidepressant medication.[9] In more severe cases, patients are usually referred for specialist treatment where relapse prevention interventions can include further high-intensity psychological treatment and lithium augmentation of antidepressant medication. There is some evidence that acute treatment with electroconvulsive therapy (ECT) and an antidepressant is more effective at preventing relapse rather than antidepressant medication alone, although the NICE Guideline Committee recognised that the evidence for this was of low quality.[8]

The availability and supply of psychological treatments as recommended by NICE is inadequate at present and it is possible that these interventions do not constitute realistic treatment options in the real-world NHS.[10] Evidence for their effectiveness and cost-effectiveness in a primary care setting is also lacking.[8] Lessons need to be learned from trials of primary care-based relapse prevention interventions and novel feasible, scalable interventions are likely to be required to ensure effective implementation and improved outcomes for patients. More research is needed to better understand relapse prevention of depression in primary care to guide optimal allocation of interventions in practice.

CAN RELAPSE BE PREDICTED?

If relapse and remission of depression could be reliably predicted at the individual patient-level, then resources can be better targeted towards relapse prevention of depression and support precision medicine, such as tailoring of intervention decisions conditional on an individual predicted risk and response to treatment.[11] This process requires prognosis research; specifically, the identification of prognostic factors and the development, validation, and impact evaluation of prognostic models for outcome risk prediction. Prognosis is 'the forecast of future outcomes for people with a particular disease or health condition.'[11]

A recent systematic review identified several prognostic factors associated with increased risk of relapse and recurrence in depression including: childhood adversity; recurrent depression; presence of residual symptoms; comorbid anxiety; rumination; neuroticism; and age of onset of depression.[12] In the UK, NICE currently highlights only a small number of these (in particular, number of previous depressive episodes and presence of residual depression symptoms) to guide prognostication in people with depression.[9]

Primary care is not yet at the point where practitioners can reliably predict outcomes for a given patient with depression based on their demographic, clinical, and disease-level characteristics. Single prognostic factors are seldom sufficient to effectively aid risk-stratification at the individual level. Rather, individualised outcome prediction is better shaped by using multiple prognostic factors in combination, in the form of a multivariable prognostic model.[11] Such risk prediction tools are increasingly recommended by policymakers and, in general practice, can be successfully built into IT systems.[11]

A robust clinical tool to risk-stratify patients and then target relapse prevention interventions to those at increased risk would be of significant benefit to patients, healthcare professionals, and the NHS as a whole.

IMPLICATIONS FOR PATIENTS AND PRACTICE

Improving risk-stratification and the allocation of relapse prevention interventions in primary care will involve discussion with patients about the risk of relapse and, for some patients, the framing of depression as a potentially chronic, on-going illness rather than something that can be 'cured'. Do patients want to have these discussions and is relapse something that concerns people with a lived experience of depression? Are such discussions required for all patients following a first episode of depression? How do clinicians decide when to adopt a chronic disease model of depression management and for which people aiming towards a more definitive treatment might be appropriate? Patient expectations and understanding may

affect outcomes and so these are important questions to consider.

The majority of existing research addressing patient preferences has been in the context of discussions around antidepressants, with fear of relapse recognised as a barrier to patients discontinuing antidepressant medication[13] and some patients confusing relapse with discontinuation symptoms.[14] Research has also shown that patients may not have full confidence in the GPs' ability to discuss discontinuation of antidepressants due to a perceived lack of knowledge and time.[15] Interestingly, GPs felt that they did have sufficient knowledge to manage continuation therapy and would be more inclined to continue antidepressant medication in patients with a history of relapse.[15] They did agree, however, that time constraints and a lack of evidence-based guidance on long-term depression management resulted in some patients being sub-optimally managed.[15]

Another consideration is whether the results of risk predictions can be used and shared in a clear and helpful manner and result in improved outcomes or lower costs when applied. To be useful in practice, prognostic models must include unambiguous prognostic factors, address a common and important problem, and have face validity (doctors must trust a model to guide their practice rather than their own experience).[11] It is possible that a statistical prediction tool aligns too closely with a biomedical model of depression that does not fully describe the course of depression in many patients. It may be that, for some patients, practitioners should be aiming to 'minimise relapse' or 'prolong remission' rather than to set the unrealistic goal of preventing relapse altogether. At the same time the authors recognise that, due to limitations imposed by the healthcare system, GPs must gather and synthesise information to aid clinical decision-making in a relatively short amount of time and a prognostic model could facilitate the identification and stratification of these different risk groups. The views and preferences of patients, healthcare professionals, and commissioners certainly need to be explored more comprehensively.

The authors hope that this editorial will encourage GPs to reflect on how relapse is currently discussed in consultations with a patient with depression. The authors highlight the need for further research into risk-stratification and more effective relapse prevention for people with depression managed in primary care.

Andrew S Moriarty,
National Institute for Health Research Doctoral Research Fellow, Department of Health Sciences; Hull York Medical School, University of York, York.

Joanne Castleton,
Patient Representative.

Simon Gilbody,
Director of the Mental Health and Addictions Research Group, Department of Health Sciences; Hull York Medical School, University of York, York.

Dean McMillan,
Professor of Clinical Psychology, Department of Health Sciences; Hull York Medical School, University of York, York.

Shehzad Ali,
Visiting Senior Research Fellow, Department of Health Sciences; Associate Professor, Department of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada.

Richard D Riley,
Professor of Biostatistics, School of Primary, Community and Social Care, Keele University, Keele, UK.

Carolyn A Chew-Graham,
Professor of General Practice Research, School of Primary, Community and Social Care, Keele University, Keele.

ADDRESS FOR CORRESPONDENCE

Andrew S Moriarty
Room 208a, Area 4, ARRC Building, University of York, Heslington, York YO10 5DD, UK.

Email: andrew.moriarty@york.ac.uk

REFERENCES

1. World Health Organization (WHO). *Depression and other common mental disorders. Global health estimates.* 2017. https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en (accessed 16 Dec 2019).

2. Rait G, Walters K, Griffin M, *et al.* Recent trends in the incidence of recorded depression in primary care. *Br J Psychiatry* 2009; **195(6):** 520–524.

3. Beshai S, Dobson KS, Bockting CLH, Quigley L. Relapse and recurrence prevention in depression: current research and future prospects. *Clin Psychol Rev* 2011; **31(8):** 1349–1360.

4. Kupfer DJ. Long-term treatment of depression. *J Clin Psychiatry* 1991; **52(Suppl):** 28–34.

5. Ali S, Rhodes L, Moreea O, *et al.* How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behav Res Ther* 2017; **94:** 1–8.

6. Kendler KS, Thornton LM, Gardner CO. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the 'kindling' hypothesis. *Am J Psychiatry* 2000; **157(8):** 1243–1251.

7. Gili M, Vicens C, Roca M, *et al* Interventions for preventing relapse or recurrence of depression in primary health care settings: a systematic review. *Prev Med* 2015; **76(Suppl):** S16–S21.

8. National Institute for Health and Care Excellence. *Depression in adults: treatment and management. Full guideline.* Methods, evidence and recommendations. 2017. https://www.nice.org.uk/guidance/GID-CGWAVE0725/documents/draft-guideline (accessed 16 Dec 2019).

9. NICE. *Depression in adults: recognition and management.* CG90. 2018. http://www.nice.org.uk/guidance/cg90 (accessed 16 Dec 2019).

10. Mental Health Taskforce. *The five year forward view for mental health.* 2016. https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf (accessed 16 Dec 2019).

11. Riley RD, van der Windt DA, Croft P, Moons KGM. *Prognosis research in healthcare: concepts, methods, and impact.* 1st edn. Oxford: Oxford University Press, 2019.

12. 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. *Clin Psychol Rev* 2018; **64:** 13–38.

13. Maund E, Dewar-Haggart R, Williams S, *et al.* Barriers and facilitators to discontinuing antidepressant use: a systematic review and thematic synthesis. *J Affect Disord* 2019; **245:** 38–62.

14. Leydon GM, Rodgers L, Kendrick T. A qualitative study of patient views on discontinuing long-term selective serotonin reuptake inhibitors. *Fam Pract* 2007; **24(6):** 570–575.

15. Bosman RC, Huijbregts KM, Verhaak PF, *et al.* Long-term antidepressant use: a qualitative study on perspectives of patients and GPs in primary care. *Br J Gen Pract* 2016; DOI: https://doi.org/10.3399/bjgp16X686641.

376

Appendix 2.1: Proof of permission to re-publish (BJGP)

CCC Marketplace™

Order Date	22-Mar-2022	Type of Use	Republish in a thesis/dissertation
Order License ID	1202211-1		
ISSN	1478-5242	Publisher	ROYAL COLLEGE OF GENERAL PRACTITIONERS
		Portion	Chapter/article

LICENSED CONTENT

Publication Title	The British journal of general practice : the journal of the Royal College of General Practitioners	Country	United Kingdom of Great Britain and Northern Ireland
Author/Editor	Royal College of General Practitioners.	Rightsholder	Royal College of General Practitioners e-Journal
Date	01/01/2002	Publication Type	
		URL	
Language	English		www.rcgp.org.uk/bjgp

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page range(s)	54-55	Distribution	U.K. and Commonwealth (excluding Canada)
Total number of pages	2		
Format (select all that apply)	Print, Electronic	Translation	Original language of publication
Who will republish the content?	Author of requested content	Copies for the disabled?	No
		Minor editing privileges?	Yes
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	GBP

NEW WORK DETAILS

Title	Predicting and prevention relapse of depression in primary care: a mixed methods study	Institution name	Hull York Medical School
		Expected presentation date	2024-01-31
Instructor name	Andrew Moriarty		

ADDITIONAL DETAILS

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	The editorial as a whole, to be incorporated in the Introduction chapter	Title of the article/chapter the portion is from	Predicting and prevention relapse of depression in primary care
Editor of portion(s)		Author of portion(s)	
Volume of serial or monograph	NA		Royal College of General Practitioners.
	NA	Issue, if republishing an article from a serial	
Page or page range of portion	54-55	Publication date of portion	N/A
			2002-01-01

CCC Terms and Conditions

1. Description of Service; De ned Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Con rmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identi ed on the Order Con rmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Con rmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Con rmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

   3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Con rmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Con rmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

   3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise speci cally set forth in the Order Con rmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms.

While User may exercise the rights licensed immediately upon issuance of the Order Con rmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Con rmation, any grant of rights to User (i) is "one-time" (including the editions and product family speci ed in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Con rmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4.
In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identi ed in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Con rmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee speci ed in the Order Con rmation, in addition to the use fee itself and any other fees and charges speci ed.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Con rmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemni es and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT,

INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, a   liates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. E ect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Con rmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

   8.1.
        User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

   8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

   8.3. The licensing transaction described in the Order Con rmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Con rmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or a liates and purporting to govern or otherwise relate to the licensing transaction described in the Order Con rmation, which terms are in any way inconsistent with any terms set forth in the Order Con rmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Con rmation, and whether such writing appears on a copy of the Order Con rmation or in a separate instrument.

8.5. The licensing transaction described in the Order Con rmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of con icts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Con rmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-7508400 or send an e-mail to support@copyright.com.

Appendix 2.2: Proof of permission to re-publish (Cochrane)

Format	Print and electronic

Portion	Full article
Will you be translating?	No

Title	NIHR Doctoral Research Fellow

Institution name	University of York and Hull York Medical School

Expected presentation date	Jan 2024

Requestor Location	Dr. Andrew Moriarty Heslington Road York, YO10 5DD United Kingdom Attn: University of York

Publisher Tax ID	EU826007151

Total	0.00 GBP

Mar 21, 2022

This Agreement between Dr. Andrew Moriarty ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5273740979473

License date	Mar 21, 2022

Licensed Content	John Wiley and Sons

Dr. Andrew Moriarty Heslington
Road

Requestor
Location

York, YO10 5DD
United Kingdom
Attn: University of York

Publisher Tax ID   EU826007151

Total   0.00 GBP

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright,

385

trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of

any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS
Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License
The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License
The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations
Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
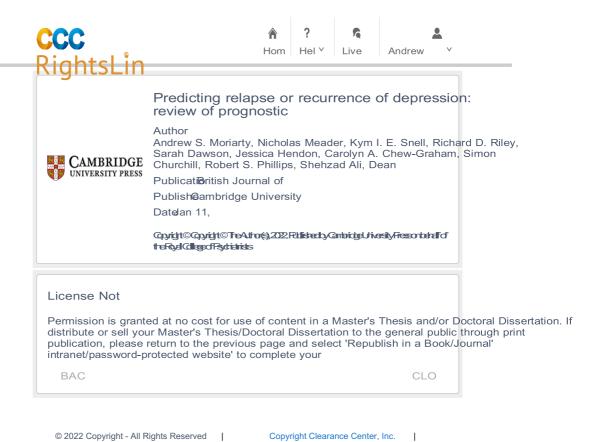
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

Appendix 2.3: Proof of permission to re-publish (BJPsych)

Appendix 2.4: Proof of permission to re-publish (BMC Diagnostic and Prognostic Research)

CCC
RightsLink®

?
Help ⌄

💬
Live Chat

The development and validation of a prognostic model to PREDICT Relapse of depression in adult patients in primary care: protocol for the PREDICTR study

SPRINGER NATURE

Author: Andrew S. Moriarty et al
Publication: Diagnostic and Prognostic Research
Publisher: Springer Nature
Date: Jul 2, 2021

Copyright © 2021, The Author(s)

Appendix 3.1: PRISMA Checklist for systematic review

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	49
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	49
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	50
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	50-54
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	54
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	395
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	55
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	55-56
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	57
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	55-56
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	58-61
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	57
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	57-59

Section and Topic	Item #	Checklist item	Location where item is reported
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	57-58
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	57-58
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	57-58
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	58
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	58
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	58
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	61-62
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	62 / Figure 3.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	62
Study characteristics	17	Cite each included study and present its characteristics.	67-71
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	88
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	NA
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	67-71
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	NA
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	NA
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	NA
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NA

Section and Topic	Item #	Checklist item	Location where item is reported
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	NA
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	90-92
	23b	Discuss any limitations of the evidence included in the review.	89-90
	23c	Discuss any limitations of the review processes used.	89-90
	23d	Discuss implications of the results for practice, policy, and future research.	92-93
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	49
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	49
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	15
Competing interests	26	Declare any competing interests of review authors.	NA
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	NA

Appendix 3.2: Database search

Database searches

• Ovid MEDLINE Search-1, (1946 to November 04, 2019), n = 2439

• Ovid MEDLINE Search-2, (1946 to March 16, 2020), n = 1518 [937 new]

• Ovid Embase (1974 to 2020 Week 19), n = 1734

• Ovid PsycINFO (1806 to May Week 1 2020), n = 1148

• Cochrane Library, (Issue 5 of 12, 2020), n = 1121

• Theses databases 8 May 2020), n = 4

Total=7964

Duplicates removed=2599

Records to screen, n = 5365 (3376 already screened (MEDLINE searches))

New records to screen (May 2020), n = 1989

Search strategies

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily <1946 to November 04, 2019>

Search Strategy: Search-1

----

1 DEPRESSION/ (112826)

2 DEPRESSIVE DISORDER/ (71437)

3 DEPRESSIVE DISORDER, MAJOR/ (28737)

4 DEPRESSION, POSTPARTUM/ (5217)

5 DEPRESSIVE DISORDER, TREATMENT-RESISTANT/ (1119)

6 (depress* adj3 (acute or clinical* or diagnos* or disorder* or major or unipolar or illness or scale* or score* or adult* or patient* or

participant* or people or inpatient* or in-patient* or outpatient* or out-patient*)).ti,ab,kf. (154965)

7 (depress* and (Beck* or BDI* or DSM* or (Statistical Manual adj2 Mental Disorders) or Hamilton or HAM-D or HAMD or MADRS or (International Classification adj2 Disease?) or ICD-10 or ICD10 or ICD-9 or ICD9 or PHQ-9 or PHQ9 or patient health questionnaire or GDS or EPDS)).ab. (48479)

8 "with depressi*".ab. (25604)

9 (depressi* or depressed).ti. (138888)

10 (depress* adj3 (postnatal* or post-natal* or postpartum* or post-partum* or pregnan*)).ti,ab,kf. (8195)

11 (depress* adj3 (refractor* or resistan* or chronic* or persist*)).ti,ab,kf. (11891)

12 (depress* and ((antidepress* or anti-depress* or SSRI* or SNRI* or serotonin or medication* or psychotropic or treatment*) adj2 (fail* or no* respon* or nonrespon* or non-respon* or unrespon* or un-respon*))).ti,ab,kf. (1539)

13 or/1-12 (298517)

14 (recurr* or relaps* or remiss* or remitt*).ti,ab,kf,hw. (900693)

15 13 and 14 (20579)

16 ((recurr* or reoccur* or re-occur* or new episode or another episode or relaps* or re-emerg* or resurg* or re-surg* or reappear* or re-appear* or flare-up) adj5 depress*).ti,ab,kf. (5822)

17 ((remiss* or remitt* or recover*) adj5 depress*).ti,ab,kf. (6368)

18 or/15-17 (24010)

19 (Prognosis/ or Decision Support Techniques/) and (Algorithms/ or Logistic Models/ or Risk Assessment/) (45685)

20 ((prognos* or predict* or decision*) and (algorithm? or model* or rule? or risk? or outcome?)).ti,kf,hw. (410679)

21 ((prognos* or predict* or decision*) adj3 (algorith? or model* or rule? or risk? or outcome?)).ab. (251570)

22 clinical prediction.mp. (2545)

23 ((prognos* or predict* or decision*) and (history or variable* or criteria or scor* or characteristic* or finding* or factor*)).ti,kf,hw. (324146)

24 ((prognos* or predict* or decision*) adj3 (history or variable* or criteria or scor* or characteristic* or finding* or factor*)).ab. (236647)

25 or/19-24 (838079)

26 18 and 25 (2455)

27 (exp animals/ or exp models, animal/) not humans.sh. (4641658)

28 (mice or mouse or murine or rat or rats or rodent* or animal model*).ti. (1421140)

29 26 not (27 or 28) (2450)

30 (comment or letter or editorial or news).sh. (1962401)

31 29 not 30 (2439)

****

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily <1946 to March 16, 2020>

Search Strategy: Search-2

----

1 *mood disorders/ or *depression/ or *depressive disorder/ or *depression, postpartum/ or *depressive disorder, major/ or *depressive disorder, treatment-resistant/ (153536)

2 (depress* or ((mood* or affective) adj disorder*)).ti. (154316)

3 limit 2 to ("in data review" or in process or publisher) (8097)

4 1 or 3 (161633)

5 exp *Recurrence/ or *Secondary Prevention/ or *Disease Progression/ (12919)

6 (predict* adj5 (longterm or long term or recurr* or reoccur* or re-occur* or new episode or another episode or relaps* or remission or re-emerg* or resurg* or re-surg* or reappear* or re-appear* or flare-up or ((future or repeat* or subsequent*) adj2 (depress* or episode?)) or ((clinical or depress* or illness) adj2 course) or ((remain* or stay*) adj (free or well or without depress*)) or (sustain* adj (recovery or remission)) or (future adj2 respon*))).ti,ab,kf. (47610)

7 5 or 6 (60278)

8 (algorithm? or decision tree? or model* or prognos* or risk? or predictors or probabilit* or ((protective or risk or sex or socioeconomic or time) adj factors)).ti,ab,kf,hw. (7782682)

9 4 and 7 and 8 (1362)

10 ((predict* adj3 (future or subsequent) adj3 (respon* or nonrespon* or treatment outcome?)) and depress*).ti,ab,kf. (67)

11 predict*.ti. and ((recurr* or relaps*) adj3 (probabilit* or likelihood? or rate? or risk?)).ti,ab,kf. and depress*.ti,kf,hw. (153)

12 predict*.ab. /freq=2 and ((recurr* or relaps*) adj3 (probabilit* or likelihood? or rate? or risk?)).ti,ab,kf. and depress*.ti,kf,hw. (227)

13 or/9-12 (1532)

14 (exp animals/ or exp models, animal/) not humans.sh. (4680637)

15 (mice or mouse or murine or rat or rats or rodent* or animal model*).ti. (1436358)

16 (comment or letter or editorial or news).sh. (2003836)

17 or/14-16 (6850921)

18 13 not 17 (1518)

****

Ovid Embase <1974 to 2020 Week 19>

Search Strategy:

----

1 *depression/ or chronic depression/ or late life depression/ or major depression/ or "mixed anxiety and depression"/ or exp perinatal depression/ or post-stroke depression/ or recurrent brief depression/ or treatment resistant depression/ (204720)

2 (depress* adj3 (acute or clinical* or diagnos* or disorder* or major or unipolar or illness or scale* or score* or adult* or patient* or participant* or people or inpatient* or in-patient* or outpatient* or out-patient*)).ti,ab,kw. (229377)

3 (depress* and (Beck* or BDI* or DSM* or (Statistical Manual adj2 Mental Disorders) or Hamilton or HAM-D or HAMD or MADRS or (International Classification adj2 Disease?) or ICD-10 or ICD10 or ICD-9 or ICD9 or PHQ-9 or PHQ9 or patient health questionnaire or GDS or EPDS)).ab. (80479)

4 "with depressi*".ab. (38096)

5 (depressi* or depressed).ti. (179033)

6 (depress* adj3 (postnatal* or post-natal* or postpartum* or post-partum* or pregnan*)).ti,ab,kw. (11865)

7 (depress* adj3 (refractor* or resistan* or chronic* or persist*)).ti,ab,kw. (17689)

8 or/1-7 (366616)

9 (recurr* or relaps* or remiss*).ti,ab,kw,hw. (1428856)

10 (longterm or long term or recurr* or reoccur* or re-occur* or new episode or another episode or re-emerg* or resurg* or re-surg* or reappear* or re-appear* or flare-up or ((future or repeat* or subsequent*) adj2 (depress* or episode?)) or ((clinical or depress* or illness) adj2 course) or ((remain* or stay*) adj (free or well or without depress*)) or (sustain* adj (recovery or remission)) or (future adj2 respon*)).ti,ab,kw. (1980780)

11 (recover* adj5 depress*).ti,ab,kw. (4130)

12 or/9-11 (2505910)

13 8 and 12 (54304)

14 prognostic index/ (47)

15 (prognosis/ or prognostic assessment/ or prediction/ or predictor variable/) and (algorithm/ or statistical model/ or risk assessment/) (84998)

16 ((prognos* or predicti* or probabilit* or decision?) and (algorithm? or model? or tool? or risk assessment?)).ti,kw. (55679)

17 prediction/ and recurrent disease/ (3181)

18 (8 and 14) or (13 and (15 or 16 or 17)) (553)

19 depressi*.ti. and (recurr* or relaps* or remiss* or recovery).ti,hw. and ((predicti* or predictor? or probability or prognostic).ti,kw,hw. or ((predicti* or predictor? or probability or prognostic) adj (index or model or tool)).ab.) and (follow up or followup or followed or months or longterm or long term).mp. (520)

20 depressi*.ti. and ((predicti* or predictor? or probability or prognostic) adj3 (recurr* or re-occur* or relaps* or remiss* or recovery)).ab. and (follow up or followup or followed or months or longterm or long term).mp. (346)

21 predict*.ti. and ((recurr* or relaps*) adj3 (probabilit* or likelihood? or rate? or risk?)).ti,ab,kw. and depress*.ti,kw,hw. (268)

22 or/18-21 (1347)

23 predict*.ab. /freq=2 and ((recurr* or relaps*) and (probabilit* or likelihood? or rate? or risk?)).ti,ab,kw. and depress*.ti,kw,hw. (1038)

24 ((prognos* or predict* or decision*) and (algorithm? or model* or rule? or tool? or risk? or outcome?)).ti,kw,hw. (848336)

25 ((prognos* or predict* or decision*) and (history or variable* or criteria or scor* or characteristic* or finding* or factor*)).ti,kw,hw. (574927)

26 ((prognos* or predict* or decision*) adj3 (algorith? or model* or rule? or tool? or risk? or outcome?)).ab. (411074)

27 ((prognos* or predict* or decision*) adj3 (history or variable* or criteria or scor* or characteristic* or finding* or factor*)).ab. (376282)

28 "decision tree"/ (12592)

29 or/23-28 (1468888)

30 13 and 29 (6371)

31 limit 30 to exclude medline journals (446)

32 22 or 31 (1734)

****

Ovid APA PsycInfo <1806 to May Week 1 2020>

Search Strategy:

----

1 (depression or depressive disorder?).hw,id. (167436)

2 (depress* adj3 (acute or clinical* or diagnos* or disorder* or major or unipolar or illness or scale* or score* or adult* or patient* or participant* or people or inpatient* or in-patient* or outpatient* or out-patient*)).ti,ab,id. (141555)

3 (depress* and (Beck* or BDI* or DSM* or (Statistical Manual adj2 Mental Disorders) or Hamilton or HAM-D or HAMD or MADRS or (International Classification adj2 Disease?) or ICD-10 or ICD10 or ICD-9 or ICD9 or PHQ-9 or PHQ9 or patient health questionnaire or GDS or EPDS)).ab. (45623)

4 "with depressi*".ab. (23223)

5 (depressi* or depressed).ti. (112230)

6 (depress* adj3 (postnatal* or post-natal* or postpartum* or post-partum* or pregnan*)).ti,ab,id. (6619)

7 (depress* adj3 (refractor* or resistan* or chronic* or persist*)).ti,ab,id. (10062)

8 or/1-7 (221741)

9 (recurr* or relaps* or remiss* or recovery).ti,ab,id,hw. (135977)

10 (longterm or long term or recurr* or reoccur* or re-occur* or new episode or another episode or re-emerg* or resurg* or re-surg* or reappear* or re-appear* or flare-up or ((future or repeat* or subsequent*) adj2 (depress* or episode?)) or ((clinical or depress* or illness) adj2 course) or ((remain* or stay*) adj (free or well or

400

without depress*)) or (sustain* adj (recovery or remission)) or (future adj2 respon*)).ti,ab,id. (181037)

11 or/9-10 (273220)

12 8 and 11 (32711)

13 (prognos* or predicti*).hw,id. and (algorithms/ or models/ or risk assessment/ or risk factors/ or at risk populations/ or *treatment outcomes/) (7659)

14 ((prognos* or predicti* or probabilit* or decision?) and (algorithm? or model? or tool? or risk)).ti,id. (24658)

15 (prediction and (recurrent depression or ((relapse or remission or recovery) adj disorders))).hw. (402)

16 ((decision trees or prognosis) and recurrence).mh. (652)

17 12 and (13 or 14 or 15 or 16) (539)

18 depressi*.ti. and (recurr* or relaps* or remiss* or recovery).ti,hw. and ((predicti* or predictor? or probability or prognostic).ti,id,hw. or ((predicti* or predictor? or probability or prognostic) adj (index or model or tool)).ab.) and (follow up or followup or followed or months or longterm or long term).mp. (168)

19 depressi*.ti. and ((predicti* or predictor? or probability or prognostic) adj3 (recurr* or re-occur* or relaps* or remiss* or recovery)).ab. and (follow up or followup or followed or months or longterm or long term).mp. (225)

20 predict*.ti. and ((recurr* or relaps*) adj3 (probabilit* or likelihood? or rate? or risk?)).ti,ab,id. and depress*.ti,id,hw. (135)

21 predict*.ab. /freq=2 and ((recurr* or relaps*) and (probabilit* or likelihood? or rate? or risk?)).ti,ab,id. and depress*.ti,id,hw. (481)

22 or/17-21 (1148)

****

Cochrane Library, Issue 5 of 12, 2020

#1 (depression or depressive):kw 39486

#2 (depress* near/3 (acute or clinical* or diagnos* or disorder* or major or unipolar or illness or scale* or score* or adult* or patient* or participant* or people or inpatient* or in-patient* or outpatient* or out-patient*)):ti,ab 37753

#3 (depress* and (Beck* or BDI* or DSM* or (Statistical Manual and Mental Disorders) or Hamilton or HAM-D or HAMD or MADRS or (International Classification and Disease*) or ICD-10 or ICD10 or ICD-9 or ICD9 or PHQ-9 or PHQ9 or patient health questionnaire or GDS or EPDS)):ab 19009

#4 (with next depressi*):ab 3851

#5 (depressi* or depressed):ti 28545

#6 (depress* near/3 (postnatal* or post-natal* or postpartum* or post-partum* or pregnan*)):ti,ab 1671

#7 (depress* near/3 (refractor* or resistan* or chronic* or persist*)):ti,ab 2896

#8 (#1 or #2 or #3 or #4 or #15 or #6 or #7) 63071

#9 (recurr* or relaps* or remiss* or recovery):ti,ab,kw 163236

#10 (longterm or "long term" or recurr* or reoccur* or re-occur* or "new episode" or "another episode" or re-emerg* or resurg* or re-surg* or reappear* or re-appear* or flare-up):ti,ab,kw 154244

#11 ((future or repeat* or subsequent*) near/2 (depress* or episode*)):ti,ab,kw 726

#12 ((clinical or depress* or illness) near/2 course):ti,ab,kw 3212

#13 ((remain* or stay*) next (free or well or without depress*)):ti,ab,kw 1272

#14 (future near/2 respon*):ti,ab,kw 159

#15 (#9 or #10 or #11 or #12 or #13 or #14) 243893

#16 #8 and #15 13670

#17 ((prognos* or predicti* or probabilit* or decision or decisions) and (algorithm* or model* or index or score or scores or tool* or risk or risks or rule or rules or tree*)):kw,ti 22750

#18 ((prognos* or predict* or decision*) and (history or variable* or criteria or characteristic* or finding* or factor* or outcome or outcomes)):ti,kw 28633

#19 ((prognos* or predicti* or probabilit* or decision or decisions) near (algorithm* or model* or index or score or scores or tool* or risk or risks or rule or rules or tree*)):ab and depress*:ti 235

#20 ((prognos* or predict* or decision*) near (history or variable* or criteria or characteristic* or finding* or factor* or outcome or outcomes)):ab and depress*:ti 864

#21 (predicti* near/3 (recurren* or relapse or remission or recovery)):ti,ab 786

#22 (#17 or #18 or #19 or #20 or #21) 40389

#23 (#16 and #22) 1056

#24 depressi*:ti and (recurr* or relaps* or remiss* or recovery):ti,kw and ((predicti* or predictor* or probability or prognostic):ti,kw or ((predicti* or predictor* or probability or prognostic) next (index or model or tool)):ab) and (follow up or followup or followed or months or longterm or long term):ti,ab,kw 93

#25 depressi*:ti and ((predicti* or predictor* or probability or prognostic) near (recurr* or re-occur* or relaps* or remiss* or recovery)):ab and ("follow up" or followup or followed or months or longterm or "long term"):ti,ab,kw 117

#26 (#23 or #24 or #25) 1121

****

Appendix 3.3: Detailed Characteristics of Included Studies

Table A3.1: Detailed characteristics of studies included in systematic review

Backs-Dermott 2010

Study details	**Sponsorship source:** Canadian Institutes of Health Research **Country:** Canada **Setting:** Community setting **Year of recruitment:** Not reported
Methods	**Type of study:** Model development study **Source of data:** Prospective longitudinal cohort study **Method used for model development:** Differential Function Analysis **Method used for internal validation:** Not reported **External validation:** Not done **Handling of missing data:** Not reported **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** 49 (29) **Number of candidate predictor parameters:** 11 **Number of predictors in final model:** 5 **Number of events per candidate predictor parameter (EPP):** Not applicable
Population	**Inclusion criteria:** Female Aged 18 - 65 Diagnosis of DSM-IV-TR current Major Depressive Episode (MDE) or MDE within the past 8 week **Exclusion criteria:** Ever experienced a manic or mixed episode Meeting criteria for a psychotic disorder, or ever experienced 2 or more psychotic symptoms

	Meeting criteria for depression with psychotic features Meeting criteria for substance abuse disorder or dependence
Baseline characteristics	**Mean age (SD):** Relapse group: 43.1 (10.87); Stable remitted group: 43.65 (11.72) **Gender (% Female):** 100
Start-point (diagnosis of depression and remission)	**Depression:** Diagnosis of DSM-IV-TR current Major Depressive Episode (MDE) or MDE within the past 8 weeks **Remission:** "per Frank 1991 criteria": 1) reported less than 2 symptoms of depression on the SCID-I for at least 2 weeks; and 2) scored ≤ 13 on the BDI-II
End-point (diagnosis of relapse/recurrence)	**Relapse** within 12 months: meeting current criteria for MDE according to SCID-I
Timing (length of follow-up)	12 months
Notes	

Berlanga 1999

Study details	**Sponsorship source:** Not reported **Country:** Mexico **Setting:** Secondary care (outpatients) **Year of recruitment:** 1994 - 1996
Methods	**Type of study:** Model development study **Source of data:** Post-RCT* prospective follow-up study **Method used for model development:** Logistic regression (multivariable analysis with a stepwise backward method in which variables that were significant in the univariable analysis were introduced into the model) **Method used for internal validation:** Not reported **External validation:** Not done **Handling of missing data:** Not reported **Evaluation of clinical utility:** Not assessed

Sample size	**Total number of participants (Number with event):** 42 (18) **Number of candidate predictor parameters:** Not reported **Number of predictors in final model:** 3 **Number of events per candidate predictor parameter (EPP):** Unclear

Population	**Inclusion criteria:**
	
	  Between 18 and 65 years old
	  DSM-IV criteria for diagnosis of major depressive disorder
	  Scoring at least 18 points on the first 17 items of the 21-item version of the Hamilton Rating Scale for Depression (HAM-D)
	
	Exclusion criteria:
	
	  Psychotic symptoms
	  Substantial suicide risk
	  If any other situation required hospitalisation

Baseline characteristics	**Mean age (SD):** Recurrence group: 34.8 (11.1); No-recurrence group: 37.2 (11.2) **Gender (% Female):** Recurrence group: 83; No-recurrence group: 71
Start-point (diagnosis of depression and remission)	**Depression:** Major depressive disorder according to DSM-IV criteria and at least 18 points on the first 17 items of the 21-item HAM-D **Remission:** Definition of remission not reported
End-point (diagnosis of relapse/recurrence)	**Recurrence:** Fulfilling criteria for MDD (clinical interview) per Frank 1991
Timing (length of follow-up)	12 months
Notes	*The RCT compared the clinical efficacy and tolerance of the antidepressants nefazodone and fluoxetine. A 'washout period' of at least 3 weeks free of antidepressant medication was a requisite for all participants

Johansson 2015

Study details	**Sponsorship source:** Not reported **Country:** Sweden **Setting:** Secondary care (psychiatric outpatients) **Year of recruitment:** Not reported
Methods	**Type of study:** Model development study **Source of data:** Prospective cohort study **Method used for model development:** Logistic regression (the 2 predictor variables were chosen which showed the strongest independent correlations with relapse/recurrence) **Method used for internal validation:** Not reported **External validation:** Not done **Handling of missing data:** Not reported **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** 51 (31) **Number of candidate predictor parameters:** 4 (based on univariable analysis) **Number of predictors in final model:** 2 **Number of events per candidate predictor parameter (EPP):** 7.75
Population	**Inclusion criteria:** Outpatients with a primary diagnosis of depressive episode or recurrent depressive disorder (ICD-10 criteria) At least 18 years of age In remission **Exclusion criteria:** Psychotic features Diagnosis of bipolar disorder Received ECT for the index period
Baseline characteristics	**Mean age (SD):** 47 (SD = 17)

	Gender (% Female): 71
Start-point (diagnosis of depression and remission)	**Depression:** ICD-10 criteria for depressive episode or recurrent depressive disorder **Remission:** determined by psychiatrist at discharge and confirmed by structured clinical interview Partial remission defined as not fulfilling the criteria of DSM-IV depressive episode but having more than minimal symptoms (i.e. Montgomery–Asberg depression rating scale—self rating scale (MADRS-S) score > 9) Full remission is defined as not fulfilling the criteria of DSM-IV depressive episode and showing only minimal symptoms (i.e. MADRS-S < 10)
End-point (diagnosis of relapse/recurrence)	**Relapse/recurrence:** (per Frank 1991) Relapse defined as having a depressive episode within 2 months of discharge Recurrence defined as having a depressive episode after a period of recovery (at least 2 months after discharge) Relapse/recurrence and current depressive status established using the sections Mood Episodes and Mood Disorders from The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)
Timing (length of follow-up)	12-14 months
Notes	

Judd 2016

Study details	**Sponsorship Source:** Not reported **Country:** US **Setting:** Secondary care (academic centres) **Year of Recruitment:** 1978-1981
Methods	**Type of study:** Model development study **Source of data:** Prospective cohort study (the National Institute of Mental Health Collaborative Depression Study)

	Method used for model development: Forward and backward selection of pre-selected predictors using stepwise mixed-model logistic regression **Method used for internal validation:** Not reported **External validation:** Not done **Handling of missing data:** Multiple imputation **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** 188 (58)* 514 SCL-90 assessments (73 with relapse) **Number of candidate predictor parameters:** 17 **Number of predictors in final model:** 12 **Number of events per candidate predictor parameter (EPP):** 4.29 (17 candidate predictors to 73 "relapses")
Population	**Inclusion criteria:** White IQ > 70 Speak English Entered the National Institute of Mental Health Collaborative Depression Study in an active major depressive episode **Exclusion criteria:** Lifetime bipolar disorder or schizophrenia
Baseline characteristics	**Mean age (SD):** 37.8 (14.4) **Gender (% female):** 58.5
Start-point (diagnosis of depression and remission)	**Depression:** Major depression, assessed by Research Diagnostic Criteria based on Schedule for Affective Disorders and Schizophrenia interviews (no lifetime bipolar disorder, schizoaffective disorder or schizophrenia)

	Remission: Psychiatric status rating of 1 (asymptomatic, returned to usual self with no symptoms of the episode) for at least 8 weeks
End-point (diagnosis of relapse/recurrence)	**Relapse (within 6 months):** 2 consecutive weeks of psychiatric status ratings at threshold for defining episode of major or minor/dysthymic depression
Timing (length of follow-up)	6 months
Notes	*There were 514 SCL-90 assessments taken from 188 participants. 73 of these assessments (from 58 participants) were identified as having relapsed

Klein 2018

Study details	**Sponsorship source:** Not reported **Country:** The Netherlands **Setting:** Primary care **Year of recruitment:** Development data: 2010 - 2013; Validation data: 2009 - April 2015
Methods	**Type of study:** Model development study with external validation **Source of data:** Prospective data from 2 pragmatic RCTs **Method used for model development:** Cox proportional hazards regression (backward selection at P < 0.05) **Method used for internal validation:** Bootstrapping; shrinkage determined for all statistics **External validation:** Separate RCTs formed development and validation datasets **Handling of missing data:** Multiple imputation **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** Development dataset: 235 (104); Validation dataset: 205 (116) **Number of candidate predictor parameters:** 8 **Number of predictors in final model:** 4 **Number of events per candidate predictor parameter (EPP):** 13

Population	**Inclusion criteria:**
	
	  Aged 18 to 65 years
	  Experienced at least 2 episodes of major depressive disorder (the last one within 2 years)

	Remitted according to DSM-IV criteria and HRSD < 10 **Exclusion criteria:** Mania/hypomania Psychotic or bipolar disorder (past or present) Alcohol/drug abuse Primary diagnosis of an anxiety disorder Organic brain damage
Baseline characteristics	**Mean age (SD):** Development dataset: 46.8 (10.6); Validation dataset: 48.3 (9.9) **Gender (% female):** Development dataset: 74.5; Validation dataset: 66.5
Start-point (diagnosis of depression and remission)	**Depression:** DSM-IV criteria **Remission:** Assessed using SCID-I and HRSD score ≤ 10
End-point (diagnosis of relapse/recurrence)	**Recurrence (time to) within 2 years:** assessed using SCID-I
Timing (length of follow-up)	2 years
Notes	

Mocking 2021

Study details	**Sponsorship source:** Not reported **Country:** US **Setting:** Community setting **Year of recruitment:** 2011-2014
Methods	**Type of study:** Model development study **Source of data:** Cross-sectional study comparing people with remitted recurrent MDD (rrMDD) with never depressed controls (rrMDD population followed up for 2.5 years) **Method used for model development:** Cox proportional hazards regression

	Method used for internal validation: Repeated double cross validation (rdCV), with bootstrapping 100 times to test random subsamples of 2/3 in and 1/3 out, and by permutation analysis **External validation:** Not done **Handling of missing data:** Not reported **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** 62 (35) **Number of candidate predictor parameters:** 399 intracellular and plasma metabolites (number of parameters unclear) **Number of predictors in final model:** Unclear **Number of events per candidate predictor parameter (EPP):** Unclear
Population	**Inclusion criteria:** ≥2 episodes of MDD according to DSM-IV Stable remission – not meeting SCID criteria for MDD and HAM-D<8 Aged 35-65 **Exclusion criteria:** Current diagnoses of alcohol and/or drug dependence, psychotic or bipolar symptoms, predominant anxiety or severe personality disorder. Also standard MRI-exclusion criteria, history of severe head trauma or neurological disease, or severe general physical illness. All participants had to be without psychoactive medication for ≥4weeks.
Baseline characteristics	**Mean age (SEM):** Males: 54 (1.4); Females: 53 (1.2) **Gender (% female):** 66.1
Start-point (diagnosis of depression and remission)	**Depression:** DSM-IV criteria **Remission:** Assessed using SCID-I and HRSD score ≤ 10
End-point (diagnosis of relapse/recurrence)	**Recurrence:** ≥5 depressive symptoms lasting at least 2 weeks according to the DSM-IV criteria (SCID).
Timing (length of follow-up)	2.5 years
Notes	

Pintor 2009

Study details	Sponsorship source: Not reported Country: Spain Setting: Secondary care (outpatients) Year of recruitment: 2001 - 2005
Methods	Type of study: Model development Source of data: Prospective cohort study Method used for model development: Logistic regression Method used for internal validation: Not reported External validation: Not done Handling of missing data: Not reported Evaluation of clinical utility: Not assessed
Sample size	Total number of participants (Number with event): 43 (18) Number of candidate predictors: Not reported Number of predictors in final model: 3 Number of events per candidate predictor parameter (EPP): Unclear

Population	Inclusion criteria:
	
	  Experienced a depressive episode according to DSM-IV (SCID)
	  Aged 30 - 65
	
	Exclusion criteria:
	
	  Alcohol or drug dependence
	  Current or history of severe psychiatric disorders except MDD
	  Severe physical health disorders
	  Body weight > 150% of ideal weight

	Taking antiepileptics Needle phobia Pregnant
Baseline characteristics	**Mean age (SD):** Relapsed group: 50.67 (8.04); Non-relapsed group: 51.88 (8.54) **Gender (% female):** Relapsed group: 50; Non-relapsed group: 56
Start-point (diagnosis of depression and remission)	**Depression:** SCID-IV diagnosis for unipolar major depressive episode (first or recurrent) **Remission:** identified using Hamilton Depression Rating Scale (HDRS-21); "Frank 1991 criteria were applied" (does not describe exactly how)
End-point (diagnosis of relapse/recurrence)	**Presence versus absence of relapse over 2-year follow-up:** identified using Hamilton Depression Rating Scale (HDRS-21); "Frank 1991 criteria were applied" (does not describe exactly how)
Timing (length of follow-up)	2 years
Notes	

Ruhe 2019

Study details	**Sponsorship source:** Not reported **Country:** The Netherlands **Setting:** Primary care **Year of recruitment:** Not reported
Methods	**Type of study:** Model development study **Source of data:** Prospective cohort study **Method used for model development:** Machine learning support vector machine (SVM); data-driven model (classification-based algorithm) **Method used for internal validation:** "Leave-one-out" validation procedure **External validation:** Not done **Handling of missing data:** Mean imputation **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** 64 (35) **Number of candidate predictors:** Not reported

	Number of predictors in final model: 4 Number of events per candidate predictor parameter (EPP): Unclear

Population	**Inclusion criteria:**
	
	  Voluntarily free of anti-depressants for past weeks
	  Between 35 and 65 years old
	  2 or more previous episodes of MDD
	
	Exclusion criteria:
	
	  Alcohol or drug dependence
	  Primary anxiety disorder
	  Psychotic or bipolar disorder
	  Received ECT within 2 months of assessment
	  History of head trauma, neurological disease or severe physical illness

Baseline characteristics	**Mean age (SD):** 53.4 (7.7) **Gender (% female):** 65.8
Start-point (diagnosis of depression and remission)	**Depression:** Recurrent MDD: 2 or more MDD episodes according to the SCID-I **Remission:** ≤ 7 on the HDRS) for ≥ 8 weeks and not fulfilling the criteria for a current MDD episode (SCID-I)
End-point (diagnosis of relapse/recurrence)	**Recurrence:** MDD according to SCID-I.
Timing (length of follow-up)	**Median follow up:** 233 days (IQR 92 - 461)
Notes	

Van Loo 2015

Study details	**Sponsorship source:** Not reported **Country:** USA

	Setting: Community setting **Year of recruitment:** 1988 - 1997
Methods	**Type of study:** Model development study with external validation **Source of data:** Prospective longitudinal data* **Method used for model development:** Elastic net penalised Cox proportional hazards regression **Method used for internal validation:** 10-fold cross-validation and shrinkage of beta-coefficients **External validation:** Temporal validation **Handling of missing data:** Single imputation **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (Number with event):** Development dataset: 194 (45); Validation dataset: 133 (57) **Number of candidate predictor parameters:** 81 candidate predictors (number of parameters unclear) **Number of predictors in final model:** 26 **Number of events per candidate predictor parameter (EPP):** Unclear
Population	**Inclusion criteria:** Female twins DSMIII MD episode in the previous year **Exclusion criteria:** Not listed.
Baseline characteristics	**Mean age (SD):** Development dataset: 30.7 (7.1); Validation dataset: 32.4 (7.1) **Gender (% female):** 100
Start-point (diagnosis of depression and remission)	**Depression:** DSM-III MD episode in previous year (self-report and confirmed by SCID) **Remission:** No longer meeting criteria according to SCID
End-point (diagnosis of relapse/recurrence)	**Recurrence:** first episode meeting DSM-III-R criteria after a period of not meeting the criteria (remission or recovery) for at least 4 months

	Time to recurrence: Number of months between initial interview and recurrence
Timing (length of follow-up)	Development dataset: median follow-up 5.5 years; Validation dataset: median follow-up 6.1 years
Notes	*Data from prospective longitudinal studies of Caucasian female-female twin pairs (Virginia Adult Twin Study of Psychiatric and Substance Use Disorder)

Van Loo 2018

Study details	**Sponsorship source:** Not reported **Country:** USA **Setting:** Community setting **Year of recruitment:** 1988 - 1997
Methods	**Type of study:** Model development study **Source of data:** Longitudinal cohort study* **Method used for model development:** Cox proportional hazards model with elastic net penalised regression analysis **Method used for internal validation:** Random split "test" sample. The final model was selected based on minimal prediction error as assessed in 10-fold cross-validation **External validation:** Not done **Handling of missing data:** Multiple imputation by chained equations **Evaluation of clinical utility:** Not reported
Sample size	**Total number of participants (Number with event): Total sample (men and women):** 653** **Number of candidate predictor parameters:** 70 predictors (number of parameters unclear) **Number of predictors in final model:** 24 **Number of events per candidate predictor parameter (EPP):** Unclear
Population	**Inclusion criteria:** Episode of MD in year prior to baseline interview **Exclusion criteria:** No MD episode in year prior to baseline interview

	Those who reported an interval of 60 days or less between the offset of their last MD episode at baseline and their first depressive episode during the follow-up
Baseline characteristics	**Mean age (SD):** 35 (8.8) **Gender (% female):** 34.6
Start-point (diagnosis of depression and remission)	**Depression:** A diagnosis of MD in the year prior to baseline interview was based on the DSM-III-R criteria as assessed by the Structured Clinical Interview for DSM-III-R **Remission:** All participants reported a period of > 60 days of (partial) remission or recovery
End-point (diagnosis of relapse/recurrence)	**Recurrence:** First reported episode meeting DSM-III-R criteria in the year prior to follow-up interview **Time to recurrence:** Time at risk for recurrence (follow-up) was defined as the interval between the offset of MD in the year prior to baseline interview and the onset of MD in the year prior to follow-up interview
Timing (length of follow-up)	5 years
Notes	*Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), a population-based longitudinal study of male–male and male–female Caucasian twin pairs **This was the full sample size, including men and women. There were also separate analyses in women (n = 226) and in men (n = 427). The male cohort was split further into a training sample (n = 277) and a test sample (for external validation)

Van Loo 2020

Study details	**Sponsorship source:** Funding for NESDA reported in paper **Country:** Netherlands (NESDA); USA (VATSPSUD) **Setting:** Primary care, secondary care and community setting (NESDA); Community setting (VATSPSUD) **Year of recruitment:** 2004 - 2007 (NESDA); 1988-1997 (VATSPSUD)
Methods	**Type of study:** External validation study using NESDA data (internal validation also performed on VATSPSUD data) **Source of data:** 2 longitudinal cohort studies* **Method used for model development:** Not applicable **Method used for internal validation:** Random split sample of VATSPSUD data used in Van Loo 2018* **External validation:** Logistic regression using NESDA dataset** **Handling of missing data:** Multiple imputation by chained equations

	Evaluation of clinical utility: Not done
Sample size	Total number of participants (Number with event): NESDA Test sample (n = 1925); VATSPSUD Test sample (n = 2301). Number with event not clear Number of candidate predictor parameters: Not applicable Number of predictors in final model: 24 Number of events per candidate predictor parameter (EPP): Not applicable

Population	For external validation (NESDA):
	Inclusion criteria:
	
	  Dutch general population, primary care, and specialised mental health care, aged 18 – 65 at baseline assessment
	
	Exclusion criteria:
	
	  No MD episode in year prior to baseline interview.
	  Those who reported an interval ⩽ 60 days between the offset of their last MD episode at baseline and their first depressive episode during the follow-up
	
	For internal validation (VATSPSUD):
	Female-female twins (n = 757) and male-male/male-female twins (n = 1544) from the VATSPSUD study (only those not included in the original training sample used to develop the prediction model in Van Loo 2018)

Baseline characteristics	Mean age (SD): NESDA Test sample: 42 (12.4); VATSPSUD Test sample: 34.9 (8.6) Gender (% female): NESDA Test sample: 68.6; VATSPSUD Test sample: 53.2
Start-point (diagnosis of depression and remission)	Depression: Lifetime episode of MD at baseline Remission: Not described
End-point (diagnosis of relapse/recurrence)	Recurrence: Any episode of MD during follow-up Time to recurrence: Follow-up to 9 years
Timing (length of follow-up)	

Notes	*Two independent test samples from Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD) and the Netherlands Study of Depression and Anxiety (NESDA) **External validation performed on NESDA cohort

Wang 2014

Study details	**Sponsorship source:** Not reported **Country:** USA **Setting:** Community setting **Year of recruitment:** 2001 - 2005
Methods	**Type of study:** Model development study with external validation **Source of data:** Prospective longitudinal dataset* **Method used for model development:** Logistic regression with combined forward and backward selection (compared C-statistic with and without each predictor, then used Net Reclassification Improvement to examine if the predictor could correctly reclassify participants into appropriate categories) **Method used for internal validation:** Application of heuristic shrinkage factor **External validation:** Geographical validation **Handling of missing data:** Single imputation **Evaluation of clinical utility:** Not assessed
Sample size	**Total number of participants (number with event):** Development dataset: 1518 (362); Validation dataset: 1195 (307) **Number of candidate predictor parameters:** Not reported **Number of predictors in final model:** 24 **Number of events per candidate predictor parameter (EPP):** Unclear
Population	**Inclusion criteria:** Current or lifetime MDE Remitted from MDE for at least 2 months Went to health professionals (councillors and/or medical doctors) for help to improve mood, were hospitalised for depression, or went to emergency room because of depression **Exclusion criteria:**

	Lifetime manic or hypomanic episodes
Baseline characteristics	**Mean age (SEM):** Development dataset: 45.38 (0.37); Validation dataset: 45.37 (0.41) **Gender (% Female):** Development dataset: 77.4%; Validation dataset: 74.9%
Start-point (diagnosis of depression and remission)	**Depression:** DSM-IV **Remission:** "Having remitted from recent depressive episode for at least 2 months"
End-point (diagnosis of relapse/recurrence)	**Recurrence, within 3 years:** Meeting DSM-IV diagnostic criteria for MDE
Timing (length of follow-up)	3 years
Notes	*Data from the US National Epidemiological Survey on Alcohol and Related Conditions

Appendix 3.4: Risk of bias and applicability

Table A3.2: Detailed risk of bias and applicability (PROBAST) assessment

	Study														
	Backs-Dermott 2010	Berlanga 1999	Johansson 2015	Judd 2016	Klein 2018		Mocking 2021	Pintor 2009	Ruhe 2019	Van Loo 2015		Van Loo 2018	Van Loo 2020	Wang 2014	
Type of study	Dev	Dev	Dev	Dev	Dev	Val	Dev	Dev	Dev	Dev	Val	Dev	Val	Dev	Val
	Domain 1: Participants														
	A. Risk of bias														
1.1. Appropriate data sources?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1.2. Appropriate inclusions and exclusion?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk of bias	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
	B. Applicability														

Concern about applicability	Low	Unclear	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Domain 2: Predictors															
A. Risk of bias															
2.1. Defined and assessed in similar way for all participants?	Yes	Yes	Yes	Yes	Yes	Yes	Probably yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.2. Assessments made without knowledge of outcome?	Probably yes	Probably yes	Yes	Yes	Yes	Yes	No information	No information	Yes	Probably yes	Probably yes	No information	No information	Probably yes	Probably yes
2.3. All available at time of model's intended use?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Risk of bias	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Unclear**	**Unclear**	**Low**	**Low**	**Low**	**Unclear**	**High**	**Low**	**Low**

423

	B. Applicability														
Concern about applicability	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
	Domain 3: Outcome														
	A. Risk of bias														
3.1. Determined appropriately?	Yes	Yes	Yes	Probably yes	Yes	Yes	Probably yes	No information	Yes	Yes	Yes	Yes	Yes	Probably yes	Probably yes
3.2. Pre-specified or standard definition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.3. Predictors excluded from outcome definition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.4. Defined and determined similar for all	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

participants?															
3.5. Determined without knowledge of predictors?	No information	No information	No information	Probably yes	Yes	Yes	Probably yes	No information	No information	No information	No information	No information	No information	No information	No information
3.6. Appropriate time interval between predictor assessment and outcome determination?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk of bias	**Unclear**	**Unclear**	**Unclear**	**Low**	**Low**	**Low**	**Low**	**Unclear**	**Unclear**	**Unclear**	**Unclear**	**Unclear**	**Unclear**	**Unclear**	**Unclear**
B. Applicability															
Concern about applicability	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
Domain 4: Analysis															

4.1. Reasonable number of participants with outcome?	Probably yes	No	No	No	Probably yes	Yes	No	No	No	No	No	No	Probably yes	No information	Yes
4.2. Predictors handled appropriately?	Yes	Probably yes	Yes	No	Probably yes	Probably yes	No information	No	Probably no	No	No	Probably yes	Probably yes	No	No
4.3. All enrolled participants included in analysis?	No	No	Yes	No	Yes	Yes	No	Yes	No	Probably yes	Probably yes	No	Yes	Yes	Yes
4.4. Missing data handled appropriately?	No information	No information	No information	Yes	Yes	Yes	No information	No information	No	No	No	Yes	Yes	Probably no	Probably no
4.5. Univariable analysis avoided?	No	No	No	No	Yes	NA	Probably yes	No	Yes	Yes	NA	Yes	NA	No	NA

4.6. Complexities in data accounted for?	Probably yes	Probably yes	Probably yes	Yes	Yes	Yes	No information	Probably yes	Yes	Probably yes	Probably yes	Yes	Probably yes	Probably yes	Probably yes
4.7. Relevant performance measures?	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No
4.8. Overfitting and optimism accounted for?	No	No	No	No	Yes	NA	No information	No	No	Yes	NA	Yes	NA	Yes	NA
4.9. Final model corresponds to multivariable analysis?	No information	No information	Probably yes	No information	Yes	NA	No information	No information	No information	Probably no	NA	Probably yes	NA	No information	NA
Risk of bias	**High**	**High**	**High**	**High**	**Low**	**Low**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**
Overall assessment of	**High**	**High**	**High**	**High**	**Low**	**Low**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**

risk of bias															
Overall concern for applicability	Low	Unclear	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low

Appendix 5.1: Pre-registered, published protocol for quantitative study

PROTOCOL   **Open Access**

The development and validation of a prognostic model to PREDICT Relapse of depression in adult patients in primary care: protocol for the PREDICTR study

Andrew S. Moriarty[1,2][*], Lewis W. Paton[1], Kym I. E. Snell[3], Richard D. Riley[3], Joshua E. J. Buckman[4,5], Simon Gilbody[1,2], Carolyn A. Chew-Graham[6], Shehzad Ali[1,7], Stephen Pilling[4,8], Nick Meader[9], Bob Phillips[9], Peter A. Coventry[1], Jaime Delgadillo[10], David A. Richards[11,12], Chris Salisbury[13] and Dean McMillan[1,2]

Abstract

Background: Most patients who present with depression are treated in primary care by general practitioners (GPs). Relapse of depression is common (at least 50% of patients treated for depression will relapse after a single episode) and leads to considerable morbidity and decreased quality of life for patients. The majority of patients will relapse within 6 months, and those with a history of relapse are more likely to relapse in the future than those with no such history. GPs see a largely undifferentiated case-mix of patients, and once patients with depression reach remission, there is limited guidance to help GPs stratify patients according to risk of relapse. We aim to develop a prognostic model to predict an individual's risk of relapse within 6–8 months of entering remission. The long-term objective is to inform the clinical management of depression after the acute phase.

Methods: We will develop a prognostic model using secondary analysis of individual participant data drawn from seven RCTs and one longitudinal cohort study in primary or community care settings. We will use logistic regression to predict the outcome of relapse of depression within 6–8 months. We plan to include the following established relapse predictors in the model: residual depressive symptoms, number of previous depressive episodes, co-morbid anxiety and severity of index episode. We will use a "full model" development approach, including all available predictors. Performance statistics (optimism-adjusted C-statistic, calibration-in-the-large, calibration slope) and calibration plots (with smoothed calibration curves) will be calculated. Generalisability of predictive performance will be assessed through internal-external cross-validation. Clinical utility will be explored through net benefit analysis.

Discussion: We will derive a statistical model to predict relapse of depression in remitted depressed patients in primary care. Assuming the model has sufficient predictive performance, we outline the next steps including independent external validation and further assessment of clinical utility and impact.

Study registration: ClinicalTrials.gov ID: NCT04666662

Keywords: Relapse, Recurrence, Depression, Prognosis, Prognostic model, Predictive model

* Correspondence: andrew.moriarty@york.ac.uk
[1]Department of Health Sciences, University of York, York, England
[2]Hull York Medical School, University of York, York, England
Full list of author information is available at the end of the article

Introduction

Depression is the leading cause of disability worldwide [1], and most people with depression are treated in primary care [2]. Around half of patients will experience a re-emergence of depressive symptoms at some point after their initial symptoms have improved, and for the majority of these, this occurs within the first 6 months [3]. Relapse and recurrence are both terms used to describe the re-emergence of depressive symptoms following some level of improvement. Generally, relapse occurs after some improvement (remission) but before full recovery [4], whereas recurrence is the onset of a further, separate episode after full recovery. While there is no empirically derived temporal cut-off to distinguish relapse from recurrence, recovery is most commonly operationalized as following an extended period of remission; between 6 and 12 months [5]. Relapse, then, occurs within 6–12 months, while recurrence occurs beyond 6–12 months [4, 6].

The distinction between relapse and recurrence provides a useful theoretical framework and there may be some clinical relevance. The implication is that the re-emergence of symptoms in relapse is part of the unsuccessfully treated index episode of depression, while in recurrence, it is attributable to a new and separate episode of depression. When the MacArthur Foundation Research Network defined these terms, or "change points," in 1991 [4, 6], their aim was to provide a framework that might be more consistently applied in the empirical literature, but also that the framework and definitions themselves be validated empirically by researchers. There have been limited attempts to do this, though where this has been attempted researchers have found some evidence to support their validity [7]. Given the wide variability in the way in which the terms relapse and recurrence have been operationalized by researchers, however, Bockting et al. [5] suggested using the terms interchangeably to describe the "re-emergence of symptoms following a period of relative wellness". We will use the term *relapse* throughout this paper.

There is some evidence that the severity of depression [8] and risk of further relapse [9–11] increases with each subsequent depressive episode, highlighting the potential benefits of intervening early to prevent relapse and recurrence with a view to improving the overall trajectory of depression. Efforts to prevent relapse could be improved by an increased ability to predict prognosis and identify high-risk individuals. Prognosis can be shaped by multiple factors, and once it is established which factors are associated with an outcome, the information can be used to create a multivariable prognostic model. Prognostic models aim to provide individualised risk estimates for a specified outcome by a particular time conditional on the individual's values for multiple prognostic factors (or predictors) [12]. We currently lack evidence-based tools to assist clinicians with risk predictions of depressive relapse.

There have been some previous attempts to develop relapse prediction models for depression [13–17]. These pre-existing prognostic models have some drawbacks with respect to successfully predicting relapse in a primary care context. Critical appraisal of these studies, using the Prediction model Risk Of Bias Assessment Tool (PROBAST), found that the majority of these studies were at high overall risk of bias [18]. The most significant limitations were inadequate sample size, inappropriate handling of missing data and presentation of inappropriate performance statistics (calibration and discrimination not assessed) [18]. Furthermore, the developed models have either demonstrated insufficient predictive performance on external validation [13], or they could not be feasibly implemented in a primary care setting due to the large number and type of included predictors [16].

This protocol outlines the methods for the development and validation of a novel prognostic model to predict an individual's risk of relapse of depression in a primary care setting. The long-term aim, beyond this study, is to implement the prognostic model in clinical practice for use by primary care health professionals to enable optimal shared decision making with patients. The model must, therefore, be accurate, generalisable and effective (i.e. result in demonstrably improved outcomes for patients). In order to be implemented in practice, it must also be clinically credible and have face validity to healthcare professionals and patients.

Objective

The objective is to develop and validate a multivariable prognostic model to predict relapse within 6 to 8 months in patients with remitted depression in primary care.

Methods

The methods have been developed in accordance with those recommended by the PROGnosis RESearch Strategy (PROGRESS) initiative [19, 20], and the prognostic model will be published according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) Statement [21]. This study will use individual participant data (IPD) from RCTs and a cohort study; therefore, elements of the Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA-IPD) statement are also relevant [22]. However, this study is not a systematic review and the aim is not to provide a summary of a complete body of research and so not all items are applicable. A Patient and Public

Involvement (PPI) group of service users have informed several aspects of this study, including selecting predictors and their measurement (for example, commenting on the acceptability of validated diagnostic instruments for depression and anxiety symptoms), definition of outcome, target patient population and clinical application. The study has been registered prospectively on ClinicalTrials.gov (available: ClinicalTrials.gov ID: NCT04666662).

Source of data

We have formed a cohort using IPD from UK primary care-based datasets. Along with cohort studies, RCTs are a recommended source of data for development of prognostic models [23]. We had IPD readily available in a pragmatic sample of four RCTs (CASPER Plus [24], REEACT [25], REEACT 2 [26] and COINCIDE [27]), derived from RCTs carried out within our own research group. In order to increase the sample size available for model development, we identified further studies: first, by searching all the National Institute for Health Research (NIHR)-funded RCTs of primary care-based interventions for depression, and second, by reference to the search results from a recent IPD meta-analysis of RCTs of depression interventions (this meta-analysis had searched for studies that had used the CIS-R as a measure of baseline severity and provided a recent search of relevant studies) [28].

To be included, we specified that RCTs must:

- Include adult patients (18 years and over) with depression and measure depressive symptoms at a minimum of three time-points (to enable diagnosis of depression, remission, relapse/no relapse). We excluded RCTs in patient groups with significant psychiatric or medical comorbidity. We also excluded feasibility studies (due to limited sample size and shorter follow-up time associated with those identified);
- Have sufficient follow-up to allow us to detect relapse within at least 6 months;
- Use only non-pharmacological interventions (e.g. psychological, social, behavioural). We excluded RCTs of pharmacological interventions, as these were felt likely not to be comparable to the pharmacological interventions that patients would be receiving from their primary care healthcare providers as usual treatment. Trials of pharmacological interventions often use medication combinations that would not be routinely prescribed in primary care and would therefore potentially reduce the generalisability of the model. Non-pharmacological interventions are more likely to affect outcomes in a comparable way;

- Use the Patient Health Questionnaire (PHQ-9) as a measure of depression.

This search added three further RCTs: COBRA [29], CADET [30] and Healthlines Depression RCT [31].

Finally, we contacted the authors of the West Yorkshire Low Intensity Outcome Watch (WYLOW) study, a longitudinal cohort study following-up patients after low-intensity cognitive behavioural therapy (LiCBT) through the Improving Access to Psychological Therapies (IAPT) service [3]. See Table 1 for details of the final included studies.

All of the included studies had pragmatic and unrestrictive inclusion criteria, and so are expected to be representative of the target population. The final PREDICTR dataset is derived from all arms (control and intervention) of seven randomised controlled trials (RCTs) of low-intensity primary care-based interventions for depression (CASPER Plus, REEACT, REEACT 2, COINCIDE, CADET, COBRA, Healthlines) and one observational cohort study (WYLOW).

Participants

Adult participants (aged 18 years and over) with depression. The included participants do not have significant psychiatric comorbidity (e.g. schizophrenia, bipolar affective disorder).

Setting

All data sources are primary care or community-based.

Start-point (remission)

There are three important time-points: baseline for the RCT (i.e. the point at which patients were depressed); follow-up 1 (FU1; to diagnose remission; t=0 for our prediction model study and corresponds with a 4-month follow-up for RCTs) and follow-up 2 (FU2; the intended prediction time and occurs at either 6 or 8 months after t=0; patient either relapses or does not relapse).

In all RCTs, the majority of participants are expected to meet criteria for a diagnosis of depression at baseline. Any participants identified to have a baseline PHQ-9 less than 10 will be excluded from the analysis. As described, FU1 is required to detect "remission" and FU2 to detect "relapse/no relapse."

The start-point (or time of intended prediction) is FU1, the point at which a patient, who started treatment with case-level depression, has entered remission. The PHQ-9 is a screening tool for major depressive disorder and a cut-off of 10 or more is used to detect clinically significant depressive symptoms [32]. Remission will be identified as a participant who had case-level depression at baseline (a PHQ-9 score of 10 or more) having (i) a post-treatment PHQ-9 score below the established cut-off of 10 at 4 months after trial baseline (this is consistent

Table 1 Summary of primary sources of IPD

Study	N	Study type	Inclusion criteria	Length of follow-up	Follow-up points	Mean age (SD)	Gender (% female)	RCT intervention	Duration of RCT intervention
CASPER Plus	485 (358 at 12 m)	RCT	65 years or older with depression	18 months	0, 4, 12 and 18 months	Intervention group: 71.9 (6.03) Control: 71.6 (5.96)	Intervention group: 59.1 Control: 63.1	Collaborative care	8–10 weeks
REEACT	461	RCT	Adults with depression	24 months	0, 4, 12 and 24 months	39.86 (12.65)	67	cCBT	6 weeks
REEACT 2	369	RCT	Adults with depression	12 months	0, 4 and 12 months	40.6 (13.8)	64.5	cCBT	4 months
COINCIDE	387	RCT	Adults with depression and multi-morbidity	24 months	0, 4, 12, 24	58.5 (11.7)	38	Collaborative care	3 months
COBRA	440	RCT	Adults with depression	18 months	0, 6, 12, 18	43.5 (14.1)	66	Behavioural activation vs CBT	16 weeks
CADET	581	RCT	Adults with depression	12 months	0, 4, 12	44.4 (13.3)	71.9	Collaborative care	14 weeks
Healthlines Depression	609	RCT	Adults with depression	12 months	0, 4, 8, 12	Intervention group: 49.1 (12.9) Control: 50 (12.8)	Intervention group: 69 Control: 68	Complex intervention (integrated telehealth)	12 months
WYLOW	439	Longitudinal observational cohort study	Adults with depression	12 months (Start-point = Remission)	Monthly	41.28 (14.59)	59.7	None (cohort are followed-up after LiCBT)	NA

with clinical recovery [30] as currently operationalized in the NHS Improving Access to Psychological Therapies (IAPT) service [3]) and (ii) an improvement of ≥5 points on the PHQ-9 (which aligns with the established reliable change index used to identify those with "reliable improvement" [33]).

End-point (relapse)

Patients will be coded as relapsed if they fulfil the following criteria within 6 to 8 months post-remission: (i) PHQ-9 score above the diagnostic cut-off (10 or more) and (ii) ≥5 points greater than their symptom score at the time of remission. As above, this is consistent with accepted criteria for reliable and clinically significant deterioration [33, 34].

The main reason for specifying the prediction end-point at 6 to 8 months rather than a single-time point is pragmatic and based on the available data (the time between FU1 and FU2 is 8 months for six of the seven RCTs and 6 months for COBRA). As discussed in the "Introduction" section, relapse is most commonly operationalized as occurring between 6 and 12 months post-remission [35] and the majority of patients who do relapse do so within the first 6 months [3]. Relapse by 6–8 months is felt to be an appropriate and sufficiently short-term timeframe for predictions to be meaningful

and clinically useful for patients and primary care professionals.

Predictors

We identified predictors based on literature review and on clinical grounds through discussion of a multidisciplinary group including members of the research team and the PPI group supporting the project. Umbrella reviews (reviews of other systematic reviews and meta-analyses) are one of the highest levels of evidence for determining associations between predictors and outcomes when selecting predictors for inclusion in a prognostic model [36]. A recently published umbrella review of prognostic factors associated with increased risk of relapse and recurrence guided the selection of candidate predictors for inclusion in the model [37]. A further systematic review of prognostic factors, published after the umbrella review, supported those findings and was also used to guide our included predictors [38]. In addition to this, we reviewed all existing prognostic models for predicting relapse or recurrence to explore other predictors used [18]. All candidate predictors are based on self-report or clinical information, and we have not included, for example, biomarkers and in-depth neuropsychological testing in an effort to ensure that the model is acceptable and usable in a primary care setting [39].

All included studies have information about key predictors, measured using reliable and validated tools. See the "Missing Data" section for details of how missing predictor information will be handled. Categorisation of continuous predictors will be avoided in order to avoid loss of information and power to detect an association between predictors and outcomes [19].

The following variables have robust evidence for their role as relapse predictors and will be included in the model (Table 2).

PHQ-9 score at remission (residual depressive symptoms)
Residual depressive symptoms is a strongly established predictor of relapse [37, 38] and will be operationalized in this study using the Patient Health Questionnaire (PHQ-9) score. The PHQ-9 is a validated tool for screening and case-finding for depression [32], routinely used in primary care. Remission is defined as a PHQ-9 score below 10 (remission), and residual symptoms are defined as a PHQ-9 at remission of between 5 and 9 [33]. Per the inclusion criteria for this study, all participants will meeting criteria for remission (i.e. PHQ-9 score of below 10); PHQ-9 score at remission (0–9) will be modelled as a continuous variable rather than binary (e.g. presence or absence of residual symptoms).

Number of previous episodes of depression
There is strong evidence that this is a significant predictor [37, 38], albeit slightly less strong than for residual symptoms. We plan to model this as a dichotomous predictor. The coding of this variable in the original RCTs is variable (i.e. a combination of continuous and dichotomous), and so it would not be possible to model as a continuous variable in this study. While there is some weak evidence that the relapse risk increases with each successive depressive episode, the prognostic effect of previous episodes on recurrence is strongest when comparing any number of previous episodes to no previous episodes [37]. This finding from the pre-existing literature is likely to be helpful for a primary care-based prognostic model, as there is potential difficulty in achieving a precise number of previous episodes in clinical practice. In this study, we will model this predictor as a dichotomous variable (0=no previous episodes, 1=one or more previous episodes) and will accept patient report, GP report or documentation in GP records.

Comorbid anxiety
There is good evidence that comorbid anxiety predicts relapse or recurrence of depression and will be included as a predictor in the model [37, 38]. The GAD-7 is a valid tool for screening and assessing severity of Generalised Anxiety Disorder score in clinical practice [40]. Pre-treatment symptoms (i.e. those at baseline) seem to be more predictive of relapse than those at depressive remission [38]. The pre-treatment GAD-7 score will be used provided it is available for all datasets; otherwise, we will use the GAD-7 at remission (t=0). GAD-7 score will be modelled as a continuous predictor.

Severity of episode
There is reasonable evidence that the baseline severity of the *index* episode is a prognostic indicator of greater odds of relapse [37]. This will be measured using the PHQ-9 score at baseline (pre-treatment) rather than that at the point of prognostication (remission). The PHQ-9 score at the point of depression diagnosis will be modelled as a continuous predictor.

RCT intervention
Because the data are drawn from RCTs, we must be mindful of the fact that approximately half of the participants have received a treatment (above usual care) and the other half have not. Where the treatments were found to be effective, not modelling the effect of different treatments can lead to unreliable risk predictions when the model is validated in a different population. Excluding the treated individuals would mean losing half of the available data, and so a preferable option is to explicitly model for treatment effect when developing a prognostic model [41, 42]. The treatments in all RCTs were acute-phase psychological treatments rather than relapse prevention interventions, and therefore, we do not know what their effect on relapse outcomes were.

Table 2 Summary of selected predictors

Predictor	Type of data	Method of measurement	Range of values and coding of predictors
PHQ-9 score at point of remission (residual depressive symptoms)	Continuous	PHQ-9 score at remission (t=0)	Range from 0 to 9
Number of previous episodes of depression	Categorical	Patient or GP report (No previous episodes vs any previous episodes)	No previous episodes=0; 1 or more previous episodes=1
Comorbid anxiety	Continuous	GAD-7 Score	0–21
Severity of episode	Continuous	PHQ-9 score at baseline	10–27
RCT intervention	Categorical	Presence or absence of effective treatment	Remission after receiving control=0; Remission after receiving intervention=1

One of the studies [31] did include an element of relapse prevention beyond the acute phase treatment (advisors phoned the patients every 2 months to check how they were getting on and encourage them to keep following the intervention advice). The interventions are also heterogeneous, and so it is possible that they affected relapse outcomes in different ways. To avoid overcomplicating the model, we will code the presence or absence of an effective intervention as a dichotomous variable. We will define an effective intervention by whether individual participants entered remission after receiving an RCT intervention (code=1) or whether they entered remission after receiving a control (code=0).

Exploratory predictors

We also plan to conduct an exploratory analysis investigating the role of the following less well established predictors: age; gender; ethnicity; employment status; relationship status; and multi-morbidity (Table 3). Age, gender and ethnicity are not well supported by the pre-existing evidence as being associated with relapse [37, 43, 44], but are routinely collected during RCTs and

Table 3 Summary and categorisation of exploratory predictors

Variable	Type of data	Method of measurement	Original categories (in RCTs)	New categories (PREDICTR)	Range of values and coding of predictors
Age	Continuous	RCT data collection/self report	Not applicable	Not applicable	
Gender	Categorical	RCT data collection/self report	Unchanged	Unchanged	Men=0; Women=1
Ethnicity	Categorical	RCT data collection/self report	White	White	White=0; Other=1
			Mixed	Other	
			Black		
			Asian		
			Chinese		
			Other		
Employment status	Categorical	RCT data collection/self report	Employed (Full time or part time)	Employed/not seeking employment	Unemployed=0; Employed or not seeking employment=1
			Student		
			Retired		
			House-person		
			Other		
			Unemployed job-seeker	Unemployed	
			Unemployed due to ill-health		
Relationship status	Categorical	RCT data collection/self report	Married/civil partnership/ cohabiting/relationship	In a relationship	Single=0; In a relationship=1
			Single	Single	
			Separated		
			Divorced		
			Widowed		
Multi-morbidity	Categorical	RCT data collection/self report	None	No long-term physical health condition	No long-term physical health conditions=0; One or more long-term physical health conditions=1
			Mental health only		
			Diabetes	One or more long-term physical health conditions	
			Asthma or COPD		
			Degenerative or inflammatory arthritis		
			Heart Disease		
			Stroke		
			Cancer		
			Kidney Disease		

often included as predictors in prognostic models [19]. There is weak evidence that employment status [45] and relationship status [46, 47] may be associated with an increased risk of relapse or recurrence. The National Institute for Health and Care Excellence (NICE) defines multi-morbidity as the presence of two or more long-term mental or physical health conditions [48]. The extant literature suggests that this is not associated with an increased risk of relapse or recurrence [44, 49]. The exploratory predictors described here are relevant to a primary care setting and, therefore, will be investigated outside of the planned principal analysis, depending on the completeness of the data and final sample size.

Sample size

Ensuring an adequate sample size will allow for more accurate estimation of regression coefficients and reduce the potential for overfitting. Rules of thumb for calculating required sample size for prediction models with binary outcomes (such as ten events per candidate predictor parameter (EPP)) are now considered too simplistic to provide robust estimates of minimum required sample size [50]. The actual required sample size is context-dependent and is informed by several factors. We used the *pmsampsize* package in Stata (available online: https://riskcalc.org/pmsamplesize/) to calculate our required minimum sample size [51].

The Cox-Snell R^2 is a measure of overall model fit and based on the method of Riley et al. [51] an anticipated Cox-Snell R^2 must be specified when calculating sample size, usually based on previous studies of similar patient groups/outcomes. No previous prognostic model study predicting relapse of depression identified so far has reported a Cox-Snell R^2 and so, to ensure an adequate minimum sample size, we used the recommended conservative estimated Nagelkerke R^2 of 15% [52]. This corresponds to a Cox-Snell R^2 of 0.0945, assuming an overall outcome proportion of 0.2, which again is a conservative estimate based on the literature [3]. We targeted an expected shrinkage factor (*S*) of 0.9 (to reflect small optimism in predictor effect estimates), as recommended [52].

To include all predictors, we require 8 predictor parameters (*P*), which corresponds to PHQ-9 score at remission, previous depressive episodes, co-morbid anxiety, severity of index episode and RCT Intervention (including 2 parameters for each continuous predictor to account for potential non-linear trends). Therefore, our minimum required sample size (*n*) is 722 (with 145 events) for these predictors. Our actual sample size exceeds this, and therefore, we anticipate that the study will be of a sufficient size to require minimal shrinkage and provide meaningful estimates of predictive performance.

Missing data

To avoid loss of power and precision, missing data will be handled using multiple imputation with chained equations (MICE) [53]. Missing values will be imputed based on other predictor and outcome values, under a missing at random assumption, and multiple copies of the dataset will be created with identical known information and different imputed values, reflecting the uncertainty associated with imputation. Imputation will be undertaken for each RCT separately, to preserve the clustering of participants within trials and any between-trial heterogeneity in predictor effects and outcome prevalence. We will assume that data are missing at random, unless this appears inappropriate upon inspection and discussion with original trialists. We will use the percentage of participants with one or more missing values to determine the number of imputations needed, in line with current guidance; at least 20, as long as this is greater than or equal to the percentage of participants with one or more missing values [19, 54]. Results from non-imputed and imputed data will be compared as a form of sensitivity analysis. Given the selection criteria, we do not anticipate any systematically missing predictors across datasets.

Statistical analysis methods
Data pre-processing

The datasets will be combined and harmonised to ensure consistency across trials. To assess IPD integrity, we will compare numbers of participants in each treatment arm with those reported in the primary references. We will check the relapse rate within each arm and compare these across datasets. To define the quality of the IPD for prognostic modelling, we will perform risk of bias assessment on the included datasets using the PROBAST [55]. Only the participants, predictors and outcome domains are pertinent; the analysis domain is used for assessment of prognostic model development and validation studies which do not apply to the RCTs included in this study.

Once remission has been identified this will represent time t=0. Relapse will then be coded as 0=no relapse, 1= relapse as described in "End-point" section. Descriptive statistics will be produced for all predictors and outcome data. Exploratory univariable analysis will be performed to evaluate the unadjusted relationship between each predictor variable and the outcome variable, but not for the purpose of informing predictor selection. We will explore percentage of cases that relapse over the different studies to assess comparability of data sources.

Model development

The model will be developed using a multilevel multivariable logistic regression, with a binary (relapse/no relapse) outcome. Model parameters will be estimated via

435

unpenalised maximum likelihood estimation and then penalised post-estimation using a uniform shrinkage factor (see later). The modelling will preserve the clustering of patients within trials, by having a random effect on the intercept, a random intervention effect and a random control effect, also allowing for between-study correlation in these pair of effects. If it is not possible to fit random effects in the multilevel logistic regression model, as originally planned, we will explore alternative modelling approaches. This would initially consist of a Generalised Estimating Equation model to control for the clustering without a random intercept. If this is also not possible, we will perform a single-level analysis with robust standard errors and accept that the limitation is that there may be a clustering effect that we are unable to properly control for.

Stepwise methods for predictor selection are not generally recommended for prediction models as this has been reported to remove judgment of the analyst from the process of model development as well as leading to *estimation bias* (estimating the performance of a prediction model after testing for statistical significance of predictors in the same data) [36]. We have selected our key predictors on the grounds of best available evidence and clinical acceptability, as well as practical reasons related to data availability. The list of predictors is felt to be of appropriate length, so we will avoid predictor selection techniques during model development and include all predictors regardless of their statistical significance ("full model" approach) [56]. This will also apply in the presence of multi-collinearity, which is not an issue for prediction purposes. We will only consider the need to exclude predictors due to collinearity if this is preventing convergence of the estimated model.

The full model approach described has the advantage of not being overly data-dependent and avoids the risk of removing clinically important predictors from the final model [56]. Calibration plots with loess smoothed calibration curves will be provided. Optimism will be measured and adjusted for using bootstrapping.

We will explore non-linear relationships in the modelling process using multivariable fractional polynomials (MFPs), a flexible and recommended approach for modelling continuous predictors in medical datasets. The other recommended method for modelling continuous predictors is the use of restricted cubic splines, and while these two methods often result in similar models, there is some evidence that MFPs perform better than restricted cubic splines in the presence of simpler relationships and medium amounts of information, as is the case here [20, 57]. We have factored in two predictor parameters (beta coefficients) per continuous variable to account for this approach, as described in the "Sample size" section.

Beyond the primary analysis outlined, and dependent on final sample size, an exploratory analysis will be performed investigating the role of less established relapse predictors (Table 3). Univariable associations between these predictors and outcome will be explored and, because the role of these variables as relapse predictors is less well understood, predictor selection through stepwise backward elimination will be used to develop an exploratory model. With sufficient sample size, stepwise backward elimination is an acceptable form of variable selection, performs similarly to other predictor selection approaches (for example, LASSO [58]) and is more compatible with our planned approaches for handling missing data and exploring non-linear trends. Guidance suggests using a more liberal p-value than the standard 0.05 for retention [19]; we will use a p-value of 0.157 or less as a stopping rule (consistent with Akaike information criteria (AIC) at one degree of freedom) [59].

Internal validation

The predictive performance and optimism of the developed model will be assessed. Calibration (a measure of the agreement between predictions from the model and observed outcomes) will be assessed by plotting observed vs predicted risks for groups defined by tenths of individual predicted risk (calibration plot) and by including a loess smoothed calibration curve across individuals (avoiding grouping). Apparent and optimism-adjusted calibration-in-the-large and calibration slope will be estimated. Discrimination (the ability of the model to differentiate between those who do or do not relapse) will be assessed using the C (concordance) index. The C-index assesses the extent to which the model assigns a higher probability of relapse to a patient who did eventually relapse in contrast to a patient who did not. The optimism-adjusted C-index will be derived using bootstrapping.

Optimism describes the risk of obtaining misleading measures of predictive accuracy when this is assessed in the same dataset used for model development, mainly due to overfitting. Internal validation can be used to provide optimism-corrected performance statistics can mitigate for this effect. Non-parametric bootstrapping will be used as a means of resampling the original dataset. This has the advantage, for example over a single split-sample approach, of allowing all of the data to be used in model development. Bootstrapping will be performed within each individual study, and then, these will be combined to create a new bootstrap sample to ensure studies are represented evenly for the final analysis. Multiple imputations for missing data will be performed within each bootstrap sample.

A bootstrap sample will be created in which the model development process will be repeated. The performance

of this model will be evaluated in the bootstrap sample (bootstrap, or apparent, performance) and in the original sample (test performance). This process will be repeated hundreds of times and the average difference between the bootstrap and test performance for each performance statistic provides the estimate of optimism for that statistic. Optimism-adjusted performance statistics will be derived by subtracting the average optimism estimate (from bootstrapping) from the apparent performance *of the original model*. The uniform shrinkage factor (calculated as the optimism-adjusted calibration slope) will then be applied to all estimated predictor effects to produce a penalised logistic regression model, and the intercept updated to ensure calibration-in-the-large.

Sensitivity, specificity and positive and negative predictive values for the model will be calculated at risk thresholds considered potentially clinically relevant. It is unclear whether the creation of risk groups is in the best interests of patients, but they are often used to guide clinical decision making [21]. In the absence of a gold standard test (as is the case here), the need for and definition of risk groups will be determined based on discussion within the research team and through consultation with our PPI group during model development. We will avoid basing risk thresholds on the data used to develop the model. The net benefit of the model at particular thresholds will also be examined using decision curve analysis and compared to treat all and treat none decisions [60].

External validation

External validation is the assessment of a model's predictive performance in data not used in the development process and is a measure of a model's generalisability and performance in a range of populations and settings. To conserve information and to allow for all data to be used for model development, we do not plan to perform a conventional external validation as part of this study. We do, however, have IPD from multiple studies, and therefore, generalisability and heterogeneity of the model performance will be examined using internal-external cross-validation (IECV) [61], as follows. We will exclude data from each primary study in turn and develop the risk prediction model using the remaining data. We will then externally validate the developed model using the data from the excluded study. This process will be repeated, each time omitting a different study, until the model has been fitted excluding each study once. Random effects meta-analysis will then be used to summarise the performance across studies, to obtain summary measures of the model performance and estimates of heterogeneity in performance across studies. We will ensure that each cycle of the IECV approach retains sufficient sample size for model development; in this manner, each cycle will retain the majority of the available IPD for model development, and so the final models produced in each cycle are likely to be similar to each other. A consistent model development strategy will be used in each cycle of the IECV approach [62].

A sensitivity analysis will be performed measuring predictive performance statistics omitting, first, the observational cohort data (WYLOW) and, secondly, the RCT (COBRA) with relapse at 6, rather than 8, months. If our risk of bias (PROBAST) assessment identifies any studies that are not at overall low risk of bias, we will also perform a sensitivity analysis omitting these studies.

Discussion

We have reported a protocol for the development and validation of a novel prognostic model to predict depressive relapse in a primary care setting. As discussed, we have used an up-to-date review of the extant literature to guide predictor selection and our sample size is in excess, relative to the number of predictor parameters, of those used in previous prognostic model studies. We now briefly discuss our anticipated next steps, beyond this prognostic model study.

It is envisaged that this statistical model could form the basis for a clinical tool, to be embedded in GP IT systems, to help identify patients who are at higher risk of relapse. Longer term, and with further research, a decision tool could be developed to help inform decisions as to which patients with remitted depression should receive relapse prevention interventions. Provided we are able to demonstrate sufficient predictive accuracy during the validation stages, the model should undergo external validation (in a different dataset, to assess generalizability) and, ideally, independent validation (by a different research team, to reduce risk of bias). External validation could be done on either an unrelated retrospective dataset or, preferably, a prospective dataset collected specifically for this purpose. Finally, the impact of the model should be evaluated, and the gold standard way of doing this is through a randomised controlled trial with clinically meaningful outcome measures [63].

Qualitative work with stakeholders will be used to decide the extent to which the model can be implemented and will guide the evaluation of the model in practice, including plans for impact assessment. In particular, Cuijpers recently highlighted the importance of assessing the effect of mental health treatments on patient-defined outcomes (e.g. quality of life and functional outcomes) as well as those determined to be important by researchers or clinicians [64]. This is applicable to health technologies, like prognostic models, and exploring patient-defined outcomes will form a part of our evaluation process beyond this study.

Limitations

The ideal dataset for developing a prognostic model is a prospective, pre-designed cohort study. The advantage of such an approach is that investigators retain control over inclusion and exclusion criteria, definition and measurement of predictors and outcomes, ensure appropriate timings, reduce missing data and minimize other potential biases (for example, selecting bias and blinding). However, the costs (financial and time) of carrying out a prospective study would be substantial and secondary analysis of good quality data from RCT and other cohorts is an accepted alternative [23]. We are mindful of the potential problems with this approach, particularly the risk of missing data (that we have planned for) and the chance that predictors and outcomes may not be recorded optimally. We are reliant on the quality of the initial data collection with respect to this latter point, and we are confident that the studies included are of a high standard.

A further common pitfall of RCTs is the narrow eligibility criteria often stipulated which can impact on the generalizability of any findings to the target population of interest (in our case, a primary care patient population). We are reassured that the eligibility criteria for included studies were inclusive and pragmatic with relatively small numbers of participants with missing data. We do however recognise that RCT participants may differ from the general population in important ways and results should be interpreted with this in mind.

In the planning stage of this project, we considered other data sources, in particular the Clinical Practice Research Datalink (CPRD), a large electronic database of routinely collected follow-up data from primary care. Following discussions with CPRD experts at the University of York, it was evident that the coding of measures of relapse and recurrence were not optimal for identifying patients who relapsed and that this would have limited our ability to develop a reliable and generalisable model.

Further limitations relate to measurement of start-point (remission) and end-point (relapse or not), which will be measured using PHQ-9 score. The gold standard would have been to use diagnostic interviews, which may have been possible with a prospective cohort study. The PHQ-9 is a validated and widely used tool with good sensitivity and specificity [65], and the large sample size (possible because of the use of secondary data analysis) should compensate for this. A further point to consider is that the start- and end-points are defined at the next time-point they were actually measured rather than necessarily capturing the precise "real-world" moment of remission. However, this reflects the situation in general practice, where diagnostic tools will be applied at patient consultation rather than in real time.

Therefore, we feel this is justifiable and actually mirrors the clinical picture accurately. We will use the reliable clinical recovery and deterioration definitions (sample size allowing) to ensure robust coding of start- and end-points.

In the event that multilevel modelling with a random intercept and random effects on the intervention/control variable is not possible, we will be required to make an assumption that the effects of the different interventions and controls in the RCTs were homogenous. It is not likely that the interventions had a significant effect on relapse rates, even where they did improve acute depression symptoms. However, it is possible that one or more of the interventions (or controls) did exert an effect on relapse of which we are not aware. We will take a pragmatic approach to modelling this, following the steps that we have outlined in the "Methods" section. A further limitation is that the data we plan to analyse do not allow for survival analysis, as the follow-up time-points were insufficiently similar and infrequent. However, time to relapse is important and would increase our understanding; future prospective work should consider this when designing strategies for data collection.

There are some predictors not included due to lack of relevance and usefulness to GPs. For example, neuroticism (the personality trait), childhood maltreatment and rumination have been found to be associated with increased risk of relapse and recurrence [37], as has duration of index episode of depression and age at onset of first episode of depression [66]. These are not routinely measured in practice and have not been coded for in our cohorts; they will therefore not be included as predictors. The cohort has been designed to be as undifferentiated as possible to represent a GP case-mix. Increased predictive performance would be more likely if we were to be very specific in defining this cohort, but this would have implications for its utility in the real-world primary care setting.

In summary, this study will derive a statistical model aiming to predict relapse. If it demonstrates sufficient predictive performance, it could be used to guide the allocation of interventions to prevent relapse in a primary care setting, improving outcomes for patients and ensuring efficient use of healthcare resources.

Moriarty et al. Diagnostic and Prognostic Research   (2021) 5:12

Page 11 of 13

CT: Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy; WYLOW: West Yorkshire Low-intensity Outcomes Watch

Authors' contributions

ASM is the lead author, conceived the study, and was responsible for writing the first draft of the manuscript. LWP, KIES and RDR have provided methodological expertise and contributed to the statistical analysis plan. NM and RPS provided specific methodological expertise. JEJB and SP provided content and methodological expertise and commented on the final manuscript. SG, DM, CCG and SA were all involved in the conception of the study. PAC, DAR, CS and JD provided data for inclusion in the PREDICTR dataset and have commented on the final manuscript. The authors read and approved the final manuscript.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

The University of York's Health Sciences Research Governance Committee confirmed that this study is exempt from full ethical approval as it entails the secondary analysis of anonymised data from studies that had already received ethical approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Health Sciences, University of York, York, England. [2]Hull York Medical School, University of York, York, England. [3]Centre for Prognosis Research, School of Medicine, Keele University, Keele, England. [4]Centre for Outcomes and Research Effectiveness, Research Department of Clinical, Educational and Health Psychology, University College London, London, England. [5]iCope – Camden and Islington Psychological Therapies Services, Camden & Islington NHS Foundation Trust, London, England. [6]School of Medicine, Keele University, Keele, England. [7]Department of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada. [8]Camden & Islington NHS Foundation Trust, St Pancras Hospital, London, England. [9]Centre for Reviews and Dissemination, University of York, York, England. [10]Department of Psychology, University of Sheffield, Sheffield, England. [11]Institute of Health Research, College of Medicine and Health, University of Exeter, Exeter, England. [12]Department of Health and Caring Sciences, Western Norway University of Applied Sciences, Inndalsveien 28, 5063 Bergen, Norway, USA. [13]Centre for Academic Primary Care, University of Bristol, Bristol, England.

References

1. Geneva: World Health Organization. Depression and Other common mental disorders: global health estimates. Geneva: World Health Organization; 2017.
2. Rait G, Walters K, Griffin M, Buszewicz M, Petersen I, Nazareth I. Recent trends in the incidence of recorded depression in primary care. Br J Psychiatry. 2009;195(6):520–4. https://doi.org/10.1192/bjp.bp.108.058636.
3. Ali S, Rhodes L, Moreea O, McMillan D, Gilbody S, Leach C, et al. How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. Behav Res Ther. 2017; 94:1–8. https://doi.org/10.1016/j.brat.2017.04.006.
4. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. JAMA Psychiatry. 1991;48(9):851–5.
5. Bockting CL, Hollon SD, Jarrett RB, Kuyken W, Dobson K. A lifetime approach to major depressive disorder: the contributions of psychological interventions in preventing relapse and recurrence. Clin Psychol Rev. 2015; 41:16–26. https://doi.org/10.1016/j.cpr.2015.02.003.
6. Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. Neuropsychopharmacology. 2006;31(9):1841–53. https://doi.org/1 0.1038/sj.npp.1301131.
7. Riso LP, Thase ME, Howland RH, Friedman ES, Simons AD, Tu XM. A prospective test of criteria for response, remission, relapse, recovery, and recurrence in depressed patients treated with cognitive behavior therapy. J Affect Disord. 1997;43(2):131–42. https://doi.org/10.1016/S01 65-0327(96)01420-6.
8. Kendler KS, Thornton LM, Gardner CO. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the "kindling" hypothesis. Am J Psychiatry. 2000;157(8):1243–51. https://doi. org/10.1176/appi.ajp.157.8.1243.
9. Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG. Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. Br J Psychiatry. 1996;168(S30):17–30. https://doi.org/10.1192/S0007125000298371.
10. Burcusa SL, Iacono WG. Risk for recurrence in depression. Clin Psychol Rev. 2007;27(8):959–85. https://doi.org/10.1016/j.cpr.2007.02.005.
11. Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT, et al. Multiple recurrences of major depressive disorder. Am J Psychiatry. 2000; 157(2):229–33. https://doi.org/10.1176/appi.ajp.157.2.229.
12. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338(mar31 1):b604. https://doi.org/10.1136/bmj.b604.
13. Klein NS, Holtman GA, Bockting CLH, Heymans MW, Burger H. Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression. J Psychiatr Res. 2018;104:1–7. https://doi.org/10.1016/j.jpsychires.2018.06.006.
14. van Loo HM, Aggen SH, Gardner CO, Kendler KS. Multiple risk factors predict recurrence of major depressive disorder in women. J Affect Disord. 2015; 180:52–61. https://doi.org/10.1016/j.jad.2015.03.045.
15. Van Loo HM, Aggen SH, Gardner CO, Kendler KS. Sex similarities and differences in risk factors for recurrence of major depression. Psychol Med. 2018;48(10):1685–93. https://doi.org/10.1017/S0033291717003178.
16. van Loo HM, Bigdeli TB, Milaneschi Y, Aggen SH, Kendler KS. Data mining algorithm predicts a range of adverse outcomes in major depression. J Affect Disord. 2020;276:945–53. https://doi.org/10.1016/j.jad.2020.07.098.
17. 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(5):451–7. https://doi.org/10.1002/da.22215.
18. Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Chew-Graham CA, Gilbody S, Churchill R, Phillips RS, Ali S, McMillan D. Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. Cochrane Database of Systematic Reviews 2021, Issue 5. Art. No.: CD013491. DOI: 10.1002/14651858.CD013491.pub2.
19. Riley RD, van der Windt D, P C, Moons K. Prognosis research in healthcare: concepts, methods, and impact. Oxford: 1st. Oxford University Press; 2019. pp 139–187.

Moriarty et al. Diagnostic and Prognostic Research   (2021) 5:12

Page 12 of 13

20. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. 2nd ed. Switzerland: Springer Nature; 2019. https://doi.org/10.1007/978-3-030-16399-0.

21. 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. https://doi.org/10.7326/M14-0698.

22. 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: the PRISMA-IPD statement. JAMA J Am Med Assoc. 2015;313(16):1657–65. https://doi.org/10.1001/jama.2015.3656.

23. Pajouheshnia R, Groenwold RHH, Peelen LM, Reitsma JB, Moons KGM. When and how to use data from randomised trials to develop or validate prognostic models. BMJ. 2019;2154.

24. Bosanquet K, Adamson J, Atherton K, Bailey D, Baxter C, Beresford-Dent J, et al. Collaborative care for screen-positive elders with major depression (CASPER plus): a multicentred randomized controlled trial of clinical effectiveness and cost-effectiveness. Health Technol Assess (Rockv). 2017; 21(67):1–251. https://doi.org/10.3310/hta21670.

25. 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. BMJ. 2015;351:1–13.

26. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, et al. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. Br J Psychiatry. 2017;210(5):362–7. https://doi.org/10.1192/bjp.bp.116.192435.

27. Coventry P, Lovell K, Dickens C, Bower P, Chew-Graham C, McElvenny D, et al. Integrated primary care for patients with mental and physical multimorbidity: cluster randomised controlled trial of collaborative care for patients with depression comorbid with diabetes or cardiovascular disease. BMJ. 2015;350:1–11.

28. 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. 2020;4:1–20.

29. Richards DA, Ekers D, McMillan D, Taylor RS, Byford S, Warren FC, et al. Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. Lancet. 2016;388(10047):871–80. https://doi.org/10.1016/S0140-6736(16)31140-0.

30. 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. 2013;347(4913). https://www.bmj.com/content/bmj/347/bmj.f4913.full.pdf.

31. Salisbury C, O'Cathain A, 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. Lancet Psychiatry. 2016;3(6):515–25. https://doi.org/10.1016/S2215-0366(16)00083-3.

32. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

33. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord. 2010;127(1–3):122–9. https://doi.org/10.1016/j.jad.2010.04.030.

34. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991; 59(1):12–9. https://doi.org/10.1037/0022-006X.59.1.12.

35. Beshai S, Dobson KS, Bockting CLH, Quigley L. Relapse and recurrence prevention in depression: current research and future prospects. Clin Psychol Rev. 2011;31(8):1349–60. https://doi.org/10.1016/j.cpr.2011.09.003.

36. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. JAMA Psychiatry. 2018;75(12):1280–8.

37. 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;64:13–38. https://doi.org/10.1016/j.cpr.2018.07.005.

38. Wojnarowski C, Firth N, Finegan M, Delgadillo J. Predictors of depression relapse and recurrence after cognitive behavioural therapy: a systematic

39. Kessler RC. The potential of predictive analytics to provide clinical decision support in depression treatment planning. Curr Opin Psychiatry. 2018;31(1): 32–9. https://doi.org/10.1097/YCO.0000000000000377.

40. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10): 1092–7. https://doi.org/10.1001/archinte.166.10.1092.

41. Groenwold RHH, Moons KGM, Pajouheshnia R, Altman DG, Collins GS, Debray TPA, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. J Clin Epidemiol. 2016;78:90–100. https://doi.org/10.1016/j.jclinepi.2016.03.017.

42. Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: a simulation study. BMC Med Res Methodol. 2017;17(1):1–12.

43. Burcusa SL, Iacono WG. Risk for recurrence in depression. Clin Psychol Rev. 2007;27(8):959–85.

44. Wojnarowski C, Firth N, Finegan M, Delgadillo J. Predictors of depression relapse and recurrence after cognitive behavioural therapy: a systematic review and meta-analysis. Behav Cogn Psychother. 2019;47(5):514–29.

45. LorimerB, Delgadillo J, Kellett S, Lawrence J. Dynamic prediction and identification of cases at risk of relapse following completion of low-intensity cognitive behavioural therapy. Psychother Res. 2021;31(1):19–32.

46. 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. Depression and Anxiety. 2014;31(5):451–7.

47. Johansson O, Lundh LG, Bjärehed J. Twelve-month outcome and predictors of recurrence in psychiatric treatment of depression: a retrospective s. Psychiatr Q. 2015;86(3):407–17.

48. National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. 2016.

49. Kok GD, Bockting CLH, Burger H, Hannig W, Pijnenborg GHM, Cuijpers P, Hollon SD, Laks J. Double trouble: does co-morbid chronic somatic illness increase risk for recurrence in depression? A systematic review. PLoS ONE. 2013;8(3):e57510.

50. Van Smeden M, De Groot JAH, Moons KGM, Collins GS, Altman DG, Eijkemans MJC, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 2016;16(1):1–12.

51. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;18:m441.

52. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2018:1–21.

53. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol. 2019;16. 1294-1304.

54. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377–99. https://doi.org/10.1002/sim.4067.

55. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1–33. https://doi.org/10.7326/M18-1377.

56. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. In: eBook. Statistical Methods in Medical Research, vol. 13. Switzerland: Springer International Publishing; 2015.

57. Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. Stat Med. 2013; 32(13):2262–77. https://doi.org/10.1002/sim.5639.

58. van Houwelingen HC, Sauerbrei W. Cross-validation, shrinkage and variable selection in linear regression revisited. Open J Stat. 2013;03(02):79–102.

59. Vergouwe Y, RoystonP, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. J Clin Epidemiol. 2010;63(2):205–14.

60. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:3–7.

440

61. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. Stat Med. 2004;23(6):907–26. https://doi.org/10.1002/sim.1691.
62. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7 Elsevier USA.
63. Hingorani AD, Van Der Windt DA, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:1–9.
64. Cuijpers P. Expert Review of Neurotherapeutics Measuring success in the treatment of depression : what is most important to patients ? Measuring success in the treatment of depression : what is most important to. Expert Rev Neurother [Internet]. 2020;00(00):1–3.
65. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry. 2015;37(6):567–76. https://doi.org/10.101 6/j.genhosppsych.2015.06.012.
66. Berlanga C, Heinze G, Torres M, Apiquián R, Caballero A. Personality and clinical predictors of recurrence of depression. Psychiatr Serv. 1999;50(3): 376–80. https://doi.org/10.1176/ps.50.3.376.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix 5.2: PRISMA-IPD checklist

PRISMA-IPD Section/topic	Item No	Checklist item	Reported on page
Title			
Title	1	Identify the report as a systematic review and meta-analysis of individual participant data.	115
Abstract			
Structured summary	2	Provide a structured summary including as applicable:	115
		Background: state research question and main objectives, with information on participants, interventions, comparators and outcomes.	116
		Methods: report eligibility criteria; data sources including dates of last bibliographic search or elicitation, noting that IPD were sought; methods of assessing risk of bias.	117
		Results: provide number and type of studies and participants identified and number (%) obtained; summary effect estimates for main outcomes (benefits and harms) with confidence intervals and measures of statistical heterogeneity. Describe the direction and size of summary effects in terms meaningful to those who would put findings into practice.	123
		Discussion: state main strengths and limitations of the evidence, general interpretation of the results and any important implications.	177
		Other: report primary funding source, registration number and registry name for the systematic review and IPD meta-analysis.	115
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	115
Objectives	4	Provide an explicit statement of the questions being addressed with reference, as applicable, to participants, interventions, comparisons, outcomes and study design (PICOS). Include any hypotheses that relate to particular types of participant-level subgroups.	115
Methods			
Protocol and registration	5	Indicate if a protocol exists and where it can be accessed. If available, provide registration information including registration number and registry name. Provide publication details, if applicable.	115
Eligibility criteria	6	Specify inclusion and exclusion criteria including those relating to participants, interventions, comparisons, outcomes, study design and characteristics (e.g. years when conducted, required minimum follow-up). Note whether these were applied at the	118, 121

		study or individual level i.e. whether eligible participants were included (and ineligible participants excluded) from a study that included a wider population than specified by the review inclusion criteria. The rationale for criteria should be stated.	
Identifying studies - information sources	7	Describe all methods of identifying published and unpublished studies including, as applicable: which bibliographic databases were searched with dates of coverage; details of any hand searching including of conference proceedings; use of study registers and agency or company databases; contact with the original research team and experts in the field; open adverts and surveys. Give the date of last search or elicitation.	118
Identifying studies - search	8	Present the full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	NA
Study selection processes	9	State the process for determining which studies were eligible for inclusion.	118
Data collection processes	10	Describe how IPD were requested, collected and managed, including any processes for querying and confirming data with investigators. If IPD were not sought from any eligible study, the reason for this should be stated (for each such study).	136
		If applicable, describe how any studies for which IPD were not available were dealt with. This should include whether, how and what aggregate data were sought or extracted from study reports and publications (such as extracting data independently in duplicate) and any processes for obtaining and confirming these data with investigators.	NA
Data items	11	Describe how the information and variables to be collected were chosen. List and define all study level and participant level data that were sought, including baseline and follow-up information. If applicable, describe methods of standardising or translating variables within the IPD datasets to ensure common scales or measurements across studies.	124-132
IPD integrity	A1	Describe what aspects of IPD were subject to data checking (such as sequence generation, data consistency and completeness, baseline imbalance) and how this was done.	137
Risk of bias assessment in individual studies.	12	Describe methods used to assess risk of bias in the individual studies and whether this was applied separately for each outcome. If applicable, describe how findings of IPD checking were used to inform the assessment. Report if and how risk of bias assessment was used in any data synthesis.	137
Specification of outcomes and effect measures	13	State all treatment comparisons of interests. State all outcomes addressed and define them in detail. State whether they were pre-specified for the review and, if applicable, whether they were primary/main or secondary/additional outcomes. Give the principal measures of effect (such as risk ratio, hazard ratio, difference in means) used for each outcome.	NA

Synthesis methods	14	Describe the meta-analysis methods used to synthesise IPD. Specify any statistical methods and models used. Issues should include (but are not restricted to): • Use of a one-stage or two-stage approach. • How effect estimates were generated separately within each study and combined across studies (where applicable). • Specification of one-stage models (where applicable) including how clustering of patients within studies was accounted for. • Use of fixed or random effects models and any other model assumptions, such as proportional hazards. • How (summary) survival curves were generated (where applicable). • Methods for quantifying statistical heterogeneity (such as I^2 and τ^2). • How studies providing IPD and not providing IPD were analysed together (where applicable). • How missing data within the IPD were dealt with (where applicable).	137-143
Exploration of variation in effects	A2	If applicable, describe any methods used to explore variation in effects by study or participant level characteristics (such as estimation of interactions between effect and covariates). State all participant-level characteristics that were analysed as potential effect modifiers, and whether these were pre-specified.	137-140
Risk of bias across studies	15	Specify any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to not obtaining IPD for particular studies, outcomes or other variables.	NA
Additional analyses	16	Describe methods of any additional analyses, including sensitivity analyses. State which of these were pre-specified.	144
Results			
Study selection and IPD obtained	17	Give numbers of studies screened, assessed for eligibility, and included in the systematic review with reasons for exclusions at each stage. Indicate the number of studies and participants for which IPD were sought and for which IPD were obtained. For those studies where IPD were not available, give the numbers of studies and participants for which aggregate data were available. Report reasons for non-availability of IPD. Include a flow diagram.	NA
Study characteristics	18	For each study, present information on key study and participant characteristics (such as description of interventions, numbers of participants, demographic data, unavailability of outcomes, funding source, and if applicable duration of follow-up). Provide (main) citations for each study. Where applicable, also report similar study characteristics for any studies not providing IPD.	152-154
IPD integrity	A3	Report any important issues identified in checking IPD or state that there were none.	154

Risk of bias within studies	19	Present data on risk of bias assessments. If applicable, describe whether data checking led to the up-weighting or down-weighting of these assessments. Consider how any potential bias impacts on the robustness of meta-analysis conclusions.	155
Results of individual studies	20	For each comparison and for each main outcome (benefit or harm), for each individual study report the number of eligible participants for which data were obtained and show simple summary data for each intervention group (including, where applicable, the number of events), effect estimates and confidence intervals. These may be tabulated or included on a forest plot.	152
Results of syntheses	21	Present summary effects for each meta-analysis undertaken, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified, and report the numbers of studies and participants and, where applicable, the number of events on which it is based.	161
		When exploring variation in effects due to patient or study characteristics, present summary interaction estimates for each characteristic examined, including confidence intervals and measures of statistical heterogeneity. State whether the analysis was pre-specified. State whether any interaction is consistent across trials.	
		Provide a description of the direction and size of effect in terms meaningful to those who would put findings into practice.	
Risk of bias across studies	22	Present results of any assessment of risk of bias relating to the accumulated body of evidence, including any pertaining to the availability and representativeness of available studies, outcomes or other variables.	NA
Additional analyses	23	Give results of any additional analyses (e.g. sensitivity analyses). If applicable, this should also include any analyses that incorporate aggregate data for studies that do not have IPD. If applicable, summarise the main meta-analysis results following the inclusion or exclusion of studies for which IPD were not available.	171-172
Discussion			
Summary of evidence	24	Summarise the main findings, including the strength of evidence for each main outcome.	176-177
Strengths and limitations	25	Discuss any important strengths and limitations of the evidence including the benefits of access to IPD and any limitations arising from IPD that were not available.	176-177; 314-318
Conclusions	26	Provide a general interpretation of the findings in the context of other evidence.	290-294
Implications	A4	Consider relevance to key groups (such as policy makers, service providers and service users). Consider implications for future research.	321-327

Funding			
Funding	27	Describe sources of funding and other support (such as supply of IPD), and the role in the systematic review of those providing such support.	15

A1 – A3 denote new items that are additional to standard PRISMA items. A4 has been created as a result of re-arranging content of the standard PRISMA statement to suit the way that systematic review IPD meta-analyses are reported.

Table A5.1: CIS-R anxiety subscale

Item	Description	Score
Compulsions	A series of questions around compulsions	+1 if repeating actions >3 days in past week +1 if attempted to stop +1 if action is upsetting +1 if an action was repeated more than twice
Anxiety	A series of questions around anxiety	+1 if anxious for >3days in past week +1 if causes feeling of unpleasantness +1 if causes physical symptoms +1 if anxious >3hrs in any day
Irritability	A series of questions around feelings of irritability, anger or short-temper	+1 if consistent during the past week (>3 days) +1 if feeling lasted >1hr in any day during past week +1 if shouted or felt like shouting +1 if lost temper without reason
Worry	A series of questions around worry	+1 if worry persists for >3days in past week +1 if excessively worried +1 if worry was unpleasant +1 if worried >3hrs in any day
Panic	A series of questions around panic	+1 if panic occurred once in past week +1 if panic occurred at least once more in past week +1 if panic attack lasted longer than 10 mins +1 if panic attack was unpleasant
Phobias	A series of questions around phobias	+1 if anxious for >3days in past week +1 if causes physical symptoms +1 if avoidance action taken for at least 1day +1 if avoidance action taken for more than 3days
Obsessions	A series of questions around the presence of obsessive thoughts	+1 if consistent during the past week (>3 days) +1 if tried to stop +1 if they are upsetting +1 lasted for at least 15 mins
Health anxiety	A series of questions about concern over health or future health	+1 if consistent during the past week (>3 days) +1 if considered excessive +1 if considered unpleasant +1 if difficult to stop worrying

Somatic concerns	A series of questions around the presence of aches/pains or bodily discomfort	+1 if consistent during the past week (>3 days). +1 if consistent and lasted at least 3 in any day during last week +1 if consistent and unpleasant +1 if consistent and bothersome during 'interesting' activity

Appendix 5.4: Master codebook for IPD Harmonisation

Table A5.2: Master codebook for IPD Harmonisation

Code	Description of code	Type of data	Method of measurement	Range of values and coding of predictors
ID_orig	ID in original study	Identifier (individual participant)	NA	NA
ID_PREDICTR	ID assigned for purpose of PREDICTR IPD dataset	Identifier (individual participant)	NA	NA
RCT	Coding for individual studies	Identifier (cluster)	NA	CADET=1 CASPER Plus=2 COBRA=3 Healthlines=4 REEACT=5 REEACT-2=6 WYLOW=7
PHQ_baseline	PHQ-9 at Baseline of RCT	Continuous	PHQ-9	0-27
GAD_baseline	GAD-7 at baseline	Continuous	GAD-7	0-21
PHQ_FU1	t=0 for PREDICTR	Continuous	PHQ-9	0-27
PHQ_FU2	6-8 months post-FU1	Continuous	PHQ-9	0-27
followup_time_(FU2-FU1)	How long has elapsed between FU1 and FU2	NA	NA	6 or 8 (months)
remit	Has the pt remitted at FU1? If PHQ-9 score at baseline > 10 and less	Binary	NA	Yes=1, No=0

	than 10 at FU1 (plus change of 5 or more points)			
relapse	Has the pt relapsed at FU2? If remission at FU1 and PHQ-9 >10 (plus change of 5 or more points)	Binary	NA	Yes=1, No=0
residual_sympto ms (=PHQ_FU1)	PHQ-9 at remission	Continuo us	PHQ-9 score at remission	0-9
prev_eps	Number of previous episodes of depression	Categori cal	Patient or GP report (No previous episodes vs any previous episodes)	No previous episodes=0; 1 or more previous episodes=1
comorbid_anx (=GAD_baseline)	Comorbid anxiety	Continuo us	GAD-7 Score	0-21
comorbid_anx_zs core	Comorbid anxiety (z score)	Continuo us	Combined z score based on mean and SD within original study dataset GAD-7 for all studies other than REEACT For REEACT, z score of CIS-R anxiety subscale	NA

Severity (=PHQ_baseline)	Severity of depression at baseline	Continuous	PHQ-9 score at baseline	10-27
RCT_intervention	Presence or absence of RCT intervention	Categorical	Presence or absence of treatment	Absence of effective treatment (control arm or non-effective intervention) =0; Presence of treatment (effective intervention arm)=1
age	Age in years	Continuous	Self-report	
gender	Gender	Categorical	Self-report	Male=0, Female=1
ethnicity	Ethnicity	Categorical	Self-report	White=0 Other=1
employment	Employment status	Categorical	Self-report	Unemployed =0; Employed=1
relationship	Relationship status	Categorical	Self-report	In a relationship= 1; Not in a relationship= 0
multi_morbidity	Multi-morbidity	Categorical	Self-report	No long-term physical health condition=0; One or more long-term physical health conditions=1
ADM_current	Current antidepressant medication (ADM)	Categorical	Self-report	Not currently taking ADM at remission=0; Taking ADM at remission=1

DEPARTMENT OF HEALTH SCIENCES

c/o Department of Philosophy
Heslington
York YO10 5DD

Telephone	(01904) 323253
Fax	(01904) 321383
E-mail	smh12@york.ac.uk

Prof Stephen Holland
Chair, Health Sciences Research Governance Committee

www.york.ac.uk/healthsciences

7 February 2020

Andrew Moriarty
NIHR Doctoral Research Fellow
Department of Health Sciences and the Hull York Medical School
University of York
York YO10 5DD

Dear Andrew

Depression in primary care: Development and validation of a prognostic model to predict relapse and evaluation of relapse prevention interventions

Thank you for your email of 4 February, including the successful funding application for your NIHR Fellowship project.

I am writing to confirm that the project does not require HSRGC review, because it involves secondary analysis of anonymised data from trials that received ethical approval.

Yours sincerely

Stephen Holland
Chair: HSRGC

cc. *Prof Paul Galdas, Prof Dean McMillan, Dr Lewis Paton*

Appendix 6.1: Categorical predictors

Table A6.1: Summary of categorical predictors and their coding in original datasets

Predictor variable		Study (total number of participants in study)						
		CADET (n=158)	CASPER Plus (n=101)	COBRA (n=169)	Healthlines Depression (n=110)	REEACT (n=221)	REEACT-2 (n=159)	WYLOW (n=326)
RCT Intervention	Effective RCT Intervention	Collaborative care: n=87	Collaborative care: n=56	CBT: n=92 BA: n=77	Healthlines Integrated Telehealth intervention: n=67	RCT Intervention (cCBT) was not effective	MoodGym with telephone support: n=80	Not applicable. All participants received LiCBT through IAPT.
	Ineffective RCT Intervention or Control	Usual care: n=72	Usual care: n=45	No usual care arm: n=0	Usual care: n=43	cCBT ("Beating the Blues"): n=66; cCBT ("MoodGym"): n=78; Usual care: n=77	Guided self-help with telephone support: n=30; MoodGym only: n=49	
Ethnicity	White	White: n=137	White: n=100	White British: n=155; White Irish: n=4; White (other): n=6	White: n=107	White British: n=208; Any other White background: n=8	White British: n=145; White Irish: n=3; Any other White background: n=6	White: n=292
	Non-white	Asian or Asian British: n=8; Black or Black British: n=7; Mixed: n=3; Other: n=3	Black or Black British: n=1	Other Asian: n=1; Black African: n=1; Other=1; Prefer not to say: n=1	Mixed: n=2; Other: n=1	Asian or Asian British: n=1; Chinese: n=1; Japanese: n=1; Jewish: n=1	Asian or Asian British (Indian): n=1 Chinese: n=1 Other (not specified): n=2	Mixed: n=7; Asian: n=5; Black: n=4; Chinese: n=1; Other: n=2

	Missing	Missing: n=0	Missing: n=0	Missing: n=0	Missing: n=0	Missing: n=0	Missing: n=1	Missing: n=16
Relationship status	In a relationship	Married/living as married: n=69		Cohabiting: n=19; Civil partnership: n=1; Married: n=86		Married: n=100; Living with a partner: n=38; In a relationship: n=2	Married: n=71; Living with partner: n=22	
	Not in a relationship	Single: n=49; Separated: n=13; Divorced: n=23; Widowed: n=4		Single: n=32; Divorced / separated: n=31		Divorced / separated: n=21; Widowed: n=2; Never married: n=22	Divorced/separated: n=21; Widowed: n=4; Single: n=39; Other (not specified): n=1	
	Missing	Missing: n=0		Missing: n=0		Missing: n=36	Missing: n=1	
Employment status	Employed or not seeking employment	Full-time paid or self employment: n=66; Part-time paid or self employment: n=29; Voluntary employment: n=1; Student: n=3; Housewife / husband: n=12; Retired: n=13; Other=9		Employed / student: n=120; Retired: n=12	Full time work: n=35; Part time work: n=18; Full time education: n=1; Unable to work (illness): n=15; Unable to work (carer): n=1; Retired: n=20; Looking after home: n=7; Other: n=9	Employed part-time: n=37; Employed full-time: n=92; Self=employed: n=21; Retired: n=14; Looking after family or home: n=9; Not employed (ill health): n=15; Not employed but not seeking: n=3; Other: n=22	Employed part-time: n=26; Employed full-time: n=62; Self-employed: n=15; Retired: n=16; Looking after family or home: n=5; Not employed but not seeking work due to ill health: n=4; Full-time student: n=16; Other (job lined up): n=1	Employed, retired, student, homemaker: n=214

	Unemployed	Unemployed: n=27		Not employed: n=37	Unemployed: n=4	Not employed but seeking work: n=8	Not employed but seeking work: n=10	Unemployed: n=112
	Missing	Missing: n=0		Missing: n=0	Missing: n=1	Missing: n=0	Missing: n=0	Missing: n=0
Multimorbidity	Multi-morbidity	Diabetes: n=2; Asthma: n=14; Arthritis: n=7; Heart disease: n=3; High blood pressure: n=11; More than one of the above: n=11; Other: n=40	At least one of diabetes, osteoporosis, hypertension, rheumatoid arthritis, osteoarthritis, stroke, cancer, respiratory condition, eye condition, or heart disease: n=86	One or more long-term conditions (in addition to depression): n=96				Long-term condition (self-reported): n=87
	No multi-morbidity	No long-standing illness, disability or infirmity: n=70	None of above: n=15	None: n=73				No long-term condition (self-reported): n=206
	Missing	Missing: n=0	Missing: n=0	Missing: n=0				Missing: n=33
Age (categorical)	<40 years old	n=64	n=0	n=59	n=22	n=96	n=68	n=145
	40-49 years old	n=40	n=0	n=46	n=25	n=69	n=39	n=85
	50-59 years old	n=35	n=0	n=37	n=32	n=39	n=24	n=64
	60-69 years old	n=17	n=43	n=18	n=22	n=12	n=25	n=24

	70 years and over	n=2	n=58	n=9	n=9	n=4	n=3	n=8

Appendix 6.2: REEACT and CIS-R anxiety subscale

Table A6. 2: Summary statistics for CIS-R anxiety subscale in REEACT

Item	Mean score (SD)*	
	REEACT Total (n=685)	REEACT PREDICTR (n=221)
Compulsions	0.72 (1.19)	0.47 (0.97)
Anxiety	2.19 (1.51)	2.01 (1.55)
Irritability	2.16 (1.36)	1.94 (1.38)
Worry	2.46 (1.33)	2.38 (1.38)
Panic	0.74 (1.25)	0.44 (1.03)
Phobias	1.34 (1.28)	1.15 (1.15)
Obsessions	1.26 (1.60)	1.06 (1.49)
Health anxiety	0.88 (1.15)	0.70 (1.03)
Somatic concerns	1.51 (1.44)	1.33 (1.38)
Total	13.27 (6.66)	11.45 5.86)

*For REEACT dataset as a whole (REEACT Total) and for those included in PREDICTR study (i.e. those who have remitted) (REEACT PREDICTR):

Appendix 6.3: Risk of bias (IPD)

Table A6.3: Detailed risk of bias assessment (PROBAST) for sources of IPD

	Study						
	CADET (Richards et al., 2013)	**CASPER Plus** (Bosanquet et al., 2017)	**COBRA** (Richards et al., 2016)	**Healthlines Depression** (Salisbury et al., 2016)	**REEACT** (Gilbody et al., 2015)	**REEACT-2** (Gilbody et al., 2017)	**WYLOW** (Ali et al., 2017)
Domain 1: Participants							
1.1. Appropriate data sources?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1.2. Appropriate inclusions and exclusion?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk of bias	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
Domain 2: Predictors							
2.1. Defined and assessed in similar way for all participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.2. Assessments made without knowledge of outcome?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

2.3. All available at time of model's intended use?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk of bias	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
Domain 3: Outcome							
3.1. Determined appropriately?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.2. Pre-specified or standard definition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.3. Predictors excluded from outcome definition?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.4. Defined and determined similar for all participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.5. Determined without knowledge of predictors?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.6. Appropriate time interval between	Yes	Yes	Yes	Yes	Yes	Yes	Yes

predictor assessment and outcome determination?							
Risk of bias	Low	Low	Low	Low	Low	Low	Low
Overall assessment of risk of bias	Low	Low	Low	Low	Low	Low	Low

Appendix 6.4: Multiple imputation check

Number of previous episodes
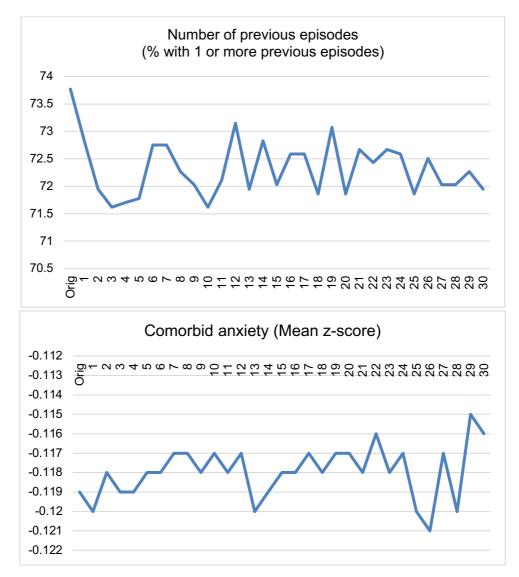(% with 1 or more previous episodes)

Comorbid anxiety (Mean z-score)

Figure A6.1: Summary of original and imputed datasets (1-30) for predictors with missing data*

*Number of previous episodes:

% of participants with 1 or more previous episodes in original dataset = 73.8%

Mean % of participants with 1 more previous episodes (imputed) = 72.3% (SD: 0.51)

Comorbid anxiety:

Mean z-score in original dataset = -0.119

Mean z-score (averaged over 30 imputed datasets) = -0.118 (SD: 0.001)

Appendix 6.5: Predicted probability distributions in each cluster (IECV)

1) CADET

2) CASPER Plus

3) COBRA

4) Healthlines Depression

5) REEACT
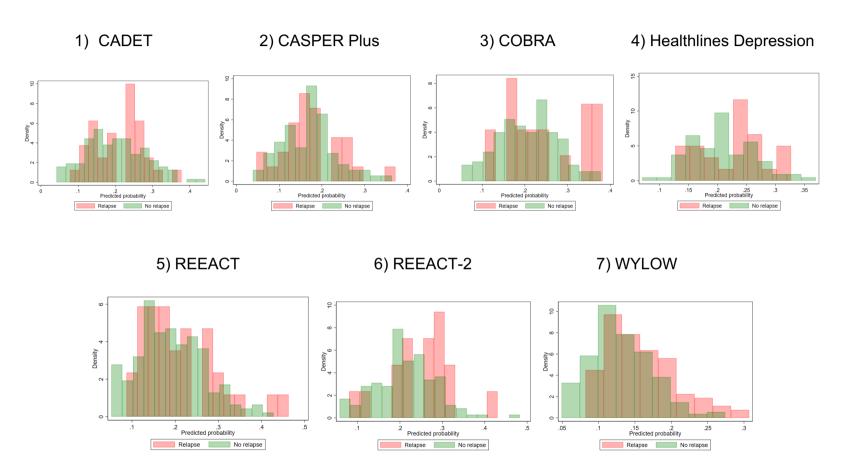
6) REEACT-2

7) WYLOW

Figure A6. 2: Predicted probability distributions in each cluster (IECV)

Appendix 6.6: Sensitivity analysis

This sensitivity analysis excluded REEACT and modelled comorbid anxiety as GAD-7 rather than z-score in the other six studies. All other variables and modelling approach were the same as in the primary analysis.

Modelling of continuous predictors

MFPs were used to model continuous predictors and explore non-linear relationships within the imputed datasets, as for the primary analysis (Table A6.4).

Table A6.4: Transformations and mean-centring of continuous predictors following MFP modelling

Predictor	Transformation and centring
Residual symptoms	X^2-28.9576543 (X = (residual_symptoms+1))
Residual symptoms 2	X^2*ln(X)-48.73333691 (X = (residual_symptoms+1))
Severity	severity-16.21515152
Comorbid anxiety (GAD-7)	comorbid_anx-12.42807717

(Adjusted) results from the multivariable analysis are presented in Table A6.5.

Table A6.5: Results from multilevel multivariable associations (adjusted) between outcome and predictors

Predictor	Beta coefficient (95% CI)	p-value
Number of previous episodes	0.08 (-0.35 to 0.51)	0.715
Residual symptoms	0.15 (0.09 to 0.22)	<0.001
Residual symptoms 2	-0.06 (-0.09 to -0.03	<0.001
Severity	0.10 (0.05 to 0.14)	<0.001
Comorbid anxiety (GAD-7)	-0.04 (-0.08 to 0.00)	0.047
RCT intervention	0.01 (-0.67 to 0.69)	0.979

Intercept (baseline risk): -1.36 (95% CI: -2.06 to -0.67)

Standard deviation of random effect on intercept: 0.45 (95% CI: 0.12 to 1.72)

Standard deviation of random effect on slope (RCT Intervention): 0.54 (95% CI: 0.11 to 2.72)

Correlation between random effects: -0.32 (95% CI: -0.96 to 0.86)

I calculated pooled performance statistics (C-statistic, C-slope and calibration-in-the-large) and also average within-cluster statistics to assess heterogeneity in model apparent performance during model development. 95% Prediction intervals were also calculated.
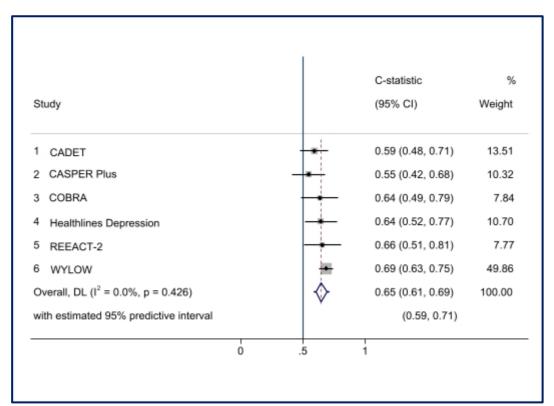
(a)

Figure A6.3: Pooled C-statistic for sensitivity analysis (apparent performance)
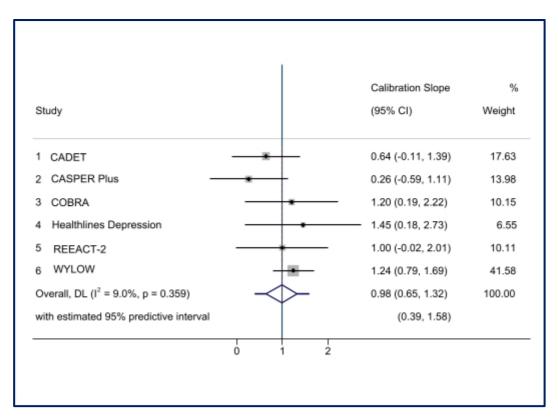

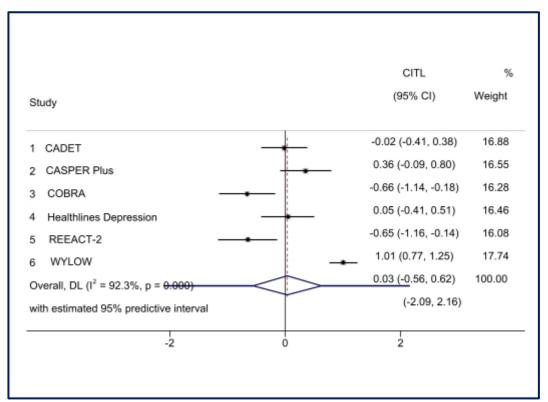Figure A6.4: Pooled calibration slope for sensitivity analysis (apparent performance)

Figure A6.5: Pooled CITL for sensitivity analysis (apparent performance)

Table A6.6: Summary of within-cluster and pooled apparent performance statistics for sensitivity analysis

Study	Number in study	C-statistic (95% CI)	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)
CADET	158	0.59 (0.48 to 0.71)	0.64 (-0.11 to 1.39)	-0.02 (-0.41 to 0.38)
CASPER Plus	101	0.55 (0.42 to 0.68)	0.26 (-0.59 to 1.11)	0.36 (-0.09 to 0.80)
COBRA	169	0.64 (0.49 to 0.79)	1.20 (0.19 to 2.22)	-0.66 (-1.14 to -0.18)
Healthlines Depression	110	0.64 (0.52 to 0.77)	1.45 (0.18 to 2.73)	0.05 (-0.41 to 0.51)
REEACT-2	159	0.66 (0.51 to 0.81)	1.00 (-0.02 to 2.01)	-0.67 (-1.16 to -0.14)
WYLOW	326	0.69 (0.63 to 0.75)	1.24 (0.79 to 1.69)	1.01 (0.77 to 1.25)
Pooled results	1023	0.65 (0.61 to 0.69)	0.98 (0.65 to 1.32)	0.03 (-0.56 to 0.62)

Appendix 6.7: Secondary analyses

Following the univariable analysis undertaken as part of the secondary analysis, the analysis here explores relationship status as a predictor in the model. The purpose of this exploratory analysis was to assess the impact of including relationship status as a predictor within the model, given its statistically significant association with relapse on univariable analysis. The same modelling procedures were followed as for the primary analysis. I retained a multilevel logistic regression model with random intercept to preserve the clustering, but did not include the random slope due to convergence issues with a lower sample size. There were 707 participants in the four clusters with data available for the relationship status variable (CADET, COBRA, REEACT and REEACT-2). I initially developed the model in this data, without relationship status (Part 1), to enable me to quantify the effect of adding relationship status (Part 2) and provide a like-for-like comparison.

Part 1: Model development without relationship status

Again, MFPs and mean-centring were applied for continuous predictors (Table A6.7).

Table A6.7: Transformation and mean-centring of continuous predictors following MFP modelling (secondary analysis without relationship status)

Predictor	Transformation
Residual symptoms	X residual_symptoms-4.451202263
Severity	severity-16.53903819
Comorbid anxiety	comorbid_anx_zscore+.1846512682

Table A6.8: Multivariable associations (adjusted) between outcome and predictors for secondary analysis (without relationship status)

Predictor	Beta coefficient (95% CI)	p-value
Number of previous episodes	-0.09 (-0.61 to 0.43)	0.728
Residual symptoms	0.08 (0.00 to 0.16)	0.057
Severity	0.08 (0.02 to 0.14)	0.008
Comorbid anxiety	-0.14 (-0.39 to 0.12)	0.287
RCT intervention	-0.37 (-0.81 to 0.07)	0.102

Intercept (baseline risk): -1.60 (95% CI: -2.08 to -1.11)

Standard deviation of random effect on intercept: 0.01

I calculated pooled performance statistics (C-statistic, C-slope and calibration-in-the-large) and also average within-cluster statistics to assess heterogeneity in model apparent performance during model development. 95% Prediction intervals were also calculated.
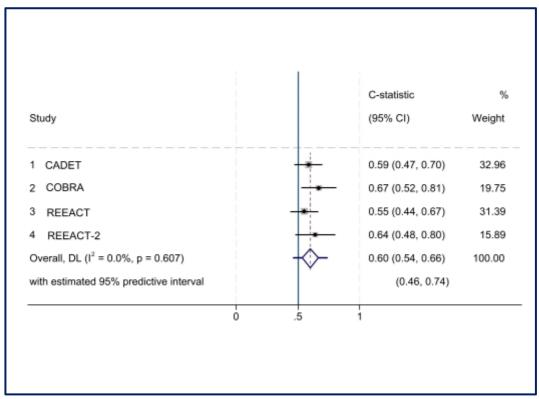
Figure A6.6: Pooled C-statistic (apparent performance) for secondary analysis without relationship status
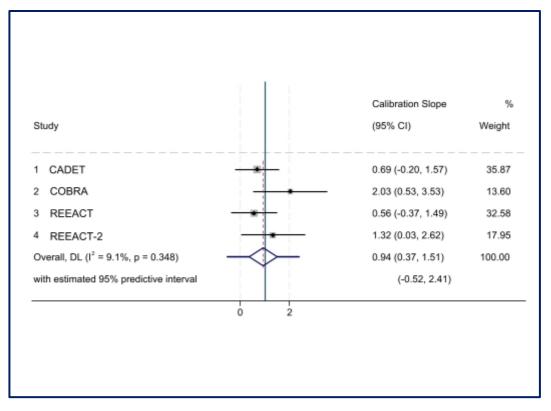

Figure A6.7: Pooled calibration slope (apparent performance) for secondary analysis without relationship status

Figure A6.8: Pooled CITL (apparent performance) for secondary analysis without relationship status

Table A6.9: Summary of within-cluster and pooled (apparent) performance statistics for secondary analysis without relationship status

Study	Number in study	C-statistic (95% CI)	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)
CADET	158	0.59 (0.47 to 0.70)	0.69 (-0.20 to 1.57)	0.36 (-0.04 to 0.75)
COBRA	169	0.67 (0.52 to 0.81)	2.03 (0.53 to 3.53)	-0.10 (-0.58 to 0.38)
REEACT	221	0.55 (0.44 to 0.67)	0.56 (-0.37 to 1.49)	-0.06 (-0.43 to 0.31)
REEACT-2	159	0.64 (0.48 tp 0.80)	1.32 (-0.03 to 2.62)	-0.27 (-0.77 to 0.24)
Pooled results	707	0.60 (0.54 to 0.66)	0.94 (0.37 to 1.51)	0.01 (-0.25 to 0.27)

Part 2: Multilevel logistic regression model with relationship status

Table A6.10: Transformation and mean-centring of continuous predictors following MFP modelling (secondary analysis with relationship status)

Predictor	Transformation
Residual symptoms	residual_symptoms-4.451202263
Severity	severity-16.53903819
Comorbid anxiety	comorbid_anx_zscore+.1855859548

Table A6.11: Multivariable associations (adjusted) between outcome and predictors for secondary analysis (with relationship status)

Predictor	Coefficient (95% CI)	P value
Number of previous episodes	-0.15 (-0.66 to 0.37)	0.582
Residual symptoms	0.07 (-0.01 0.15)	0.081
Severity	0.07 (0.01 to 0.13)	0.020
Comorbid anxiety	-0.12 (-0.37 to 0.14)	0.363
Relationship status	-0.79 (-1.23 to -0.34)	0.001
RCT intervention	-0.40 (-0.84 to 0.04)	0.076

Intercept (baseline risk): -1.11 (95% CI: -1.65 to -0.56)

Standard deviation of random effect on intercept: 2.00e—09 (SE = 155.69)

I calculated pooled performance statistics (C-statistic, C-slope and calibration-in-the-large) and also average within-cluster statistics to assess heterogeneity in model apparent performance during model development. 95% Prediction intervals were also calculated.
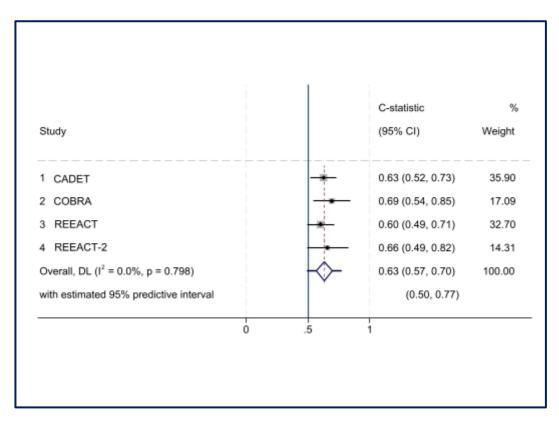
Figure A6.9: Pooled C-statistic (apparent performance) for secondary analysis with relationship status

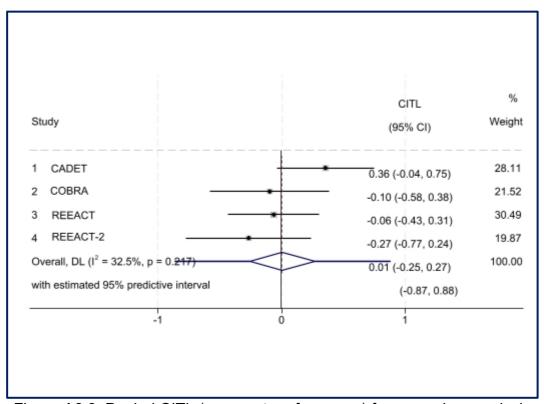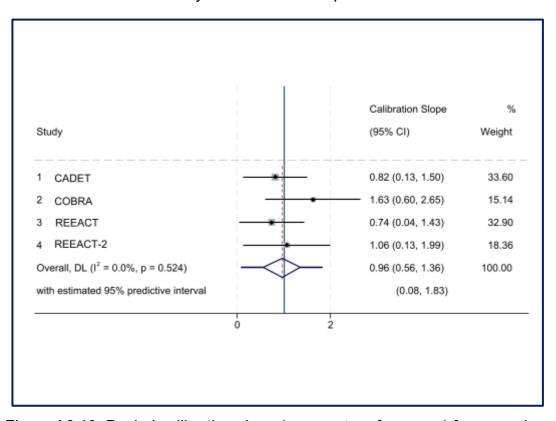Figure A6.10: Pooled calibration slope (apparent performance) for secondary analysis with relationship status
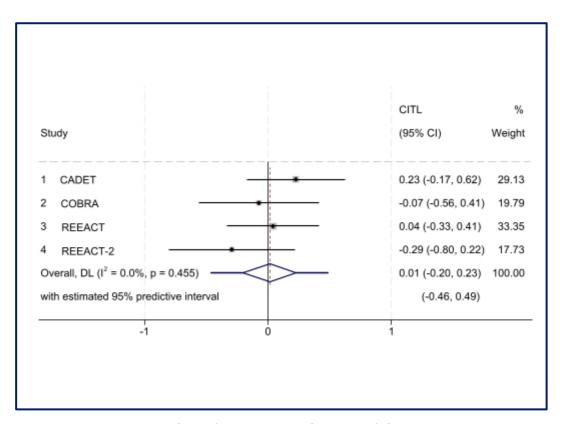
Figure A6.11: Pooled CITL (apparent performance) for secondary analysis with relationship status

Table A6.12: Summary of within-cluster and pooled (apparent) performance statistics for secondary analysis with relationship status

Study	Number in study	C-statistic (95% CI)	Calibration slope (95% CI)	Calibration-in-the-large (95% CI)
CADET	158	0.63 (0.52-0.73)	0.82 (0.13 - 1.51)	0.23 (-0.17 - 0.63)
COBRA	169	0.70 (0.54-0.85)	1.64 (0.61 – 2.67)	-0.07 (-0.56 to 0.41)
REEACT	221	0.60 (0.49 – 0.70)	0.72 (0.02 - 1.41)	0.04 (-0.33 to 0.41)
REEACT-2	159	0.66 (0.49 - 0.82)	1.08 (0.14 - 2.01)	-0.29 (-0.80 to -0.22)
Pooled results	707	0.63 (0.57-0.70)	0.96 (0.56 - 1.36)	0.01 (-0.20 to 0.23)

Appendix 7.1: Health Research Authority Approval Letter

Dr Andrew Moriarty

NIHR Doctoral Research Fellow

University of York
Department of Health Sciences and the Hull York
Medical School
Faculty of Science
University of York
YO10 5DD

11 March 2022

Dear Dr Moriarty

Email: approvals@hra.nhs.uk

HRA and Health and Care

Study title:	**Predicting and preventing relapse of depression in primary care**
IRAS project ID:	**292780**
Protocol number:	**DRF-2018-11-ST2-044**
REC reference:	**22/WM/0022**
Sponsor	**University of York**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?

HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?

HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:
- Registration of research
- Notifying amendments
- Notifying the end of the study

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

Who should I contact for further information?

Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **292780**. Please quote this on all correspondence.

Yours sincerely,

Helen Penistone

Approvals Manager

Email: approvals@hra.nhs.uk

Copy to: Dr Michael Barber

List of Documents

The final document set assessed and approved by HRA and HCRW
Approval is listed below.

Document	*Version*	*Date*
Copies of materials calling attention of potential participants to the research [Appendix A Poster for people with lived experience of depression]	1	23 November 2021
Copies of materials calling attention of potential participants to the research [Appendix B Poster for GPs]	1	23 November 2021
Copies of materials calling attention of potential participants to the research [Appendix O Receipt (lay participants)]	1	10 February 2022
Copies of materials calling attention of potential participants to the research [Appendix P Receipt (GPs)]	1	10 March 2022
Copies of materials calling attention of potential participants to the research [Appendix N Social media adverts]	2	10 February 2022
Covering letter on headed paper [Thank you Letter]	1	23 November 2021
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Sponsor indemnity]		21 December 2021
GP/consultant information sheets or letters [Appendix F PIL for GPs]	3	06 February 2022
GP/consultant information sheets or letters [Invitation letter for GPs]	1	23 November 2021
Interview schedules or topic guides for participants [Topic Guide for GPs]	1	23 November 2021
Interview schedules or topic guides for participants [Topic Guide for people with lived experience]	1	23 November 2021
IRAS Application Form [IRAS_Form_21122021]		21 December 2021
Letter from funder [Funder letter]		21 December 2021
Letter from sponsor [Sponsor Letter]		21 December 2021
Letters of invitation to participant [Participant Invitation Letter]	1	23 November 2021
Non-NHS/HSC Site Assessment Form [Risk protocol]	1	23 November 2021
Organisation Information Document [PIC agreement]	1	21 December 2021
Other [Response to HRA REC March 2022]	1	03 March 2022
Participant consent form [Appendix G: Consent Form]	3	06 March 2022
Participant consent form [Appendix H Consent Form GPs]	3	06 February 2022
Participant information sheet (PIS) [Appendix D: PIL]	3	06 February 2022
Research protocol or project proposal [Protocol version 3]	3	05 February 2022
Summary CV for Chief Investigator (CI) [A Moriarty CV]		21 December 2021

Depression

is the leading cause of disability in the world

If you have experience of depression, you could help improve the outcome of treatments

Why?

At least half of people with depression will become unwell again (relapse) after they have finished their initial treatment.

What?

We would like to understand more about the views and experiences of people with lived experience of depression and about the care and support you have received.

How?

If you are aged 18 years and over and have lived experience of depression, we would like you to get involved and share your experiences with us

A single interview will last one hour at the most and can be in-person or by telephone/online. You will be reimbursed for your time.

How can I participate?

If you have any questions or would like to get involved, please contact Dr Andrew Moriarty at andrew.moriarty@york.ac.uk.

Express your interest now

Version 1; 23rd November 2021

Depression in primary care:
Predicting and preventing relapse

Are you a General Practitioner with experience seeing patients with depression?

What is the problem?

At least half of people with depression will become unwell again (relapse) after they have finished their initial treatment.

What is the research project?

We are developing a tool to guide relapse risk prediction and prevention for GPs to use. We would like to learn more about how risk is currently assessed and discussed with patients. We would like to explore how a risk assessment tool could be used in practice.

What will it involve?

We would like to seek your views in a one-off interview. Your time will be reimbursed.

The interview will last an hour at the most and can be in-person or by telephone/online.

How can I participate?

If you have any questions or would like to get involved, please contact Dr Andrew Moriarty at andrew.moriarty@york.ac.uk.

Version 1; 23rd November 2021

Appendix 7.4: Invitation to participate letter for people with lived experience

[PRACTICE HEADED PAPER]

Dear XXX,

We are writing to you with an opportunity to take part in a one-off interview as part of a research project about depression. A team of researchers from the University of York and Keele University are looking to improve care and support for people with depression in primary care.

Please find enclosed an information sheet with more detail.

If this sounds interesting to you and you would like to be involved or to know more, then please contact the research team by emailing: andrew.moriarty@york.ac.uk. If you don't think this is relevant to you, please ignore this letter - you do not need to do anything more.

Yours sincerely,

[NAME OF GP]

Predicting and preventing relapse of depression in primary care

Information for people with lived experience of depression considering taking part in the study

We would like to invite you to take part in a research study that is looking at relapse of depression in primary care. Before you decide to take part, it is important that you understand why this research is being done and what it would involve for you.

Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part in the study.

The research is sponsored by the University of York and conducted by researchers from the University's department of Health Sciences, in collaboration with the Hull York Medical School and Keele University. It is funded by the National Institute for Health Research. The study is led by Dr Andrew Moriarty, a GP in York and researcher at the University of York and Hull York Medical School.

Part 1 tells you the purpose of this study and what will happen if you take part.
Part 2 gives you more detailed information about the conduct of the study.
Please ask us if there is anything that is not clear.
Thank you for reading this.

PART ONE
What is the purpose of the study?

There is insufficient evidence about how to support people with depression to stay well. After initial improvement, some people may become unwell again ("relapse"). The aim of our research is to develop a way of identifying patients who are at risk of relapse. General Practitioners (GPs) would like to learn about your experiences to understand how to predict and prevent relapse.

Why am I being asked to take part?

If you are aged 18 years and older, with lived experience of depression, we would like to talk to you about your treatment, care and support (this could be from your GP or elsewhere). We would also like to explore your understanding of relapse risk and prevention and how you think we might improve this for patients in the future.

Prior to speaking to the researcher, it would be helpful to think about:
- Your experiences with depression;
- The treatments and support you have received;
- Whether anybody has discussed risk of relapse with you in the past;
- Whether you have received relapse prevention treatments and how think GPs could help you to make decisions around these.

This will involve a one-off discussion with a researcher, lasting for up to an hour. This could be in-person, on the telephone or using an online virtual platform, according to your preference. All of the information shared with us will be treated confidentially and will be used in our effort to improve care for patients with depression in primary care. You will be reimbursed for your time at a rate of £25 for one interview, lasting an hour.

Do I have to take part?
It is up to you to decide whether or not to join the study. Your decision will **NOT** affect your care in any way. It will not, for example, impact on any decisions made by care professionals.

You may wish to discuss the study further with the research team (contact details are provided at the end of this information sheet). You would be free to withdraw from the study at any time, without giving a reason.

If I wish to take part, what do I need to do?
If you would like to take part you must complete a consent form. There is nothing further you would need to do.

What will happen to me if I take part?

The study involves a member of the research team talking to you about your experience. This will last about 60 minutes and is likely to be online. You can decide when the researcher talks to you. You can end the interview at any time (see *Part 2*).

With your permission we would like to record the interview. This recording will be destroyed once the interview is transcribed. This transcription will be kept safe and secure. It will not be shared with those involved in your care or with anyone outside the research team (see *Part 2*).

Your interview, along with others we have spoken to, will inform our analysis and write-up. You will be given a unique study number for the duration of the study so that your name will not be used in any publications. Nor will it be made available outside the research team.

What are the possible disadvantages or risks of taking part?

Due to the nature of this study, it is possible that you will find the discussion upsetting. In designing this research, we have consulted with advisors, including people who have experienced depression. This should help ensure interviews are undertaken in a sensitive manner.

In the event of any distress, you may pause or stop the interview at any time and are not under any pressure to answer any questions you do not want to. If necessary, we will direct you to resources for support. We can also arrange for a care professional to talk to you.

What are the possible benefits of taking part?

Some people find it helpful to talk about their experiences. Aside from this, you probably won't get any direct benefits from taking part in this study. However, we plan to publish the results of our research and it may help people understand more about the experiences of depression.

Who has reviewed this study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed by the Health Research Authority (IRAS 292780).

Will my taking part in the study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in *Part 2*. This completes *part 1*.

If the information in *Part 1* has interested you and you are considering taking part, please read the additional information in *Part 2* before making any decision.

PART TWO

How will we use information about you?

We will use information from our interview with you for this research project. Your transcribed interview will have a code and will not contain any contact details.

We will also have access to your name and contact details, although this information will be kept separate from your interview transcript.

Your contact details would be stored electronically on a secure server and only authorised individuals at the University of York will have access to it.

People who do not need to know who you are, will not be able to see your name or contact details. Nor will they have access to your interview transcript. We will keep all information about you safe and secure.

Once we have finished the study, we will keep some of the data for up to ten years. This will enable us to complete our publications and reports. We will write our reports in a way that no-one can work out that you took part in the study.

What are your choices about how your information is used?

The University of York is a publicly funded organisation that conducts research in the public interest to improve health and healthcare services. The ability to change this material, however, is limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have.

You can find details in the links to further information below how the use of your data in this research follows the data protection laws in the UK.

- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/

- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/research-partcipants/

- by emailing the University of York's data protection officer on dataprotection@york.ac.uk

Will my taking part in this study be kept confidential?

If you decide to take part in the study, what you tell us will be kept confidential and not shared with those involved in your care and treatment. No one outside the research team will know that you have taken part in the study. We will keep all information that we have about you safe and secure.

We will write our reports in a way that no-one can work out that you took part in the study.

Data collected for the study may be looked at by authorised persons who are organising the research. Data may also be looked at by other authorised

people to check that the study is being carried out correctly. All have a duty of confidentiality to you as a research participant.

The only time we would break our duty of confidentiality is if we are worried that you – or someone else – was being, or was likely to be, harmed. If that happens, we will talk with you about it.

What will happen to the results of the research study?
Researchers from the University of York will analyse the material collected.

The results of the study will be published in academic journals.

Anonymised study data may be reused by the research team or researchers in other institutions for secondary research purposes and in future research.

What if there is a problem?
Complaints
If you have a concern about any aspect of this study, you should ask to speak to the research team who will do their best to answer your questions (contact details below). You can also contact the Chief Investigator Dr Andrew Moriarty. He will be happy to discuss your concerns. If you are unhappy with Dr Moriarty's response, you can contact Professor Patrick Doherty, who is Head of Department (Health Sciences).

If you remain unhappy following this and you wish to complain formally, you can do this by contacting the Parliamentary and Health Service Ombudsman, who is independent of the NHS and government, at 0345 015 4033.

Data Protection
If you are unhappy with the way your personal data has been handled, please contact the University's Data Protection Officer at dataprotection@york.ac.uk. If you are not satisfied with our response, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see www.ico.org.uk/concerns.

Dr Andrew Moriarty can be contacted at andrew.moriarty@york.ac.uk.

Thank you for reading this information sheet and for considering whether to take part in this study.

Dear Colleague,

We are writing to ask for your help with a research study. We are hoping to learn from your experience and knowledge to improve care and support for people with depression in primary care. This research project is run by a team of researchers from the University of York and Keele University.

Please find enclosed an information sheet with more detail.

If this sounds interesting to you and you would like to be involved or to know more, then please contact the research team by emailing: andrew.moriarty@york.ac.uk. If you don't think this is relevant to you, please ignore this letter - you do not need to do anything more.

Yours sincerely,

Dr Andrew Moriarty

Predicting and preventing relapse of depression in primary care

Information for general practitioners considering taking part in the study

We would like to invite you to take part in a research study that is looking at relapse of depression in primary care. Before you decide to take part, it is important that you understand why this research is being done and what it would involve for you.

Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part in the study.

The research is sponsored by the University of York and conducted by researchers from the University's department of Health Sciences, in collaboration with the Hull York Medical School and Keele University. It is funded by the National Institute for Health Research. The study is led by Dr Andrew Moriarty, a GP in York and researcher at the University of York and Hull York Medical School.

Part 1 tells you the purpose of this study and what will happen if you take part.
Part 2 gives you more detailed information about the conduct of the study. Please ask us if there is anything that is not clear.
Thank you for reading this.

PART ONE
What is the purpose of the study?

People with depression are at high risk of having a relapse or recurrence after entering remission or recovery.

We are developing a tool to help GPs to stratify patients according to risk of relapse. We are interested in the perspectives of GPs on how they currently assess, communicate and manage relapse risk with patients.

We would then like to receive GP input on the tool as it currently is and how we can best look to implement it in primary care. We are aiming to make the tool as easy to use as possible, including only routinely collected information.

Why am I being asked to take part?

We would like to talk to GPs who have experience in the care of people with depression. Interviews will take no longer than 30 minutes and can be in-person or remote, and all information will be treated confidentially.

GPs will be reimbursed at a rate of £44.00 for half an hour or £88.00 for an hour of their time, in the form of a voucher. Conversations can be in-person, on the telephone or using a remote platform, according to your preference. We will be flexible and can work around your clinical commitments. All of the information shared with us will be treated confidentially and will be used in our effort to improve care for patients with depression in primary care.

Do I have to take part?
It is up to you to decide whether or not to join the study.

You may wish to discuss the study further with the research team (contact details are provided at the end of this information sheet). You would be free to withdraw from the study at any time, without giving a reason.

If I wish to take part, what do I need to do?
If you would like to take part you must complete a consent form. There is nothing further you would need to do.

What will happen to me if I take part?
The study involves a member of the research team talking to you about your experience. This will last about 30 minutes and is likely to be online. You can decide when the researcher talks to you. You can end the interview at any time (see *Part 2*).
With your permission we would like to record the interview. This recording will be destroyed once the interview is transcribed. This transcription will be kept safe and secure. It will not be shared with those involved in your care or with anyone outside the research team (see *Part 2*).

Your interview, along with others we have spoken to, will inform our analysis and write-up. You will be given a unique study number for the duration of the study so that your name will not be used in any publications. Nor will it be made available outside the research team.

What are the possible benefits of taking part?
We hope that your input will contribute towards tools that GPs will find helpful and easy to use in practice. We plan to publish the results of our research and it may help people understand more about the experiences of depression.

Who has reviewed this study?
All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed by the Health Research Authority (IRAS 292780).

Will my taking part in the study be kept confidential?
Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in *Part 2*. This completes *part 1*.

If the information in *Part 1* has interested you and you are considering taking part, please read the additional information in ***Part 2*** before making any decision.

PART TWO
How will we use information about you?
We will use information from our interview with you for this research project. Your transcribed interview will have a code and will not contain any contact details.

We will also have access to your name and contact details, although this information will be kept separate from your interview transcript.

Your contact details would be stored electronically on a secure server and only authorised individuals at the University of York will have access to it.
People who do not need to know who you are, will not be able to see your name or contact details. Nor will they have access to your interview transcript. We will keep all information about you safe and secure.

Once we have finished the study, we will keep some of the data for up to ten years. This will enable us to complete our publications and reports. We will write our reports in a way that no-one can work out that you took part in the study.

What are your choices about how your information is used?
The University of York is a publicly funded organisation that conducts research in the public interest to improve health and healthcare services. The ability to change this material, however, is limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate.
You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have.

You can find details in the links to further information below how the use of your data in this research follows the data protection laws in the UK.

- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/

- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/research-partcipants/

- by emailing the University of York's data protection officer on dataprotection@york.ac.uk

Will my taking part in this study be kept confidential?
If you decide to take part in the study, what you tell us will be kept confidential and not shared with those involved in your care and treatment. No one outside the research team will know that you have taken part in the study. We will keep all information that we have about you safe and secure.

We will write our reports in a way that no-one can work out that you took part in the study.
Data collected for the study may be looked at by authorised persons who are organising the research. Data may also be looked at by other authorised people to check that the study is being carried out correctly. All have a duty of confidentiality to you as a research participant.

The only time we would break our duty of confidentiality is if we are worried that you – or someone else – was being, or was likely to be, harmed. If that happens, we will talk with you about it.

What will happen to the results of the research study?
Researchers from the University of York will analyse the material collected.

The results of the study will be published in academic journals.

Anonymised study data may be reused by the research team or researchers in other institutions for secondary research purposes and in future research.

What if there is a problem?
Complaints

If you have a concern about any aspect of this study, you should ask to speak to the research team who will do their best to answer your questions (contact details below). You can also contact the Chief Investigator Dr Andrew Moriarty. He will be happy to discuss your concerns. If you are unhappy with Dr Moriarty's response, you can contact Professor Patrick Doherty, who is Head of Department (Health Sciences).

If you remain unhappy following this and you wish to complain formally, you can do this by contacting the Parliamentary and Health Service Ombudsman, who is independent of the NHS and government, at 0345 015 4033.

Data Protection
If you are unhappy with the way your personal data has been handled, please contact the University's Data Protection Officer at dataprotection@york.ac.uk. If you are not satisfied with our response, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see www.ico.org.uk/concerns.

Dr Andrew Moriarty can be contacted at andrew.moriarty@york.ac.uk.

Thank you for reading this information sheet and for considering whether to take part in this study.

Predicting and preventing relapse of depression in primary care

Please initial the box next to each consent statement if in agreement

1. I confirm I have read the Participant Information Sheet. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.

3. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers.

4. I understand that my interview will be digitally recorded and transcribed (written out). I consent for the research team to use this anonymized material, possibly including word-for-word quotes.

5. I give permission for my interview to be stored securely on a password-protected computer, in line with the regulations of the Data Protection Act (2018).

☐

6. I agree that I may be contacted and invited to participate in other research studies that may follow on from the findings of this study (this is optional).

☐

7. I agree to take part in the above study.

☐

Participant:

Name: _____

Signature: _____

Date: _____

Individual taking informed consent:

Name: _____

Signature: _____

Date: _____

Predicting and preventing relapse of depression in primary care

Please initial the box next to each consent statement if in agreement

1. I confirm I have read the Participant Information Sheet. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.

3. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers.

4. I understand that my interview will be digitally recorded and transcribed (written out). I consent for the research team to use this anonymized material, possibly including word-for-word quotes.

5. I give permission for my interview to be stored securely on a password-protected computer, in line with the regulations of the Data Protection Act (2018).

6. I agree that I may be contacted and invited to participate in other research studies that may follow on from the findings of this study (this is optional).

7. I agree to take part in the above study.

Participant:

Name: _____

Signature: _____

Date: _____

Individual taking informed consent:

Name: _____

Signature: _____

Date: _____

Appendix 7.10: Summary of PAG meetings

Date of meeting	Attendees	Summary of meeting
July 2017	AM, EW, SP, PM, GP, JC (In-person at University of York)	Initial meeting to discuss proposed NIHR funding application. Meeting funded by grant from NIHR Research Design Service Public Involvement Fund. Identified problem of relapse and mapped initial ideas for workstreams. The group reported this is an important area and they all reported experiences of feeling "forgotten" or "abandoned" after initial episode. In particular, the group thought it was important that patient perspectives were explored through qualitative work. The group were all very keen to remain involved throughout, if funding were awarded. The group agreed to read and comment on application.
28th October 2019	AM, HG, SP, PM, EW, JG, GP (In-person at University of York)	Discussed overall plans for project. consensus was that the phrases "preventing relapse" or "reducing relapse" were appropriate - felt that it was important to aim for this positive outcome. The group agreed to read and comment on a lay summary for Cochrane Prognosis review. Reviewed literature around relapse predictors and began to select candidate predictors for model. Early discussions around focus of qualitative work. Agreed future meetings.
27th October 2020	AM, JC, EW, SP, GB, GP (Remote online, Zoom)	Discussed relapse and the group highlighted that it would be important to explore preferences

		around terminology and how to communicate risk of relapse in qual work.
		We approved the plain language summary for Cochrane review.
		Agreed "PREDICTR" acronym for prognostic model study, after considering different options and discussing as group.
		Determined new categories for harmonising IPD. Discussed qual work – agreed need for PAG to collaborate on topic guide and for one or two "pilot" interviews. Discussed member checking and AM agreed to explore pros/cons. Highlighted need for risk protocol.
1st June 2021	AM, JC, EW, GP, SP (Remote online, Zoom)	Discussion mainly aimed at qualitative materials with a view to IRAS application. Agreed on phrase "people with lived experience of depression" for lay participants. Agreed mainly recruit via GP search but agreed would be acceptable to use study advertisements for some participants as well.
7th September 2021	Present: AM, JC, EW, GP (Remote online, Zoom)	Continuation of collaborative work on participant-facing materials for qualitative study. Worked on and agreed first draft of topic guide for interviews with people with lived experience of depression.
14th September 2021	Present: AM, EW (Remote online, Zoom)	Pilot interview No. 1. Modified topic guide as a result.
28th September 2021	Present: AM, JC (Remote online, Zoom)	Pilot interview No. 2. Modified topic guide as a result.
12th September 2022	Present: AM, EW, GP, JC (Remote online, Zoom)	Initial review of qualitative data and initial PAG input into thematic analysis.

		Of particular note – life events/circumstance driving depression, "is depression an illness?", GP-patient relationship.
24th April 2023	AM, EW, GP, JC, PM (Remote online, Zoom)	Determined final themes and theme names for qualitative research. "Perceived determinants of depression course" – there was a strong steer from the group to separating this into external and internal factors. Remove "demonstrating" empathy- importance is on empathy rather than the demonstrating. "General practice at the centre of care" became "Patients at the centre of care within general practice". The PAG expressed the view that the process had been collaborative and themes were more person-centred as a result.

Appendix 7.11: Topic Guide for people with lived experience of depression

(Version 1.0, 23rd November 2021)

Introduction and consent

Introduction and consent

Demographics: age, post code, gender identity, ethnicity, level of education
reached, (previous) employment, rural/urban/inner city, who is at home with you,
other long-term conditions
GP details (in case risk protocol needs to be activated)
Please can you tell me why you decided to take part in this study?

1. Please can you tell me about your depression?

    Possible prompts:
    Are you currently experiencing depression?

    When do you think you had your first experience of depression?

    How many times have you experienced depression? How often?

    Do you have any thoughts on what is the cause/trigger for your depression?

2. Can you tell me who is/has been involved in helping you with your
    depression?

    Prompts:
    can you tell me what sorts of treatment you have had?

    How do you feel about your depression treatment, care and support?

    Is there anything you'd like to try or to have been offered - and what are the
barriers to this?

3. Can I ask, were you aware or advised about the risk of having depression again after your previous experience(s) of depression?

    Can you tell me about any discussions you had about this with your GP or anybody else?
    Would this have been useful? Why/why not?

    If you could be given more specific, individual information about your
personal risk of becoming unwell again by your GP, how would you feel about that?

        If available, how might this kind of information be delivered to you?
    Can you explain why?
    e.g. by a healthcare professional, online (app, website)

    In terms of having risk explained to you, what makes most sense to you? For example, some people talk in terms of percentage (60%), risk proportion (i.e. 6 in 10), or categories such as high, medium, or low risk.
    What words should be used or not used? (e.g. high risk, significant risk)

[Read explanation of PREDICTR study]**

Are there any important things that you think are missing from this? Why do you think they are important?

4. Would you have found it helpful to be offered support to prevent relapse after your depression improved?
Prompts:
    What kind of support you would prefer?
        Examples: CBT, IPT, MBCT, medication
What sort of help do you think would help you decide the kind of support?

WOULD YOU LIKE THIS IN PRIMARY CARE??

5. Anything else you would like to add?

Close interview: Thank you, re-check consent, arrange reimbursement.

**(Example explanation of the PREDICTR tool): We are developing a tool to help GPs to assess a person's risk of relapse/recurrence after having depression. We have included the following information to help to make this assessment:

- whether a person has had depression before;
- whether they still have some symptoms of depression when they are feeling better;
- whether they have also experienced anxiety as well as depression;
- how severe their episode of depression was.

These were chosen after looking at the evidence and research literature. However, we understand that there may be factors that are important to people that we have not included.

Appendix 7.12: Topic Guide for GPs

(Version 2.0, 19th May 2022)

Introduction and consent

Demographics: type of practice, urban/rural, training/teaching

AGE

GP – years of experience as a GP, Salaried/locum/Partner, ethnicity, gender

Does the GP have any other roles alongside clinical commitments

Does GP have any specialist interest/expertise in MH?
Why did you agree to be interviewed?

1. Can you reflect on your experiences of managing people with depression
    a. diagnosis
        i. Do you/would you routinely do PHQ-9/GAD-7? How do you use these instruments in practice?
    b. management options

    c. Monitoring (and explore their behaviour compared to colleagues in their practice)

    d. PT comment re unprompted follow up
    e. defining remission/recovery
    f. monitoring/review after remission/ recovery
    g. OTHER STAFF???
    h. What would help from a resource point of view??? Premises?? ARRS??
2. COVID

3. Can you tell me about if and how you assess risk of relapse in patients with depression?

4. Please can you tell me about a time that you had a discussion with a patient about relapse?

    a. What advice do you give? When?
    b. Do you advice a person about staying well and to self-monitor mood (eg using PHQ9)

5. Do you feel confident in being able to assess a patient's risk of relapse/recurrence once they have entered remission/recovery?

What sources of information or guidelines have you used/would you use to help you to assess this risk?
Do you think you can communicate risk to patients? Why? Why not?

6. What is your understanding of the type/availability of relapse prevention interventions for patients?
7. AWARE OF BOOSTER SESSIONS?? RELAPSE PREVENTION??
8. WHAT IS INTERFACE LIKE WITH IAPT / SECONDARY CARE???
is this primary care or need additional funding???

(Example explanation of the PREDICTR tool): We are developing a tool to help GPs to assess a patient's risk of relapse/recurrence after having depression. We have included the following information to help to make this assessment: previous depression (patient report/GP record); residual depressive symptoms (PHQ-9 at remission); whether they have also experience anxiety as well as depression (GAD-7); how severe their episode of depression was (PHQ-9 at baseline). These were chosen after looking at the evidence and research literature.

9. Are there any factors that you think should be included in this tool that we have not mentioned?

10. how you might use such a tool in practice? Do you think your colleagues would

What would be the best way of implementing this tool? (e.g. IT, paper, internet, patient self-complete)

Would you be prepared to use this tool if there was evidence for the effectiveness of such an approach?

What might be the barriers to using such a tool?

Would you have confidence in using this tool?

8. How do you think risk is best communicated to GPs and patients?

       e.g. percentage risk, proportion, categories (high, medium, low risk)

Prognostic models:

- Do they want predicted risks, or a 'high' or 'low/normal' output
- If and how do they use and communicate predicted risks in practice
- How do they define risk thresholds for guiding treatment choices, and what if someone is just above or just below the threshold
- How do they implement models when some information (Eg smoking, BMI) is missing?
- Do they use computer, or a chart, or nomogram etc for implementation?
- Can patient's handle information from prediction models? How do they facilitate discussions of risk?
- What, if any, models do they use
- What makes them trust some models and not others

Is there anything you would like to ask or suggest?

Close interview: Thank you, re-check consent, arrange reimbursement.

Dear XXX,

We would like to thank you very much for your time and for helping us to improve care for people with depression.

This was very much appreciated by all of us on the research team. Please let us know if you'd like to be informed when the research is published.

Yours sincerely,

Dr Andrew Moriarty
On behalf of the research team

Risk protocol

This qualitative study will use interviews to explore the experiences of people with lived experience of depression. It is possible that participants may disclose intent of self-harm or suicide in the interview. This risk protocol outlines steps that will be taken if this arises. The study will also involve interviews with General Practitioners (GPs). It is unlikely that the risk protocol will be needed in these interviews.

General statement:

The participant's GP is responsible for the ongoing clinical care of participants. The research team therefore has a duty of care to ensure that the GP is aware of any suicidal ideation expressed by participants.

At the beginning of the interview the researcher will confirm the name of participant's registered GP surgery and explain why this may be needed, should the researcher have concerns about the participant, as outlined below. The researcher will initiate the risk protocol if a participant expresses suicidal or self-harm intent.

Definition of self-harm intent:

Self-harm intent is defined as any expression from the study participant to the researcher stating he/she is planning to self-harm. Self-harm is defined as an intentional act of self-poisoning or self- injury irrespective of the type of motivation or degree of suicidal intent (RCPsych, 2010).

Definition of suicidal intent:

Suicidal intent is defined as the serious wish from the study participant to the researcher stating he/she is planning to take their own life.

Action required if self-harm or suicidal intent is expressed

Should the interview participant disclose self-harm or suicidal intent, the researcher will first ask if the participant has spoken about the situation with his/her GP. The researcher will reiterate the importance of discussing this with their GP and urge the participant to contact his/her GP. Suggested scripts are included later in the protocol.

If the participant declines to share their suicidal or self-harm intent with their

GP, the researcher will provide advice and support at the time of the interview. Andrew Moriarty (AM), GP and clinical academic (with clinical experience managing people who are distressed) is study lead and will be conducting the majority, or all, interviews. If a second opinion or additional support is required, ASM will discuss with CCG (Professor of Primary Care).

If the researcher feels it is appropriate or necessary, (s)he will contact AM whilst she is still with the study participant (face to face) or on the telephone to the participant, using a separate telephone, or as soon as possible after the interview. The researcher may also contact the participant's GP or local Crisis Team following discussion with AM.

Suggested researcher scripts:

I am concerned about some of the things you have disclosed to me and the risk of harm towards yourself. Have you talked about them with your doctor or anyone else? It is important that your doctor is aware about how you are feeling so that he/she can make sure there is the necessary support in place for you. Could you please share your thoughts with your doctor as soon as possible so they can support you? Would you like me to help you make a GP appointment for today? Will you be able to travel to your GP surgery?

a) If the participant is hesitant or declines:

It is common for people to find it hard to talk about these feelings during a visit to their family doctor, but your GP can help you with these feelings. If he/she learns of how you are feeling, he/she will be able to discuss it with you, and decide the best treatment and support for you.

b) If the participant continues to decline:

I understand you don't wish to share these thoughts with your family doctor. Would you be willing to talk to me about how you are feeling? It may help you. I will however have to let my colleague know, who is a GP, and she may be in contact with you over the next days. I will now contact Andrew Moriarty to discuss further.

References

ROYAL COLLEGE OF PSYCHIATRISTS. 2010. Self-harm, suicide and risk: helping people who self-harm. [Online] Available from: https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh- policy/college-reports/college-report-cr158.pdf?sfvrsn=fcf95b93_2 [Accessed 7th September 2019]

Appendix 7.15: Text for social media advertisement for GP participants

For GPs (Twitter):

We are looking for GPs to contribute to a research project to help improve

care for people with depression. Can you spare 30 minutes of your time?

Find out more about the research project here:

Predicting and preventing relapse of depression in primary care

I…………………………………………………... (name in full)
confirm that I participated in an interview for this study and have
received a voucher to the value of £25 as reimbursement for my
time [Voucher code…………]

Signed…………………………………………. Date……………

Confirmed by: ……………………………   Date: ……………

Predicting and preventing relapse of depression in primary care

I…………………………………………………... (name in full) confirm that I participated in an interview for this study and have received a voucher to the value of £44.00 / £88.00 [delete as appropriate] as reimbursement for my time.

[Voucher code: ……………….]

Signed………………………………………… Date……………

Confirmed by: ……………………………   Date: ………

Appendix 7.18: Example of mind map analysis

Making diagnosis and assessing progress

GAD-7

GP records, coding and technology

GP

GPs - other skills and experience

GPs – ways of working

PHQ-9 – benefits, downsides and how it is used

Conceptualization of depression

Face to face consultations vs others

Covid and depression in primary care

Primary care and depression

Non-GP health professionals

Remote consulting

Antidepressant medication in primary care

Life stress and pre-disposing factors

Causes of depression

Life stress and depression

Work

History of depression

Relationships

Childhood adversity

Barriers to seeking primary care

Patient perspectives and expectations

Patient

Previous experiences with healthcare

Barriers and facilitators to access GP

Managing own depression

Experiences of depression and diagnosis

Risk and protective factors for relapse

Factors associated with increased relapse risk

Factors considered protective against relapse

Experiences of psychological therapy

Treatments

Self management and care

Self referral

Ways of thinking about 5 R's

Terminology – 5 R's

Conceptualization of relapse

Significance of relapse

Relapse risk

Relapse prevention

Discussion of relapse

Relapse risk prediction

Relapse risk prediction model

Relapse risk and prevention

Triggers and Early Warning Signs

Barriers to primary care based relapse prevention

Experiences of GP and primary care

GP-patient relationship

Continuity of care

Holistic care

Continuity, listening and empathy

Shared decision making and patient choice

Listening and empathy

Importance of individualized care

Relationships and communication

Limited GP time and resource

Limited GP time

Limited GP resource

Patient taking responsibility

Quality and type of follow up

Medication reviews

General practice at center of care

Type and quality of follow up

External factors

Secondary care

Waiting time for psychological treatment

QOF

Health economic considerations

Primary care networks

Wider NHS and GP policy

Appendix 8.1: Qualitative study participants

Table A8.1: Details of qualitative study participants (people with lived experience of depression)

Participant number	Recruitment through (M/GP list)	Gender	Age (years)	Ethnicity	Education status	Relationship status	Employment status	Co-morbidities	IMD* Decile (patient level)	IMD Decile (practice level)
P1	GP Invitation Letter	Female	45	White British	GCSEs	Single	Employed (Administrator)	None	2	1
P2	GP Invitation Letter	Female	40	White British	A Levels	Single	Not employed (cares for children)	Asthma, Hypercholesterolaemia, Arthritis (unspecified)	5	6
P3	GP Invitation Letter	Male	67	White British	Diploma	Married	Retired	None	7	8
P4	GP Invitation Letter	Female	50	White British	Undergraduate degree	Married	Employed	Gastro-oesophageal Reflux	1	2
P5	GP Invitation Letter	Female	46	White British	Undergraduate degree	Married	Employed	Hypothyroidism, Osteoarthritis, Hypertension	9	9
P6	GP Invitation Letter	Female	52	White British	O Levels	Married	Employed	Long Covid	8	3
P7	GP Invitation Letter	Male	48	White British	GCSEs	Married	Employed	Type 2 Diabetes, Bowel cancer (in remission), Anxiety	8	4
P8	GP Invitation Letter	Male	57	White British	Undergraduate degree	Divorced	Employed	None	3	2
P9	GP Invitation Letter	Male	75	White British	O Levels	Married	Retired	Back pain	10	9
P10	GP Invitation Letter	Male	29	Asian (Indian) and British mixed	GCSEs	Co-habiting	Employed	Chronic Fatigue Syndrome, back pain	8	10

P11	GP Invitation Letter	Male	31	White British	Undergraduate Degree	Co-habiting	Employed	None	9	7
P12	GP Invitation Letter	Male	39	White British	GCSEs	Divorced	Employed	Asthma	1	2
P13	GP Invitation Letter	Female	58	White British	GCEs	Divorced	Employed		2	3
P14	GP Invitation Letter	Male	74	White British	GCE (A Level)	Married	Retired	COPD, Back pain	9	9
P15	GP Invitation Letter	Female	37	White British	Postgraduate	Co-habiting	Employed	None	10	9
P16	GP Invitation Letter	Female	24	White British	Postgraduate	Co-habiting	Employed	Migraine	5	9
P17	GP Invitation Letter	Female	36	White British	Postgraduate	Co-habiting	Employed	Polycystic Ovary Syndrome	5	3
P18	GP Invitation Letter	Male	34	White British	GCSEs	Divorced	Employed	None	6	2
P19	Poster advertisement (GP waiting room)	Female	60	White British	Undergraduate degree	Married	Self-employed	Fibromyalgia, Irritable Bowel Syndrome, Knee pain (torn cartilage), Restless legs syndrome, Generalised anxiety disorder	9	9
P20	GP Invitation Letter	Male	25	White British	Foundation Degree	In a Relationship	Self-employed	Chronic back pain,	6	2
P21	Poster advertisement (GP waiting room)	Female	37	White British	Undergraduate Degree	Single	Employed	None.	1	1
P22	GP Invitation Letter	Male	51	White Scottish	Postgraduate	Married	Employed	None.	6	10
P23	GP Invitation Letter	Male	37	British Pakistani	AS Level	Single	Employed	Long Covid.	1	1

Table A8.2: Details of qualitative study participants (GPs)

GP number	Recruitment through (SM/PN/CRN/SB)*	Gender	Age (years)	Ethnicity	Number of clinical sessions per week	Experience (years since qualifying as a GP)	Role (GP contractual status)	Areas of interest and expertise	IMD** Decile (practice level)
GP1	SB	Female	36	White British	5	2.5	Salaried	CCG/ICS work on referral pathways.	9
GP2	SB	Female	38	White British	6	10	Partner	Safeguarding. Women's health.	5
GP3	PN	Female	32	White British	7	2	Salaried	GP Wellbeing.	9
GP4	SB	Female	37	British Asian	5	11	Salaried	Women's health. Medical school tutor.	9
GP5	PN	Male	30	White British	4	1	Salaried	None.	9
GP6	SB	Male	33	British Asian (Pakistani)	2	5	Salaried	Researcher.	2
GP7	CRN	Female	36	White British	6	8	Partner	GP Trainer. Practice Research Lead and	8

								CRN role. Safeguarding. Diabetes Lead.	
GP8	CRN	Female	52	South Asian	4	20	Partner	GP Trainer. Women's Health.	7
GP9	SB	Male	45	White British	10	15	Partner	PCN Clinical Director. Federation work. Community Lead.	9
GP10	SB	Female	44	White British	5	13	Partner	Medical school tutor.	8
GP11	CRN	Male	40	Asian	6	10	Salaried	Clinical Lead.	1
GP12	CRN	Female	43	Black African	6	13	Partner	Practice Research Lead	7
GP13	SB	Male	46	White British	3	11	Partner	Researcher.	2
GP14	CRN	Female	45	White British	2	17	Salaried	Researcher.	9
GP15	CRN	Female	47	White British	6	17	Partner	GP trainer. Asthma. Practice Research Lead.	7
GP16	PN	Male	43	White British	8	11	Locum	None.	N/A
GP17	CRN	Male	35	White British	8	6	Partner	Practice Research Lead.	10
GP18	SB	Female	34	Indian	6	1	Salaried	Emergency medicine.	8

GP19	CRN	Male	61	Italian	9	22	Salaried	Prescribing Lead.	2
GP20	CRN	Male	53	Asian and British mixed	11	26	Partner	Diabetes Lead.	9
GP21	SB	Female	33	White British	4	3	Salaried	Researcher. Women's Health.	4
GP22	SM	Male	36	White British	7	4	Partner	Medical school senior role, urgent treatment centre	10

*PN=Professional Network; SM=Social media; CRN=Recruitment via Clinical Research Network; SB=Snowballing)

** English indices of deprivation 2015 Lower Layer Super Output Area (LSOA) http://imd-by-postcode.opendatacommunities.org/ Decile (1=most deprived, 10=least deprived).
https://fingertips.phe.org.uk/profile/general-practice/data

Appendix 10.1: Thesis dissemination

Peer-reviewed articles

Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Dawson S, Hendon J, Chew-Graham CA, Gilbody S, Churchill R, Phillips RS, Ali S, McMillan D. (2022). Predicting relapse or recurrence of depression: systematic review of prognostic models. *British Journal of Psychiatry.* 1-11. doi:10.1192/bjp.2021.218

Moriarty AS, Paton LW, Snell KIE, Riley RD, Buckman JEJ, Gilbody S, Chew-Graham CA, Ali S, Pilling S, Meader N, Phillips RS, Coventry PA, Delgadillo J, Richards DA, Salisbury C, McMillan D. (2021). The development and validation of a prognostic model to PREDICT Relapse of depression in primary care: protocol for the PREDICTR study. *Diagnostic and Prognostic Research.* 5:12

Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Chew-Graham CA, Gilbody S, Churchill R, Phillips RS, Ali S, McMillan D. (2021). Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. *Cochrane Database of Systematic Reviews.* Issue 5. Art. No.: CD013491. DOI: 10.1002/14651858.CD013491.pub2

Moriarty AS, Castleton J, Gilbody S, McMillan D, Ali S, Riley RD, Chew-Graham CA. (2020). Predicting and preventing relapse of depression in primary care. *British Journal of General Practice.* 70 (691): 54-55. https://doi.org/10.3399/bjgp20X707753. (Commissioned Editorial)

Moriarty AS, Meader N, Gilbody S, Chew-Graham CA, Churchill R, Ali S, Phillips RS, Riley RD, McMillan D. (2019). Prognostic models for predicting relapse or recurrence of depression [Protocol]. *Cochrane Database of Systematic Reviews.* Issue 12. Art. No.: CD013491.

Other Publications

Moriarty A, Meader N, Williams E, Chew-Graham C. "Predicting depression relapse: is this possible and useful?" *Evidently Cochrane* blog, 06 August 2021. https://www.evidentlycochrane.net/predicting-depression-relapse

Moriarty A. Primary Care Mental Health Conference 2021. *Society of Academic Primary Care* blog, 26 May 2021. https://sapc.ac.uk/blog/2021/primary-care-mental-health-conference-2021-blog-andrew-moriarty

Published abstracts

Moriarty AS, Paton LW, Meader N, Snell KIE, Riley RD, Chew-Graham CA, Gilbody S, McMillan D. (2021). Towards an evidence-based decision tool to guide relapse prediction and prevention of depression in primary care. *SAPC North 2021: Abstract Book*

Moriarty AS, Paton LW, Meader N, Snell KIE, Riley RD, Chew-Graham CA, Gilbody S, McMillan D. (2021). Can we PREDICT Relapse of depression in primary care? (Protocol for the PREDICTR Study). *Society of Academic Primary Care ASM 2021: Abstract Book.* doi: 10.37361/asm.2021.1.1

Moriarty AS, Paton LW, Meader N, Snell KIE, Gilbody S, Riley RD, McMillan D, Chew-Graham CA. (2021). Development and validation of a prognostic model to PREDICT Relapse of depression in primary care: the PREDICTR study. *Primary Care Mental Health Research Conference 2021: Abstract Book.* doi:10.37361/sigpcmh.2021.1.1

Moriarty AS, Paton LW, Meader N, Snell KIE, Riley RD, Gilbody S, McMillan D, Chew-Graham CA. (2020). Development of a prognostic model to PREDICT Relapse of depression in primary care (the PREDICTR study). Society of Academic Primary Care North Conference 2020: Abstract Book. doi: 10.37361/n.2020.1.1

Moriarty AS, Meader N, Gilbody S, Churchill R, Phillips B, Ali S, Riley RD, McMillan D, Chew-Graham CA. (2020). Prognostic models for predicting relapse of depression: A Cochrane Prognosis Review and implications for primary care. Society of Academic Primary Care ASM 2020: Book of Abstracts. doi: 10.37361/asm.2020.1.1

Moriarty AS, Meader N, Gilbody S, Churchill R, Phillips B, Ali S, Riley RD, McMillan D, Chew-Graham CA. (2020). Prognostic models for predicting relapse or recurrence of depression: results from a Cochrane Prognosis Review. *Primary Care Mental Health Research Conference 2020: Abstract Book.* doi:10.37361/sigpcmh.2020.1.1

Oral presentations

"Predicting and preventing relapse of depression in primary care: a mixed methods study". *51st Annual Meeting of the North American Primary Care Research Group,* 30 October-3 November 2023, Hilton San Francisco Union Square, San Francisco, California, US. (Accepted for oral presentation).

"Can we reduce the risk of relapse of depression in primary care? A mixed methods study." *Society for Academic Primary Care Annual Summer Meeting 2023*, 19 July 2023. Hilton Brighton Metropole Hotel, Brighton.

"Predicting and preventing relapse of depression in primary care: a mixed methods study." *Hull York Medical School Postgraduate Research Conference 2023*, 5 July 2023. Kingsley Suite, University of Hull. **(HYMS Researcher of the Year Presentation).**

"Predicting and preventing relapse of depression in primary care: a mixed methods study." *Hull York Medical School Research Away Day 2023*, 20 June 2023. King's Manor, York.

"Understanding relapse and improving the ongoing care of people with depression in primary care: A qualitative study." *Primary Care Mental Health Research Conference 2023*, 16 May 2023. Engineer's House, Bristol

"Relapse of depression in primary care: perspectives of General Practitioners and people with lived experience of depression." *Society for Academic Primary Care North Conference 2022*, 10 November 2022. St. Mary's Conference Centre, Sheffield.

"Predicting relapse of depression in primary care: initial results from the PREDICTR prognostic model development study." *Primary Care Mental Health Research Conference 2022*, 25 May 2022. Online.

"Towards an evidence-based decision tool to guide relapse prediction and prevention of depression in primary care." *Society for Academic Primary Care North Conference 2021*, 11 November 2021. Virtual meeting. **(Recipient of Early Career Researcher Prize).**

"Can we PREDICT Relapse of depression in primary care (Protocol for the PREDICTR Study)" – Long oral - *49th Annual Scientific Meeting of the Society for Academic Primary Care*, 30 June-1 July 2021. Virtual Meeting.

"Development and validation of a prognostic model to PREDICT Relapse of depression in primary care (Introduction to the PREDICTR Study)." *Primary Care Mental Health Research Conference 2021*, 26 May 2021. Online. **(Recipient of Early Career Researcher Prize).**

"Predicting relapse of depression in primary care." *Mental Health and Addictions Research Group Seminar Series*, 8 March 2021. University of York.

"Development of a prognostic model to PREDICT Relapse of depression in primary care (the PREDICTR study)." *Society for Academic Primary Care North Conference 2020,* 12 November 2020. Online.

"Prognostic models for predicting relapse or recurrence of depression: A Cochrane Prognosis Review." *Hull York Medical School Postgraduate Research Conference 2020*, 11 Sept 2020. Online.

Definitions

ARRS	Additional roles reimbursement scheme
BMA	British Medical Association
CADET	The Clinical and Cost Effectiveness of Collaborative Care for Depression in UK Primary Care Trial
CASPER	CollAborative care for Screen-Positive EldeRs
CBT	Cognitive Behavioural Therapy
cCBT	Computerised Cognitive Behavioural Therapy
CCA	Constant comparative analysis method
COBRA	Cost and Outcome of BehaviouRal Activation versus Cognitive Behaviour Therapy for Depression
COINCIDE	Collaborative Interventions for Circulation and Depression
COVID-19	Coronavirus disease 2019
DES	Directed enhanced service
GAD	Generalised anxiety disorder
GP	General practitioner
HRA	Health Research Authority
IAPT	Improving Access to Psychological Therapies
ICS	Integrated care system
IECV	Internal-external cross validation
IIF	Investment and Impact Fund
IMD	Index of multiple deprivation
IPD	Individual participant data
IPDMA	Individual participant data meta-analysis
LiCBT	Low-intensity Cognitive Behavioural Therapy
MFP	Multivariable fractional polynomials
MHP	Mental health practitioner
MICE	Multiple imputation by Chained Equations
ML	Machine learning
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research

PAG	Patient advisory group
PCN	Primary care network
PHQ-9	Patient Health Questionnaire
PPIE	Patient and public involvement and engagement
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST	Prediction model Risk Of Bias Assessment Tool
PROGRESS	PROGnosis RESearch Strategy
QOF	Quality and Outcomes Framework
RCT	Randomised controlled trial
REC	Research ethics committee
REEACT	Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
TA	Thematic analysis
WYLOW	West Yorkshire Low Intensity Outcome Watch