****

**Understanding and Predicting Relapse of Depression and Anxiety Following Psychological Interventions Delivered in Routine Services**

**By:**

Ben Lorimer

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

Submission Date: February 2022

**ABSTRACT**

The Improving Access to Psychological Therapies programme (IAPT) in England enables a large population of people with depression and anxiety to receive evidence-based psychological treatments. The programme has been demonstrated to be effective in the short-term. However, the long-term outcomes of IAPT treatments are under-researched, despite it being well-established that depression and anxiety are highly recurrent disorders, and that relapse is relatively common following psychological interventions. Consequently, this thesis sought to contribute to the field’s understanding of relapse following psychological interventions delivered in routine practice, by aiming to improve our ability to predict its occurrence, and to assess its impact on patients and services. To begin, a systematic review (Chapter 2) of predictors of relapse of anxiety-related disorders following cognitive behavioural therapy (CBT) was conducted. An overall pooled relapse rate of 21.8% was estimated, and residual symptoms was identified as the only consistently supported risk factor of relapse. Two studies are then discussed (Chapters 3 and 4) in which a machine learning approach was implemented to predict relapse following low-intensity (LI) and high-intensity (HI) psychological interventions, with the developed models displaying promising predictive potential (AUC = 0.70-0.83). Following this, a qualitative study (Chapter 5) identified psychosocial factors, and in particular work-related stress, as being risk factors for symptomatic deterioration following LI interventions. This study also raised methodological issues regarding current approaches that are used to assess relapse occurrence. The last empirical study in this thesis (Chapter 6) examined the consequences of relapse from a health services perspective, with it being estimated that 13.7% of patients who receive treatment within IAPT subsequently return for additional treatment. Finally, Chapter 7 concludes the thesis by drawing together and synthesising the findings across the thesis, discussing their theoretical implications, and providing recommendations for clinical research and future research.

**ACKNOWLEDGEMENTS**

I would like to begin by acknowledging my supervisory team of Jaime Delgadillo and Steve Kellett. As I have said to many of my friends and family before, I honestly do not think that I could have asked for better PhD mentors. Over the course of my time in Sheffield, the confidence I have in myself and in research has grown exponentially, and this is in large part down to the guidance and support that Jaime and Steve have consistently provided. Whether it be putting my name forward for conference presentations, arranging trips to external research organisations on my behalf, or simply inviting me for a coffee or sending an encouraging email when I was feeling low, they have always made me feel valued and supported. Without their supervision, this PhD would have been a substantially lesser experience.

I would also like to acknowledge a number of people who provided me with much appreciated assistance on the studies included in this thesis. Specifically, I’d like to thank: Arthur Nye for assisting me with the systematic review; James Lawrence for hosting me in London and for teaching me how to implement extreme gradient boosting; Richard Holder, Laura Firth, and Paula Scott-Loftus for helping me source NHS data for one of my studies; Elizabeth Ruth for helping me with the framework thematic analysis; and Julia Giesemann and Wolfgang Lutz for their collaboration in the revolving door study. I would also like to thank the Economic and Social Research Council for funding this PhD project, and Jayne for answering all of my many PhD-related queries.

One of the highlights of my PhD experience has been the close friends that I have made during the process – Amber, Anton, Fidan, Haifa, and others. I am very grateful that we have been able to support each other and share so many funny memories during this process - and I have really appreciated their almost superhero levels of patience at hearing me moan and complain. I would also like to thank Mel - for making sure that conferences were some of my favourite moments of the entire PhD, and for regularly joining me in laughing at Steve’s consistently terrible football opinions. My thanks also go to my non-Cathedral Court friends, including Sam, Rob, Beth, Emily, my QUB friends, the Cratcliffe group, the Jolly Rogers, and the Afungus group, for providing me with support and welcome distractions from my PhD. I’d also like to specifically thank Rhi for using her graphic design wizardry to help me finalise two figures in this thesis.

Finally, I would like thank my family – Mum, Dad, Danny, and of course, my dogs Scout and Eddie. There are no words to describe how grateful I am for each of them. I have had many bad days over the past few years (as well as many good ones), and my parents have always been there to listen and provide support – just as they have always been throughout my life. This PhD simply would not have been possible without their constant guidance, support, and love, and I will never be able to articulate the appreciation I have for them.

I have thoroughly enjoyed the experience of undertaking this PhD. Although it has had plenty of stressful and overwhelming moments, and there have been times where my motivation has been non-existent, I don’t think it could have been a more fulfilling or rewarding experience. I have learned so much over the last three and a half years, and the PhD has at times acted as an anchor of stability during moments of instability in my personal life. I could not be more grateful for the people I have discussed in these acknowledgements - for helping me get through the last few years, and for ensuring that this time has been as rewarding as it has been.

**LIST OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| LIST OF TABLES ……………………………………...……………...………........ | | 6 |
| LIST OF FIGURES ……………………………………………………..………...... | | 7 |
| APPENDICES ………………………………………………..………….................. | | 8 |
| GLOSSARY OF COMMONLY USED TERMS ………………………………….. | | 9 |
| NOTES ON INCLUSION OF PUBLISHED WORK …………………………........ | | 11 |
|  | |  |
| CHAPTER 1: Relapse following evidence-based psychological treatments for depression and anxiety: An introduction ………………………………………........ | | 12 |
|  | The prevalence and impact of depression and anxiety ………..................... | 12 |
|  | Improving access to psychological therapies programme (IAPT) …...………….. | 15 |
|  | Long-term outcomes …...…………………...………............................................ | 17 |
|  | Relapse …………………………..………………………………...…………….. | 18 |
|  | Theories of relapse occurrence ............................................................................... | 20 |
|  | Relapse prevention ………………………………………………………………. | 21 |
|  | Relapse risk factors …………………………………………………………........ | 24 |
|  | Conclusion: Aims of the thesis ………………………………………………….. | 25 |
|  |  |  |
| CHAPTER 2: Predictors of relapse and recurrence following cognitive behavioural therapy for anxiety-related disorders: A systematic review ….................................... | | 27 |
|  | Introduction ……..……………………………...………..…………………......... | 27 |
|  | Method ………..……………………………………..…………………………... | 28 |
|  | Results ……...…………………………...…………..…………………………… | 32 |
|  | Discussion ………..……………………………..……………..………………… | 41 |
|  |  |  |
| CHAPTER 3: Dynamic prediction of cases at risk of relapse following low-intensity cognitive behavioural therapy using a machine learning approach ………………… | | 49 |
|  | Introduction ……………….……………………………………………………... | 49 |
|  | Method …….……………………………...…………………............................... | 50 |
|  | Results …………………………………................................................................ | 60 |
|  | Discussion ………………..……………………………………..……….............. | 65 |
|  |  |  |
| CHAPTER 4: Predicting cases at risk of relapse following high-intensity psychotherapy using a machine learning approach …….....……………………..….. | | 72 |
|  | Introduction ……………………………………...……..………..………............. | 72 |
|  | Method …………………………………………………..…..……………............ | 73 |
|  | Results ………………………………...………………...…..……………............ | 84 |
|  | Discussion …………………….………….…………..………..…………............ | 92 |
|  |  |  |
| CHAPTER 5: A qualitative exploration of patients’ perspectives of symptom deterioration following low-intensity cognitive behavioural therapy …….........…… | | 100 |
|  | Introduction ………………………………..………...…………..…………......... | 100 |
|  | Method ……………………………………..………………..……………............ | 101 |
|  | Results ………………………………………..……………..……………............ | 107 |
|  | Discussion ……………………………..………..……………..……………........ | 116 |
|  |  |  |
| CHAPTER 6: An investigation of treatment return in an Improving Access to Psychological Therapies (IAPT) service ...........…………………………………….. | | 125 |
|  | Introduction …….………………………………………………..……………..... | 125 |
|  | Method ……………………..………………………………..……………............ | 126 |
|  | Results ………………………………………………………..…………….......... | 133 |
|  | Discussion …………………………..….………………………..……………..... | 142 |
|  |  |  |
| CHAPTER 7: General discussion ……………………………………………........... | | 149 |
|  | Summary of thesis findings ………………………..…………………………….. | 149 |
|  | Interpretation of results ……………………………………...……………............ | 151 |
|  | Clinical implications ………………………………………………………........... | 159 |
|  | General strengths and limitations of the thesis ……………………………........... | 162 |
|  | Recommendations for future research directions ………………………………... | 167 |
|  | Overall conclusions ……………..………………………………..…………........ | 169 |
|  | |  |
| References ………………………………………………..…………………............. | | 171 |
| Appendices ……………………….………………………..………………............... | | 199 |

**LIST OF TABLES**

|  |  |
| --- | --- |
| Table 1.1. *Overview of relapse prevention interventions* .…………………………... | 22 |
| Table 2.1. *Characteristics of the included studies* ………………………………....... | 34 |
| Table 3.1. *The potential predictors input into each XGBoost model* ........................... | 54 |
| Table 3.2. *The ability of each model to predict relapse when evaluated on ‘test’ set* . | 61 |
| Table 3.3. *The predictors identified as important by each developed model* ……….. | 63 |
| Table 4.1. *The potential predictors input into the XGBoost models* ………………... | 80 |
| Table 4.2. *The ability of each model to predict relapse when evaluated on ‘test’ set .* | 87 |
| Table 4.3. *The predictors identified as important by each developed model* .............. | 89 |
| Table 5.1. *Selected demographic and clinical information of participants* ………..... | 103 |
| Table 5.2. *Qualitative themes related to the reasons behind patients’ perspectives of their relapse classifications* ......................................................................................... | 108 |
| Table 5.3. *Qualitative themes related to the risk factors participants attributed to their symptom increases* ............................................................................................... | 111 |
| Table 5.4. *Qualitative themes related to the coping strategies that participants discussed employing to address their symptom increases* …………........................... | 113 |
| Table 6.1. *The predictors identified as important by each developed model* .............. | 141 |
| Table 7.1. *Relapse rates for study chapters 3, 4, and 6 recalculated using all pairs of reliable change indices (RCI) applied across the thesis* ………………….................. | 166 |

**LIST OF FIGURES**

|  |  |
| --- | --- |
| Figure 2.1. *PRISMA diagram representing article selection procedure* .………........ | 30 |
| Figure 3.1. *Example decision tree* ………………………………................................ | 57 |
| Figure 3.2. *Distribution of predicted patient relapse probabilities made by each developed model* ........................................................................................................... | 62 |
| Figure 4.1. *Flow chart representing data management and curation process*  …....... | 76 |
| Figure 4.2. *Distribution of lengths of time between patients’ discharge from treatment and first follow-up review appointment* ………........................................... | 77 |
| Figure 4.3. *Distribution of predicted patient relapse probabilities made by each developed model* ………………................................................................................... | 88 |
| Figure 6.1. *Flow chart displaying different pathways taken by patients through the psychological service* ................................................................................................... | 134 |
| Figure 6.2. *Proportions of sample who did not return for treatment, returned for treatment at least once, or returned multiple times* …………………......................... | 137 |
| Figure 6.3. *Average change scores of PHQ-9, GAD-7, and WSAS after each treatment episode received by patients* ........................................................................ | 138 |
| Figure 6.4. *Distribution of predicted treatment return probabilities made by each developed model* ………………................................................................................... | 142 |

**APPENDICES**

|  |  |
| --- | --- |
| Appendix A. *Example systematic review search strategy (Chapter 2)* .……….......... | 199 |
| Appendix B. *Risk of bias ratings for included studies (Chapter 2)* …......................... | 200 |
| Appendix C. *Significant and non-significant predictors of relapse of anxiety disorders after CBT (Chapter 2)* ................................................................................... | 201 |
| Appendix D. *Glossary for predictors (Chapter 3) .......................................................* | 203 |
| Appendix E. *Partial dependence plots for relationships between important predictors and relapse (Chapter 3)* ……….................................................................. | 206 |
| Appendix F. *Glossary for predictors (Chapter 4)* …................................................... | 221 |
| Appendix G. *Partial dependence plots for relationships between important predictors and relapse (Chapter 4)* ……….................................................................. | 223 |
| Appendix H. *Reasons behind participants’ perceptions of relapse classifications, and risk factors and coping strategies discussed by each participant* *(Chapter 5)* .... | 233 |
| Appendix I. *Potential predictors input into XGBoost model and glossary for predictors (Chapter 6)* ................................................................................................. | 237 |
| Appendix J. *Partial dependence plots for relationships between important predictors and relapse (Chapter 6)* .............................................................................. | 239 |

**GLOSSARY OF COMMONLY USED TERMS**

**Cognitive behavioural therapy (CBT)** – psychological therapy that focusses on learning and implementing coping strategies designed to change maladaptive thoughts and behaviours.

**Extreme gradient boosting (XGBoost)** – an ensemble machine learning approach that involves combining several decision trees, which can be implemented to develop predictive models.

**Follow-up review** – an appointment following the end of treatment in which a patient is assessed in order to detect the occurrence, or early signs, of relapse.

**Generalised Anxiety Disorder Scale (GAD-7)** –a seven-item screening tool for anxiety disorders (scored between 0-21) that is commonly used for routine outcome monitoring in IAPT services.

**High-intensity** – psychotherapies delivered in IAPT that are delivered by trained and accredited therapists (e.g., CBT, counselling for depression, interpersonal psychotherapy, psychodynamic therapy).

**Improving Access to Psychological Therapies programme (IAPT)** – a nationwide program in England that enables access to evidence-based psychological interventions for the treatment of depression and anxiety via a stepped-care system.

**Low-intensity** –brief, highly structured, protocol-driven, guided self-help interventions delivered in IAPT that are facilitated by qualified Psychological Wellbeing Practitioners. Interventions based on CBT principles.

**National Institute for Health and Care Excellence (NICE)** – public body that produces guidelines for healthcare provision in English National Health Service.

**Relapse –** defined in this thesis as beingthe clinically significant return of symptoms following the completion of an episode of treatment that resulted in remission of symptoms.

**Relapse prevention** – interventions designed specifically to prevent the occurrence of relapse (e.g., mindfulness-based cognitive therapy, booster sessions).

**Stepped-care** – a service delivery system in which patients are initially offered low-intensity treatments. If patients do not respond to this treatment, or they present with more complex problems, they are ‘stepped-up’ and offered high-intensity treatments.

**Patient Health Questionnaire (PHQ-9)** - a nine-item screening tool for depression (scored between 0-27) that is commonly used for routine outcome monitoring in IAPT services.

**Work and Social Adjustment Scale (WSAS)** – a five-item questionnaire that assesses the extent of functional impairment caused by mental health problems (scored between 0-40) that is commonly used for routine outcome monitoring in IAPT services.

**NOTES OF INCLUSION OF PUBLISHED WORK**

Work that has contributed to two of the chapters in this thesis has been written up as two separate manuscripts that have been accepted for publication, and these co-authored papers are referenced below. The information from these papers has been presented in Chapters 2 and 3 in a format to fit within this body of work, and therefore, although there is some replication, they are not identical to the published papers.

**Chapter 2**

Lorimer, B., Kellett, S., Nye, A., & Delgadillo, J. (2021). Predictors of relapse and recurrence following cognitive behavioural therapy for anxiety-related disorders: A systematic review. *Cognitive Behaviour Therapy, 50*(1), 1-18. https://doi.org/10.1080/16506073.2020.1812709

**Chapter 3**

Lorimer, B., Delgadillo, J., Kellett, S., & Lawrence, J. (2021). Dynamic prediction and identification of cases at risk of relapse following completion of low-intensity cognitive behavioural therapy. *Psychotherapy Research, 31*(1), 19-32. https://doi.org/10.1080/10503307.2020.1733127

**CHAPTER 1**

**Relapse Following Evidence-Based Psychological Treatments for Depression and Anxiety: An Introduction**

The objective of this PhD was to contribute to the field’s understanding of why some patients with depression and anxiety who seemingly recover following a course of psychological treatment experience a clinically significant deterioration of symptoms (i.e., a ‘relapse’) sometime after treatment completion. Long-term outcomes following cognitive behavioural therapy (CBT) and other psychotherapies were investigated over five empirical studies, with the aims of understanding and predicting relapse occurrences. This first chapter introduces key concepts and provides important context that underpins the body of research. To begin, the symptoms of depression and anxiety are outlined, followed by a discussion of their impact on both the individuals who experience such problems and on society more generally. Following this, the English Government’s mental health service delivery system, the Improving Access to Psychological Therapies programme (IAPT), is described. Finally, issues related to the limited research into long-term outcomes and the occurrence of relapse events following psychological interventions are discussed. The chapter concludes with an outline of the aims of each of the remaining chapters that comprise the thesis.

**The Prevalence and Impact of Depression and Anxiety**

Mental health problems are widespread across the globe, with the World Health Organisation (WHO; 2017) estimating that more than 300 million people suffer from depression, and a similar number suffer from an anxiety disorder. Depression is characterised by persistent low mood and/or diminished pleasure in activities, with additional symptoms including feelings of guilt or worthlessness, fatigue, reduced movement, trouble concentrating, and disturbed sleep or appetite (Diagnostic and Statistical Manual of Mental Disorders, 5th ed. [DSM-5]; American Psychiatric Association [APA], 2013). Meanwhile, anxiety-related disorders (including generalized anxiety disorder [GAD], social anxiety disorder, panic disorder, specific phobias, among others) are generally characterised by excessive feelings of anxiety or worry, difficulty in controlling worry, and avoidance of anxiety-inducing situations or triggers (DSM-5; APA, 2013). It has been estimated that approximately 10.8% of the world’s population will experience depression at some point in their lifetime, while 16.6% will suffer from an anxiety-related disorder (Lim et al., 2018; Somers et al., 2006).

The emotional and physical impact of depression and anxiety on an individual level are considerable. Indeed, according to WHO (2017) depression is the largest contributor to global disability, and anxiety disorders are ranked as the sixth largest contributor. The observed declines in work, social, and physical functioning caused by symptoms of depression and anxiety have been demonstrated to be associated with reduced quality of life (Olatunji et al., 2007; Papakostas et al., 2004). Furthermore, depression and anxiety are also associated with an increased risk of developing serious physical illnesses (e.g., heart disease, diabetes, cancer etc.), and engaging in health risk behaviours (e.g., higher rates of smoking and less physical activity; Pratt et al., 2016). In addition, depression has been identified as a major contributor to suicide deaths (Chesney et al., 2014; Too et al., 2019; WHO, 2017). Combined, these factors contribute greatly to the excess mortality rate observed in people suffering from depression and anxiety. Specifically, a recent meta-analysis estimated that people with mental health problems have a mortality rate that is over two times higher than people without them (Walker et al., 2015). This increased risk in mortality was estimated to be associated with approximately 10 life-years being lost among this population. In summary, these findings demonstrate that depression and anxiety are commonly experienced, highly disabling conditions, which can have significantly detrimental impacts on quality of life and overall lifespan.

In addition, the wide prevalence of depression and anxiety and the personal issues caused by them also have a considerable economic impact on wider society more generally. For instance, depression and anxiety have been demonstrated to have substantial indirect economic costs, through reduced work productivity and higher reliance on government benefits (Centre for Mental Health, 2010; Trautmann et al., 2016). For example, mental health problems have been identified as one of the four main reasons behind sickness absence in the United Kingdom (Leaker & Kumar, 2021), and individuals suffering from such problems have also been demonstrated to be less productive when present at work (Johnston et al., 2019; Stewart et al., 2003). Depression and anxiety also have substantial direct economic costs, with healthcare services spending substantial amounts to provide healthcare support to individuals experiencing such problems. Indeed, healthcare systems were estimated to spend a *global* total of US$0.8 trillion on the treatment of mental health problems in 2010, with this amount projected to more than double to US$2 trillion by 2030 (Trautmann et al., 2016). More specifically, the English National Health Service (NHS) spent £14.3 billion on mental health services in 2020/21 alone (Baker, 2021).

In an attempt to address this considerable impact caused by depression and anxiety, many influential theories have been postulated to explain why such problems can develop and be maintained. For example, one theoretical perspective is the biological model of mental health, which attributes mental health problems to genetic predispositions, biochemical imbalances, and structural abnormalities of the brain (Goldstein et al., 2011; Nutt et al., 2002). Meanwhile, the earliest psychological theoretical perspective, the psychodynamic approach, proposes that mental health problems can be explained as reactions to loss, and as conflicts between conscious and unconscious processes (Freud, 1917). Conversely, interpersonal perspectives emphasise the importance of social interactions in the maintenance of mental health problems, with psychological symptoms being perceived to arise as a response to difficulties in everyday relationships with other people (Weissman et al., 2002). In contrast, the cognitive model of mental health argues that depression and anxiety are influenced by maladaptive beliefs and information processing biases that serve to maintain these beliefs (Beck, 1964). Furthermore, the behavioural perspective emphasises the importance of a person’s environment in the development of depression and anxiety, with symptoms being postulated to arise following the removal of positive reinforcement for adaptive behaviours combined with the reinforcement of maladaptive behaviours (Lewinsohn, 1974).

These influential theoretical perspectives have informed the development of different evidence-based treatments for depression and anxiety, including pharmacological treatment (biological), psychoanalytic therapy (psychodynamic), interpersonal psychotherapy (interpersonal), and cognitive behavioural therapy (CBT; cognitive and behavioural). However, despite the development of such treatments and the high expenditure of health services discussed previously, many barriers preventing access to mental health treatment still exist, with the supply of mental health practitioners often not meeting population demand (Andrilla et al., 2018; Health Resources and Services Administration, 2016), and waiting times for treatment often being excessively long (Goldner et al., 2011; Triggle, 2019).

**Improving Access to Psychological Therapies Programme (IAPT)**

To enable more people experiencing common mental health problems to be able to access recommended evidence-based interventions, the United Kingdom (UK) Government introduced the Improving Access to Psychological Therapies programme (IAPT) into the NHS in 2008 (Clark, 2011). This programme was established under the premise that, alongside relieving suffering for a large clinical population, the increased national investment in mental health services would be offset by long-term savings on welfare benefits and physical healthcare, and associated increases in economic productivity (Layard & Clark, 2014). Recent reports indicate that between 2020 and 2021 approximately 1.46 million people were referred to IAPT services, of which 1.02 million received treatment and 658,000 completed treatment (Baker, 2021).

IAPT adopts a ‘stepped care’ service delivery system that implements National Institute for Health and Care Excellence (NICE; 2011) guidelines for the treatment of depression and anxiety (Clark, 2011). In stepped care, patients are initially offered low-intensity psychological treatments, which consist of brief, highly structured, protocol-driven, guided self-help interventions facilitated by trained Primary Care mental health workers (also known as “Psychological Wellbeing Practitioners” [PWPs]). If patients do not respond to this initial treatment, or they present with more severe symptoms or complex problems, they are subsequently ‘stepped-up’ and offered high-intensity interventions, which primarily involve lengthier psychotherapeutic interventions delivered by qualified and accredited therapists. This process highlights the two primary features of stepped care: 1) the recommended treatment for a patient is the least restrictive in terms of cost and personal inconvenience, while also still being likely to provide effective treatment; and 2) the stepped care model is self-correcting, in that if a treatment is not resulting in significant health gains, then changes are made to the patient’s treatment (Bower & Gilbody, 2005).

The current treatment of choice within IAPT is CBT, a therapeutic modality based on the cognitive and behavioural perspectives of mental health. This approach is specifically grounded in the theory that maladaptive interactions between thoughts, feelings, and behaviours can result in the onset and maintenance of mental health problems (Beck, 2011). All low-intensity interventions delivered in IAPT are based on CBT principles, and CBT is the most commonly delivered high-intensity treatment (Baker, 2018; Clark, 2018). CBT’s predominance as the treatment of choice within IAPT is primarily due to it being more systematically investigated than other psychotherapeutic modalities, and it being consistently demonstrated to be both efficacious and effective in the treatment of depression and anxiety (Barth et al., 2013; Cuijpers et al., 2016; Hofmann et al., 2012, Hofmann & Smits, 2008). Importantly, however, although other high-intensity psychotherapies delivered through IAPT (e.g., counselling for depression, interpersonal psychotherapy [IPT], psychodynamic therapy; Baker et al., 2018) have been investigated to a lesser extent than CBT, they have been argued to be just as effective treatments (Barkham et al., 2021; Cuijpers et al., 2013a; King et al., 2014).

Recent research has demonstrated that IAPT can produce promising outcomes for those that access it. For example, a recent meta-analysis of practice-based studies conducted within the IAPT context demonstrated that the programme produces large pre-post treatment effect sizes for reductions in depression and anxiety, and moderate effects sizes for improvements in work and social adjustment (Wakefield et al., 2020). In addition, recent reports indicate that 68.3% of patients who complete a course of IAPT treatment display reliable improvement in their condition, with 51.4% of patients moving to recovery (i.e., no longer having clinically-significant symptoms) following treatment (Baker, 2021). This recovery rate is above the 50% target that has been set for IAPT by the UK Government and NHS England since its inception (Clark, 2011). Indeed, this target has been consistently surpassed since the beginning of 2017 at a national (aggregated) level (Clark, 2019). These findings demonstrate the benefits that the introduction of IAPT has provided to people suffering from depression and anxiety, with greater access to effective evidence-based interventions being provided and more than half of patients recovering following treatment.

However, these promising recovery rates that are associated with the IAPT programme only reflect short-term outcomes, as they only focus on how a patient feels immediately post-treatment. Evaluations of IAPT services do not generally provide information regarding longer-term outcomes, or the extent to which treatment gains are maintained over a longer period of time. Indeed, the previously discussed meta-analysis of practice-based studies conducted within the IAPT context identified a total of 60 eligible studies, and only four of these studies included post-treatment follow-up assessments of outcome data (Wakefield et al., 2020). This illustrates the limited research that has been conducted to investigate the long-term outcomes of IAPT treatments, and therefore how little is known regarding the extent to which patients maintain their gains following treatment completion.

**Long-Term Outcomes**

It is important to investigate the long-term effects of psychological interventions for multiple reasons. First, depression and anxiety are highly recurrent mental health problems, frequently returning following initial experiences of improvement (Hardeveld et al., 2010; Vervliet et al., 2013). For instance, individuals who recover from one episode of depression have an approximate 50% chance of experiencing a subsequent episode, with this probability rising with each additional experienced episode (Burcusa & Iacono, 2007). Similarly, anxiety disorders have been demonstrated to have recurrence rates ranging from 39% to 56% following treatment (Bruce et al., 2005; Vervliet et al., 2013). Therefore, considering that depression and anxiety are known to frequently return following recovery, it is important to assess whether the psychological interventions that are being delivered are effective at preventing deteriorations and returns to suffering from occurring. Second, as discussed previously, IAPT was established under the premise that increased investment in mental health services would be offset by longer-term savings on welfare benefits and physical healthcare, with patients becoming more economically active and productive following treatment. However, this argument relies on the idea that patients remain in a state of recovery following treatment completion. Considering this, in order to be able to assess whether IAPT is capable of effectively fulfilling its promise of reaping long-term economic savings, it is important to assess whether patients are able to maintain their treatment gains following treatment. The above reasons illustrate the humanitarian and health economic importance of understanding the long-term effectiveness of psychological interventions for those patients who recover following an initial episode of treatment.

Looking beyond the context of IAPT and at the wider literature of psychotherapy research more generally, research indicates that CBT and other psychotherapies may have promising long-term outcomes. For example, a recent network meta-analysis of randomised controlled trials found that, when compared to care-as-usual or waiting list at one-year follow-up, several psychotherapies (including CBT, IPT, and psychodynamic therapy) had significantly higher response rates, as indicated by a 50% reduction in symptom severity relative to baseline (Cuijpers et al., 2021). Moreover, another recent meta-analysis of randomised controlled trials found that CBT for anxiety disorders is associated with significantly lower symptoms than control conditions up to one year after completing treatment (van Dis et al., 2019). These findings indicate that CBT and other psychotherapies can have lasting benefits for some patients.

However, there are also limitations with how these reviews investigated long-term outcomes, with both approaches being unable to assess whether treatment gains are maintained long-term. For instance, the network meta-analysis of depression trials examined long-term response relative to baseline for both patients who responded to initial treatment and patients who did not initially respond (Cuijpers et al., 2021). This approach therefore does not focus on those patients who had a positive outcome following treatment, and therefore does not assess whether patients were able to maintain treatment gains following treatment completion. Meanwhile, the meta-analysis of anxiety trials primarily investigated trials that assessed long-term outcomes by aggregating means of symptom severity ratings at follow-up (van Dis et al., 2019). This process masks within-individual change, thus preventing the rate of patients that have experienced a clinically significant increase in symptoms from being understood. In summary, although both of the discussed meta-analyses indicate that CBT and other psychotherapies may be associated with positive long-term outcomes at a group level, they do not provide information regarding individual differences in the long-term maintenance of treatment gains.

**Relapse**

One approach that may allow for this important aspect of psychological treatment to be evaluated is to investigate the occurrence of relapses and recurrences. Relapse is typically defined as a clinically significant return of symptoms after an initial remission of symptoms, but before full recovery has occurred (i.e., an extended period of remission), while recurrence refers to a new episode of an illness after full recovery has occurred (Frank et al., 1991; Rush et al., 2006). Although this definition is generally agreed upon by researchers, there is substantial variability in terms of how they are operationalised (Bockting et al., 2015). For instance, when defining recovery as an extended period of time in which there is a continued remission of symptoms, different researchers have adopted periods of four, six, and 12 months (Bockting et al., 2015; Frank et al., 1991; Rush et al., 2006; Reimherr et al., 1998). Due to the substantial variability in how relapse and recurrence have been operationalised in the literature, and for purposes of simplicity, both definitions will be referred to as ‘relapse’ for the remainder of this thesis (unless otherwise stated). When relapse and recurrence are discussed as separate constructs, a recovery period of 12 months has been operationalised, according to contemporary guidelines (Bockting et al., 2015).

Unfortunately, the investigation of relapse following psychological treatment has been relatively rare. For example, the previously discussed meta-analysis of trials that investigated the long-term outcomes of CBT for anxiety disorders identified only six trials out of 69 that reported relapse rates (van Dis et al., 2019). Furthermore, the sample sizes of these studies were highly underpowered, with the largest sample of patients who received CBT in these six studies being *n*=24. This illustrates the limited research that has been conducted into relapse following psychological interventions.

However, some studies have been conducted in this area, and although psychotherapy has been demonstrated to have lower relapse rates than pharmacological treatment (Cuijpers et al., 2013b; Hollon et al., 2006; Otto et al., 2005; Vittengl et al., 2007), relapse still remains relatively common following psychotherapy. For example, a meta-analysis estimated that 29% of patients with depression who completed CBT with remission of symptoms relapse within one year, with this increasing to 54% within two years when recurrence events are considered (Vittengl et al., 2007). Moreover, a recent meta-analysis of 17 studies estimated that 14% of patients with an anxiety disorder who complete CBT with remission of symptoms relapse following treatment (Levy et al., 2021). In addition, a longitudinal cohort study conducted within IAPT also found that many patients relapse after completing *low-intensity* CBT (LiCBT; Ali et al., 2017). Specifically, they estimated that 53% of patients with depression and/or anxiety that complete LiCBT relapse within one year, with a further 13% experiencing a recurrence in the following year (Ali et al., 2017; Delgadillo et al., 2018b). Finally, although research into relapse following non-CBT-based psychotherapeutic interventions has been even more comparatively limited, a recent randomized controlled trial found that 38% of patients who respond to IPT relapse or experience a recurrence within 17 months of completing treatment (Lemmens et al., 2019). These findings demonstrate that it is relatively common for patients who initially respond following an initial course of psychological treatment to not maintain their treatment gains long-term. Furthermore, they indicate how relapse can have a significant detrimental impact on healthcare costs, leading to service inefficiencies, due to a ‘revolving door’ process whereby patients return for further treatment (Boerema et al., 2016; Roscoe, 2019).

**Theories of Relapse Occurrence**

Multiple explanations have been provided for the high relapse rates associated with depression and anxiety. For example, one suggested hypothesis is that there are individual differences related to premorbid vulnerability. Specifically, this hypothesis proposes that individuals who experience multiple episodes of a mental health problem already possessed the necessary (yet to be identified) characteristics to increase their risk of suffering from recurrent depression or anxiety, and that these characteristics existed even before their initial episode (Burcusa & Iacono, 2007). In contrast, other explanations based on the “scarring” hypothesis argue that the very experience of an episode of depression or anxiety causes a change to underlying factors (e.g., biological, cognitive, stress-related factors) that increases the risk of having additional episodes (Bockting et al., 2015; Burcusa & Iacono, 2007). One example of an explanation related to the process of scarring is the “kindling hypothesis” (Kendler et al., 2000). This model proposes that psychosocial stressors are more influential in the onset of initial episodes of mental health problems than in subsequent episodes, with later episodes being more autonomous in their onset due to a potential process of sensitization to a distressed state.

In addition to these general hypotheses, other psychotherapeutically-focussed explanations for relapse occurrence have been suggested. For instance, it has been suggested that CBT produces lower relapse rates than pharmacological treatment because it is more capable of dealing with the actual problems that are associated with the onset of common mental health problems. For example, it may be that CBT helps ameliorate the causal processes that increase the risk of depression and/or anxiety, or that it provides patients with compensatory skills and strategies that can offset such causal risk processes (Hollon et al., 2006). This hypothesis therefore proposes that relapses generally occur because a person’s initial problem was not sufficiently addressed, or because they have not developed effective coping strategies to prevent or manage the onset of subsequent mental health problems. This may also potentially explain why, although to a lesser extent than pharmacotherapy, relapse still frequently occurs following CBT. For example, it may be that some patients initially recover during therapy because they developed specific coping strategies, but that they then later experience a relapse as these strategies were learned in a very specific context, which were then unable to be utilised effectively in brand new contexts following therapy completion (Bouton, 2002; Craske et al., 2014). Indeed, this may explain the limited durability of LiCBT, considering this form of treatment has limited time to generalize learning to multiple contexts. This therefore highlights the importance of ensuring that the skills that patients learn during therapy are both consolidated and generalized before concluding treatment contact. In summary, various potential mechanisms have been proposed to explain why relapse of depression and anxiety frequently occurs, both on a general basis and more specifically following psychological treatment.

**Relapse Prevention**

Based on the various proposed explanations of relapse occurrence, a range of relapse prevention strategies have been introduced to address the high relapse rates associated with depression and anxiety (see Table 1.1 for an overview). Some of these strategies are embedded within the acute-phase treatments that are provided to patients, including the development of relapse prevention plans in collaboration with patients at the end of treatment (National Collaborating Centre for Mental Health, 2020; Rodgers et al., 2012). These plans typically involve a reflection of the key learning points from therapy, the identification of any potential stressors or setbacks, and the development of plans to manage such stressors when they arise. Additional strategies to prevent relapse that are often embedded within acute-phase interventions include the scheduling of follow-up reviews post-treatment, and providing patients with the option of receiving booster sessions. The terms “follow-up review” and “booster session” are often used interchangeably within the literature, however, according to IAPT guidelines, they represent two different concepts (National Collaborating Centre for Mental Health, 2020). Specifically, follow-up reviews are scheduled (one or more) at the end of treatment to occur at a set-date (usually three to six months after treatment completion), with their primary purpose being to act as a ‘check-in’ and thus provide an opportunity for detecting early signs of relapse. Meanwhile, therapy booster sessions are not scheduled in advance at the end of treatment, but instead occur more spontaneously following treatment completion (usually at request of the patient) when early signs of relapse begin to appear. Indeed, a follow-up review may result in a booster session to be provided if early signs of relapse are identified. Previous systematic reviews have demonstrated that the provision of booster sessions following psychotherapy can help reduce the occurrence of relapse (Gearing et al., 2013; Whisman, 1990). Considering this, the provision of follow-up reviews and booster sessions following treatment completion have been recommended as important components of psychological interventions (Beck, 2011; National Collaborating Centre for Mental Health, 2020).

|  |  |
| --- | --- |
| **Table 1.1**  *Overview of Relapse Preventions Interventions* | |
| Strategy/Intervention | Overview |
| Relapse Prevention Plan (or Blueprint) | A plan developed between patients and therapist toward the end of treatment, in which potential stressors are identified and strategies to manage such stressors are agreed-upon. |
| Follow-Up Review | An appointment scheduled after the end of treatment for a set-date, which acts as a ‘check-in’ and opportunity to detect early signs of relapse. May lead to a booster session. |
| Booster Session | An optional therapeutic session provided to patients after treatment completion. Generally requested by patient upon early signs of relapse. |
| Continuation-Phase CBT/IPT | A psychotherapeutic intervention in which patients continue to undertake the same treatment modality to which they initially responded to. Delivered over several months. |
| Preventive Cognitive Therapy | An intervention, based on CBT, specifically designed to prevent relapse for patients with recurrent depression. |
|  |  |
| Mindfulness-Based Cognitive Therapy | An intervention that integrates elements of CBT with training in mindfulness. Typically delivered in group format. |
| Abbreviations: CBT, cognitive behavioural therapy; IPT, interpersonal psychotherapy | |

Another strategy that has been developed to prevent relapse is for patients to continue undertaking the same treatment modality to which they responded to in the acute-phase. For example, Jarrett et al. (2001, 2008) developed continuation-phase CBT to be provided to those patients who achieved remission of symptoms following acute-phase CBT, with the intervention being delivered for multiple months post-treatment (e.g., eight months). In contrast to acute-phase CBT, in which the reduction of symptoms and acquisition of skills is emphasised, continuation-phase CBT emphasises relapse prevention, reducing residual symptoms, and generalising skills. Meta-analyses have indicated that continuation-phase CBT is effective for the prevention of relapse (Cuijpers, 2013b; Vittengl et al., 2009). Continuation-phase IPT has also been investigated as a possible relapse prevention strategy; however, evidence for this intervention is more mixed (Frank et al., 1990, 2007; Reynolds et al., 2006)

A final strategy that has been suggested to reduce relapse rates is to provide interventions that are specifically designed to prevent relapse. Indeed, multiple relapse prevention interventions have been developed, such as Preventive Cognitive Therapy (PCT; Bockting et al., 2005) and mindfulness-based cognitive therapy (MBCT; Segal et al., 2002). PCT is an intervention based on CBT principles that is specifically developed for patients with recurrent depression who are currently remitted. It contrasts from continuation-phase CBT by being briefer (i.e., delivered in eight weekly sessions), and not being delivered instantly following the end of acute-phase treatment by the same therapist. Meanwhile, MBCT integrates elements from CBT with training in mindfulness, a meditation practice in which one intentionally and non-judgementally focuses their attention on present moment experiences (Baer, 2003). The underlying model of MBCT adopts a kindling perspective of relapse, specifically proposing that individuals with previous episodes of mental health problems are at increased risk of experiencing recurrence, as even mild downward shifts in mood can reactivate patterns of negative, ruminative thinking similar to those that prevailed in previous episodes (Segal et al., 1996). Consequently, MBCT aims to reduce relapse risk by teaching skills that enable patients to disengage from dysfunctional cognitive habits, with ruminative thought patterns being the primary target (Segal et al., 2002). Both PCT and MBCT have been demonstrated to be effective at preventing relapse occurrence. For instance, a recent randomised controlled trial indicated that PCT significantly reduces time to relapse when compared to treatment as usual (de Jonge et al., 2019), while a previous meta-analysis demonstrated that MBCT reduces the risk of relapse of depression by 34% relative to treatment as usual and placebo controls (Piet & Hougaard, 2011).

It has been argued that relapse prevention interventions, such as those discussed above, are most effective when provided to individuals who possess the greatest risk of relapse (Bockting, et al., 2015). For example, MBCT has been demonstrated to significantly reduce relapse risk for patients who have had three of more previous episodes of depression, while no risk reduction has been found for patients who have had only two previous episodes (Piet & Hougaard, 2011). Considering this, it is therefore important that relapse prevention interventions are targeted to those patients who are in greatest need of them. This would ensure that these interventions are provided to those patients who would gain the most from them, while also potentially enabling mental health services, who often operate with limited funds, to deliver such interventions in a cost-effective manner. Therefore, the identification of risk factors associated with relapse should be a priority for future research.

**Relapse Risk Factors**

However, unfortunately relatively little is known about factors associated with relapse after psychotherapy for depression and anxiety. This was highlighted by a recent systematic review that investigated predictors of depressive relapse after CBT (Wojnarowski et al., 2019). This review identified only 13 studies as being eligible, and only two well-supported and replicated predictors of depressive relapse were identified: residual symptoms, and previous number of depression episodes. Specifically, a moderate correlation was estimated between residual symptoms and relapse (*r* = 0.34), while a small correlation was estimated between previous episodes and relapse (*r* = 0.19). This illustrates how little is currently known regarding risk factors of relapse.

Wojnarowski et al.’s (2019) review also identified a number of limitations associated with previous research that has contributed to our currently limited understanding of relapse risk factors. For example, it was identified that only four variables were assessed as potential predictors of relapse in more than one study (residual symptoms, previous episodes, cognitive reactivity, and marital status). Therefore, many potential predictors of relapse have barely been investigated by previous research. In addition, it was highlighted that clinical variables were investigated substantially more than psychosocial or demographic variables. Indeed, only two of the 13 studies investigated demographic variables (Evans et al., 1992; Thase et al., 1992), with only two other studies examining the occurrence of stressful life events as a potential predictor (Harkness et al., 2012, 2014). This demonstrates that previous research investigating relapse has focussed almost exclusively on clinical variables, and neglected other potentially useful domains of information. Furthermore, most studies were grossly underpowered to identify reliable predictors of relapse, and the adopted relapse definitions and measures varied widely across the studies. Similar issues related to sample size and inconsistent outcome measures were also identified in the recent meta-analysis of anxiety relapse rates conducted by Levy et al. (2021). Finally, the Wojnarowski et al. (2019) review itself is limited in that it focussed exclusively on depressive relapse, with no similar systematic review having been conducted to investigate risk factors associated with anxiety relapse following CBT. In summary, these limitations demonstrate that little is currently known about what causes relapse and what factors are associated with it, limiting our ability to predict its occurrence and thus identify those patients who are at greatest risk.

**Conclusion**

The introduction of IAPT in England has significantly improved access to evidence-based psychological interventions for people suffering from depression and anxiety, and the programme is associated with promising post-treatment outcomes. However, the investigation of long-term outcomes has been a neglected area of research, despite it being well-established that depression and anxiety are highly recurrent problems, and that relapse is relatively common following psychological interventions. An improved understanding of relapse following treatment, and the risk factors associated with it, could enable psychological services to offer targeted relapse prevention interventions in a cost-effective manner, and thus provide patients the greatest opportunity to maintain treatment gains long-term. This thesis therefore sought to explore factors related to relapse after psychological interventions in routine practice, and to gain an improved understanding of the impact of its occurrence on both patients and services. The rationale and aims of the following six chapters are outlined below.

**Aims of the Thesis**

* Due to the limited understanding related to risk factors associated with relapse following CBT, the aim of Chapter 2 is to extend upon the systematic review conducted by Wojnarowski et al. (2019) by reviewing and synthesising the contemporary literature on predictors of anxiety relapse following CBT. The prevalence of relapse events following CBT for anxiety-related disorders is estimated, and potential predictors of relapse are discussed.
* Chapters 3 and 4 discuss two studies in which machine learning approaches are applied using IAPT routine outcome data to develop predictive models capable of identifying cases at risk of relapse following psychological interventions. Chapter 3 explores relapse following LiCBT, while Chapter 4 examines relapse following high-intensity psychotherapy (including both CBT and non-CBT-based therapies). Within both chapters, the prevalence of relapse following both intervention formats are discussed, the predictive abilities of the developed prognostic models are evaluated, and potential individual predictors of relapse are explored.
* Chapter 5 attempts to enrich the studies discussed in Chapters 3 and 4 through a qualitative exploration of patients’ perspectives of symptom deterioration following LiCBT. Participant agreement/disagreement with a relapse classification following a longitudinal cohort study is explored, alongside risk factors perceived to be associated with symptom deterioration, and coping strategies discussed as being helpful with dealing such deterioration.
* Chapter 6 discusses an empirical study that investigates the potential consequences of relapse from a health services perspective. Specifically, the extent to which patients who receive psychological treatment within IAPT subsequently return for additional treatment is investigated. A treatment return rate is estimated, and potential mechanisms explaining why patients may return for further treatment are explored. In addition, a machine learning approach is applied in an attempt to predict the occurrence of treatment return, and the predictive ability of the developed model is evaluated.
* Finally, Chapter 7 concludes the thesis by drawing together and synthesising the findings from the systematic review and the four empirical studies. The implications of the research are discussed, and recommendations for clinical practice and future research are provided.

**CHAPTER 2**

**Predictors of Relapse and Recurrence Following Cognitive Behavioural Therapy for Anxiety-Related Disorders: A Systematic Review**

The introductory chapter highlighted that relapse is relatively common following psychological interventions and that various developed interventions are effective at preventing relapse following treatment completion (e.g., mindfulness-based cognitive therapy [MBCT]; Segal et al., 2002). These interventions also appear to provide the greatest benefit when provided to patients at greatest risk of relapse (Bockting et al., 2015). Therefore, the ability to target relapse prevention interventions to such patients would ensure that those at greatest risk receive the support that is needed, while also ensuring that the interventions can be delivered by services in a cost-effective manner. In order for the targeting of these interventions to be effective, an understanding of the risk factors associated with relapse following psychological interventions is required. Unfortunately, understanding regarding this is currently limited. However, as discussed in Chapter 1, a recent systematic review and meta-analysis attempted to address this knowledge gap by reviewing the current literature on predictors of depression relapse following cognitive behavioural therapy (CBT; Wojnarowski et al., 2019). Although limitations associated with small sample sizes and inconsistent relapse measurements were identified, this review calculated a pooled relapse rate for depression of 33.4% across 13 studies, and also found consistent support for two predictors of depressive relapse: the presence of residual depressive symptoms, and prior episodes of depression.

However, little is known regarding what factors are associated with relapse following CBT for common mental health problems other than depression (i.e., anxiety-related disorders, including obsessive-compulsive disorder [OCD], and post-traumatic stress disorder [PTSD]). Although a recent meta-analysis investigated relapse rates of anxiety-related disorders following CBT (estimating a pooled relapse rate of 14%; Levy et al., 2021), this meta-analysis did not specifically investigate predictors of relapse and instead focused primarily on relapse prevalence. Therefore, to address this gap in the literature, this chapter discusses a study in which the contemporary literature on predictors of relapse of anxiety-related disorders following CBT were reviewed. This review aimed to estimate the prevalence of relapse events and to identify predictors of relapse using systematic review and meta-analytic methods.

**Method**

**Protocol and registration**

The systematic review protocol was prospectively registered and published in the international Prospective Register of Systematic Reviews (PROSPERO) database (Protocol ID: CRD42019133033).

**Eligibility Criteria**

To be included in this review, studies must have (1) included an adult (18+) sample of patients who had been diagnosed with an anxiety disorder, PTSD, and/or OCD, and (2) who had completed a course of CBT with remission of symptoms as identified by validated measures and/or diagnostic interviews. Co-morbidity of other mental health disorders was allowed, but the primary disorder must have been an anxiety-related disorder. The review also only included (3) longitudinal cohort studies or randomised controlled trials that had been (4) published in the English language (5) in peer-reviewed journals, and that (6) included a follow-up period of at least twelve weeks and (7) investigated at least one potential predictor of relapse.

As this review aimed to improve understanding of risk factors associated with relapse following acute-phase CBT, studies were excluded if any formal maintenance intervention designed to prevent relapse had occurred. This was to ensure that any identified predictors of relapse are associated with the delivery of CBT, and not with the delivery of maintenance interventions. However, studies that did not control for participants receiving additional, external therapy during follow-up (i.e., not provided as part of the study) were included. Considering that services and trials cannot disallow participants from seeking additional psychological support, these studies were included for three reasons: (1) to allow for a wider synthesis of the literature; (2) greater external validity, as this situation likely reflects best what occurs in routine practice; and (3) to enable an assessment of the extent to which studies did not control for this factor. There were no exclusion criteria associated with medication use, however studies that introduced pharmacological treatment as a maintenance intervention were excluded.

**Search Strategy**

Four databases (PsycINFO, PubMed, Scopus, and Web of Science) were searched for relevant articles published between January 1st 1990 and May 17th 2019, using a predetermined search strategy (Appendix A). This strategy consisted of variations of the keywords: ‘cognitive behavioural therapy’; ‘relapse’; ‘predict’; and variations of each of the investigated disorders (‘separation anxiety disorder’, ‘selective mutism’, ‘specific phobia’, ‘social anxiety disorder’, ‘panic disorder’, ‘agoraphobia’, ‘generalised anxiety disorder’, ‘PTSD’, and ‘OCD’).

**Study Selection**

The search strategy identified 233 unique records. After screening of titles and abstracts by a single reviewer, 208 ineligible studies were excluded. Following this, two reviewers independently assessed the full-texts of the remaining 25 articles and, both agreed only four of these articles were eligible for the review. The most common reasons for exclusion were: different treatments being grouped together for analysis; inclusion of a maintenance intervention; and the lack of a relapse outcome measure. Reference lists of the four eligible articles were also hand-searched, and new studies that cited the eligible articles were searched in Web of Science. This step identified four additional eligible articles. Finally, one other eligible article was identified outside of the systematic search process. Therefore, a total of nine eligible articles were included in this review. The corresponding authors of these eligible articles were contacted by e-mail to request further references that may be eligible, but this did not produce any additional eligible articles. A PRISMA diagram (Moher et al., 2009) summarising the selection process is illustrated in Figure 2.1.

**Figure 2.1**

*PRISMA Diagram Representing Article Selection Procedure*

Records identified through database searching

(*n* = 383)

Records after duplicates removed

(*n* = 233)

Records screened: Titles and Abstracts

(*n* = 233)

Records excluded

(*n* = 208)

Records assessed for eligibility: Full texts

(*n* = 25)

Records excluded (*n* = 21), with reasons:

*n* = 6 - different treatments analysed together;

*n* = 5 - maintenance interventions;

*n* = 4 - no measure of relapse;

*n* = 2 - no predictors of relapse;

*n* = 2 - analysis included patients who did not respond to treatment;

*n* = 1 - no diagnosis of anxiety;

*n* = 1 - no intervention investigated

*Many studies were excluded for multiple reasons.*

Hand-searched papers identified via references lists and reverse-citing of eligible records

(*n* = 4)

Number of papers included in qualitative synthesis

(*n* = 9)

Records identified by additional hand-searching

(*n* = 1)

**Risk of Bias Assessment**

Two reviewers independently assessed the risk of bias of the nine eligible articles, using an adapted[[1]](#footnote-1) version of the CASP Cohort Study Checklist (Critical Appraisal Skills Programme, 2018). Interrater agreement for the two reviewers’ assessments was found to be fair[[2]](#footnote-2) (Cohen’s kappa = 0.34), thus demonstrating significant disagreement between the reviewers in terms of their initial risk of bias assessments. However, these disagreements were resolved through discussion, with reviewers reaching complete consensus on all items.

**Data Extraction and Synthesis**

Data were extracted and tabulated by one reviewer using a structured form developed with guidance from the Cochrane Collaboration Data Collection Form (Cochrane Collaboration, 2014). A narrative synthesis of the characteristics, methods and results of the identified studies was conducted. If the investigation of an individual potential predictor was replicated across multiple studies and there was sufficient reporting of statistical information, a random-effects meta-analysis was also conducted to enable a quantitative synthesis of data. This was performed using the ‘Meta-Essentials’ Excel workbook for meta-analysis (Suurmond et al., 2017). To enable meta-analysis and the calculation of a pooled effect size, relevant inferential statistics (e.g. t-test, chi-square, log rank) were transformed into correlation coefficients (*r*). Heterogeneity was assessed using the Q and I2 statistics (Higgins et al., 2003), while potential publication bias was examined through investigation of funnel plots using Egger’s regression and rank-correlation tests. The small number of eligible studies prevented subgroup or moderator analyses from being possible.

**Results**

**Study Characteristics**

The nine studies deemed eligible for review are described in Table 2.1, with all of the studies being longitudinal cohort studies. The mean age of participants within each study ranged from 30.6 to 42.4 years. Two pairs of studies were related. Braga et al. (2010) was an investigation that continued on from Braga et al. (2005), and therefore used the same sample. Meanwhile, Fava et al. (1995) was a preliminary investigation that preceded Fava et al. (2001b), and therefore used a sample that was later a subset of the latter article.

Panic disorder was the most commonly investigated disorder (*n*=4 studies). For the remaining five studies, two investigated OCD, two investigated social anxiety disorder, and one investigated a mix of anxiety disorders. No studies specifically examined separation anxiety disorder, selective mutism, specific phobia, generalised anxiety disorder, or PTSD. There was also variation in terms of the CBT interventions investigated; three studies explored relapse after one-to-one CBT, three following cognitive behavioural group therapy (CBGT), and three following exposure therapy. No studies reported using a treatment fidelity check or therapist competency measures. Moreover, one study exploring one-to-one CBT (Lincoln et al., 2005) used a relatively idiosyncratic treatment protocol. This protocol consisted of a five-to-seven-day intensive treatment phase consisting of high-density exposure with cognitive restructuring. Following this, participants were instructed to continue exposure in their everyday lives for six weeks, and were provided with further support from their therapist during this time if needed. The authors acknowledged that their adopted “format of treatment differed from existing approaches” (p. 212), with CBT for common mental health problems typically being delivered through weekly sessions (Arch et al., 2012; Cuijpers et al., 2013c).

***Relapse Definitions***

There were differences in how relapse was operationalised. Five studies assessed relapse status using the clinical global impressions scale (CGI; Guy, 1976). This scale assesses a clinician’s subjective perception of the global severity of a disorder based on interview, and ranges from 1-7 (minimal to severe symptoms). DiMauro et al. (2013b) used the CGI, but in contrast to other studies, used a self-report version. For every study that used the CGI, patients had to have a score greater than two (“borderline ill”) to meet criteria for relapse. However, only one study (Otto et al., 1996) used this as the only criterion for relapse; other studies had additional criteria in their outcome assessments, and these varied across studies (see Table 2.1). For example, in one study (Heldt et al., 2011) patients had to have either a CGI score of greater than two *or* be suffering from panic attacks to be classified as having relapsed. In contrast, in Braga et al.’s (2005; 2010) study patients had to have a CGI score of greater than two *and* an increase of at least 35% on the Yale-Brown Obsessive Compulsive Scale (Y-BOCS; a validated measure of OCD symptom severity; Goodman et al., 1989) to be classified as having relapsed. Conversely, DiMauro et al. (2013b) classified patients as having relapsed when they had a CGI score of greater than two *and* a score of greater than six on the Sheehan Disability Scale (SDS; Sheehan et al., 1996), which measures a patient’s functioning in work, social situations, and with family.

There were two definitions of relapse that did not use the CGI. One was used in Fava et al.’s (1995, 2001a, 2001b) studies where relapse was defined as the return of panic disorder or social anxiety disorder as defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM[[3]](#footnote-3)). Meanwhile, Lincoln et al. (2005) was the only study to assess relapse by calculating reliable change indexes (RCI), which is a psychometric criterion that assesses whether a change over time of an individual score is statistically significant and not a reflection of measurement error (Jacobson & Truax, 1991).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.1**  *Study Characteristics* | | | | | | | | | | | | |
| Study | Design | Country | Primary Disorder | Intervention | N Total | N Responded to Treatment, Completed Follow-Up | N (%) Subjects Relapsed | Mean Agea | Follow-Up Length | Relapse Definition | N Predictors | Some Patients Receive Follow-Up Therapy? |
| Braga et al. (2005) | LCT | Brazil | OCD | CBGT | 42 | 31 | 11 (35%) | 36.8 | 12 months | 1) Increase ≥35% Y-BOCS and 2) CGI > 2 | 6 | Yes |
| Braga et al. (2010)b | LCT | Brazil | OCD | CBGT | 42 | 31 | 13 (42%) | 36.8 | 24 months | 1) Increase ≥35% Y-BOCS and 2) CGI > 2c | 1 | Yes |
| DiMauro et al. (2013b) | LCT | United States | Mixed | CBT | 181 | 26 | 4 (15%) | 42.4 | 1-6 yearsd | 1) Self-report CGI-S > 2 and 2) SDS > 6 | 9 | Yes |
| Fava et al. (1995) | LCT | Italy | PD (with AG) | BE | 110 | 81 | 15 (19%) | 35.8 | 2-9 years (median=4 years) | The occurrence of DSM-III-R PD | 15 | Noe |
| Fava et al. (2001a) | LCT | Italy | SA | BE | 70 | 45 | 6 (13%) | 30.6 | 2-12 years (median=6 years) | The occurrence of DSM-IV SA | 18 | No |
| Fava et al. (2001b)f | LCT | Italy | PD (with AG) | BE | 200 | 132 | 31 (23%) | 34.5 | 2-14 years (median=8 years) | The occurrence of DSM-IV PD | 16 | No |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Heldt et al. (2011) | LCT | Brazil | PD | CBGT | 56 | 56 | 17 (30%) | 38.8 | Second follow-up year of a two-year follow-up period g | CGI > 2 or the presence of panic attacks | 20 | No mention |
| Lincoln et al. (2005) | LCT | Germany | SA | CBT | 287 | 90 | 12 (13%) | 33.9 | 12 months | RCI of SCL-IS < -1.96 | 36 | Yes |
| Otto et al. (1996) | LCT | United States | PD | CBT | 40 | 40 | - | 38.7 | 24 months | Two consecutive months with CGI > 2 | 14 | No mention |
| aThe mean age of a study’s overall sample is reported here if the mean age of the subsample investigated for the purposes of predicting relapse was not available.  bBraga et al. (2010) is a follow-up of Braga et al. (2005).  cBraga et al. (2010) report that a patient must have CGI > 1 to be classified as having relapsed. This is inconsistent with their previous study that this study is associated with (Braga et al., 2005), in which patient CGI scores must be > 2 in order for a relapse to be classified. Therefore, it is assumed that this is a typo, and the definition used by Braga et al. (2005) is reported instead.  dDiMauro et al (2013b) reported that follow-ups occurred one-year post-treatment. However, this was later corrected (DiMauro et al., 2013a) and follow-ups actually ranged from 1-6 years.  eThis is not stated in this article, but is stated in Fava et al. (2001b).  fFava et al. (2001b) is an extension of the preliminary study Fava et al. (1995).  gHeldt et al. (2011) only investigated relapses that occurred in the second follow-up year of two. Relapses that occurred within the first follow-up year were not considered.  Abbreviations: AG, agoraphobia; BE, behavioural exposure; CBGT, cognitive and behavioural group therapy; CBT, cognitive behavioural therapy; CGI, Clinical Global Impressions scale; CGI-S, severity scale of the CGI; DSM-III-R, revised version of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM); DSM-IV, fourth edition of the DSM; LCT, longitudinal cohort study; PD, panic disorder; RCI, reliable change index; SA, social anxiety disorder; SCL-IS, Symptom Checklist-90 Interpersonal Sensitivity subscale; SDS, Sheehan Disability Scale; and Y-BOCS, Yale-Brown obsessive-compulsive scale. | | | | | | | | | | | | |

***Follow-Up***

There was considerable variation in the length of follow-up periods, with there being a range of 1-14 years. Five of the nine studies had relatively brief follow-up periods that did not exceed 24 months. The three studies conducted by Fava et al. (1995, 2001a, 2001b) had median follow-up durations of four, six, and eight years respectively. These were median durations, as each of the three studies did not have standardised follow-ups, but instead had a range of follow-up durations. For example, within Fava et al.’s (2001b) study some patients were followed-up for 14 years, while some patients had only been followed up for 2 years at the time of analysis. DiMauro et al. (2013b) reported that patients were followed-up one year after treatment, but a later corrigendum corrected this error by reporting that follow-ups occurred one-year post-treatment for only 37% of patients (DiMauro et al., 2013a). In fact, 33% of follow-ups occurred two-years post-treatment, and 30% occurred between three- and six-years post-treatment. Variability in follow-up durations may have impacted upon relapse rates, as there is increased opportunity for relapse events to occur with longer follow-ups. However, the three Fava et al. studies and Di Mauro et al (2013b) had four of the five lowest relapse rates of the included studies.

It is important to note that Heldt et al. (2011) only investigated relapse events that occurred in the second year of a two-year follow-up period, and therefore relapse events that occurred in the first year post-treatment were not considered. A previous study of the same sample (Heldt et al., 2006) investigated treatment response one-year after treatment completion, but did not consider relapse in this timeframe and therefore was not eligible for this review. An additional aspect of follow-up that varied across studies was whether or not some patients received additional, external therapy during the follow-up period. Only three of the nine studies (Fava et al., 1995[[4]](#footnote-4), 2001a, 2001b) reported that no patients received additional therapy during this period, while four studies reported that some patients did receive additional therapy. The remaining two studies did not mention whether patients received further therapy post-treatment or not, with it not being possible to contact the corresponding authors of these studies for clarification.

***Risk of Bias Assessment***

Seven of the nine studies were rated as “good”, one was rated as “fair”, and one was rated as “poor”. Details about the ratings of individual studies can be found in Appendix B. Fava et al. (1995) was rated as “fair” for the following reasons: 1) exclusion of patients with co-morbid major depression despite investigation of psychiatric co-morbidity as a potential predictor of relapse; 2) failure to consider the possibility that patients received further therapy during the follow-up period; 3) patients having different lengths of follow-up (e.g. two-years versus nine-years); and 4) providing highly imprecise statistics (i.e. only reporting a vague *p* value for one of two significant predictors). DiMauro et al. (2013b) was rated as “poor” for the following reasons: 1) self-report scores were used for predictor variables, however when these were missing, therapist-rated scores were used instead; 2) it was not reported who collected outcome data, thus it may have been a researcher; 3) failure to take into consideration for analysis multiple confounding variables (i.e. the effects on relapse of diagnosis, change in follow-up data collection method (outcomes collected by phone, rather than face-to-face as previously collected), and receiving further treatment during follow-up); 4) failure to consider 50 patients who were lost to follow-up (i.e. did not investigate differences between those who dropped out and those who did not, or consider them in analysis); 5) having large variability in lengths of follow-up, which was not reported in the original article or taken into account in analysis; and 6) imprecision and inconsistency in the reporting of results. The two most common sources of bias across the nine studies were the failure to take into consideration in study design and/or analysis the potential confounding variable of patients receiving further treatment during follow-up, and consistent imprecision when reporting statistics.

**Rates of Relapse**

Relapse rates ranged from 13% to 42%, and an overall pooled relapse rate of 21.8% was calculated (excluding Otto et al., 1996, which did not report a relapse rate, and Braga et al., 2005, and Fava et al., 1995, which were both associated with two other studies). OCD was the disorder with the highest relapse rate, with 35% relapsing within 12 months (Braga et al., 2005), and 42% experiencing a relapse/recurrence within 24 months (Braga et al., 2010). Social anxiety had the lowest relapse rate, with 13% relapsing at twelve months (Lincoln et al., 2005) and two to twelve years’ follow-up (median = six years; Fava et al., 2001a).

**Predictors of Relapse**

The nine studies investigated a total of 147 variables as potential predictors of relapse, with 21 significant predictors (*p*<0.05) being identified (see Appendix C). These predictors can be grouped into seven categories: residual symptoms; personality disorders; medication; clinical features; stressful life-events; degree of improvement; and demographics. The majority of the results are discussed in narrative form. The ‘residual symptoms’ predictor was replicated across multiple studies and had sufficient statistical information for a quantitative synthesis.

***Residual Symptoms***

The presence of residual symptoms of the primary disorder at the end of treatment was found to significantly predict relapse in a majority of studies (Braga et al., 2005; Braga et al., 2010; Fava et al., 1995; Fava et al., 2001a; Fava et al., 2001b; Heldt et al., 2011). In contrast to these studies, Lincoln et al. (2005) and DiMauro et al. (2013b) both found no significant effect of post-treatment levels of anxiety symptoms on relapse. However, the study by DiMauro et al. (2013b) had a particularly small sample consisting of only four relapse events, and this may explain the lack of a significant effect.

A meta-analysis was conducted to quantitatively synthesise the results of these studies. Five of the eight studies that investigated the predictive role of residual symptoms on relapse were included in the meta-analysis. Fava et al. (1995) and Braga et al. (2005) were excluded as these studies were associated with Fava et al. (2001b) and Braga et al. (2010) respectively, while DiMauro et al. (2013) was excluded as this study provided insufficient statistical information. The pooled effect estimated by the meta-analysis represented a moderate correlation between residual symptoms and relapse, however it was not found to be statistically significant (*r =* 0.35 (95% CI -0.21, 0.74), *p* = .08). There was evidence of considerable heterogeneity (Q = 56.68, *p* < .001; I2 = 92.94%), while regression (*t* = 0.93, *p* = .42) and rank correlation tests (Kendall’s Tau = 0.40, *p* = .33) for funnel plot asymmetry suggested no evidence of likely publication bias. Given the small number of eligible studies that provided data for meta-analysis, no further sensitivity analyses were carried out.

Three studies also investigated residual symptoms of conditions different from the studies’ primary target conditions as predictors of relapse. For example, Fava et al. (1995, 2001a, 2001b) and Lincoln et al. (2005) all investigated residual levels of depression as a potential predictor of relapse of panic disorder/social anxiety disorder. Moreover, as additional potential predictors of social anxiety disorder relapse, Fava et al. (2001a) investigated residual generalised anxiety and somatic anxiety, while Lincoln et al. (2005) investigated residual levels of agoraphobia, obsessive-compulsiveness, and hypochondriasis. None of these variables were found to be significant predictors of relapse. This may potentially indicate that only residual symptoms of the primary target condition are predictive of relapse, however small sample sizes may also explain the lack of significant findings.

***Personality Disorders***

Fava et al. (1995, 2001b) and Fava et al. (2001a) found that participants with a co-morbid personality disorder were significantly more likely to relapse. The presence of a personality disorder was also the strongest predictor identified in both studies. No other study investigated this variable as a predictor of relapse.

***Medication***

Post-treatment use of medication had mixed results as a predictor of relapse, with a total of four studies investigating its effect, and two observing a significant effect. Specifically, three studies examined the effects that post-treatment use of antidepressants had on relapse (Fava et al., 2001a, 2001b; Heldt et al., 2011), with only one finding it to be a significant predictor (Fava et al., 2001b). Similarly, the same three studies investigated post-treatment use of benzodiazepines as a potential predictor, with two finding a significant effect (Fava et al., 2001a, 2001b). A different study, conducted by Otto et al. (1996), did not separate the two forms of medication in their analyses, instead investigating post-treatment use of antidepressants and/or benzodiazepines as a single predictor of relapse. Although they initially found a significant effect, this became non-significant (p>0.05) when accounting for the additional variable of agoraphobic subtype.

***Clinical Features***

Three different clinical features were found to be significant predictors of relapse in one study each. The first was baseline severity of disorder, which was found by Otto et al. (1996) to significantly predict panic disorder relapse. However, this effect was not replicated by other studies (Braga et al., 2005; Di Mauro et al., 2013; Fava et al., 1995; Fava et al., 2001a; Fava et al., 2001b; Heldt et al., 2011; Lincoln et al., 2005). The second feature found to be a significant predictor of relapse was initial levels of depressed mood (Fava et al., 2001b). However, this effect was not observed by two different studies (Fava et al., 2001a; Lincoln et al., 2005), with co-morbid depression also not being a significant predictor in two other studies (Heldt et al., 2011; Otto et al., 1996). The final clinical feature was having a specific subtype of a disorder. Lincoln et al. (2005) found that patients with social anxiety disorder who relapsed reported a more generalised subtype of the disorder (i.e. experienced fear towards a greater range of social situations) at baseline compared to patients who remained in remission. Furthermore, Otto et al. (1996) found that patients who had panic disorder with the agoraphobic subtype were significantly more likely to relapse. In contrast however, Heldt et al. (2011) found that baseline severity of agoraphobia did not predict relapse.

***Degree of Improvement***

Braga et al. (2005) found that patients with OCD who had a larger reduction in symptom severity were less likely to relapse than patients who achieved less intense improvement. This finding was not replicated by DiMauro et al. (2013b) in a study investigating patients with a range of anxiety disorders. However, the small sample size of this study (*n*=4 relapse cases) may explain the lack of a significant finding.

***Stressful Life-Events***

Only one study, conducted by Heldt et al. (2011), explored the role of stressful life-events on relapse. They found that the experience of a stressful life-event characterised by “conflict” (i.e., interpersonal relationship difficulties, or occupational or financial problems) during a two-year follow-up period was a significant predictor of a relapse that occurred in the second-year of the follow-up period. Stressful life-events characterised by “loss” (e.g., death of a loved one, divorce), “medical illness” (i.e., onset or exacerbation of medical condition), or “other” (i.e., events that could not be categorised into the other three groups) did not predict relapse. However, more participants had experienced a “conflict” event (*n*=17) than a “loss” event (*n*=6), “medical” event (*n*=7), or “other” event (*n*=12), and this relatively larger subsample may have increased the opportunity of an effect to be identified.

***Demographics***

Age was the only demographic predictor to be found significant by any study, with younger participants suffering from panic disorder with agoraphobia being more likely to relapse (Fava et al., 2001b). Lincoln et al. (2005) also reported that younger patients with social anxiety disorder were more likely to relapse than older patients. However, this difference was no longer significant when the additional predictor of generalised subtype was taken into consideration. Indeed, age was found to be a non-significant predictor in three other studies (Fava et al., 2001a; Heldt et al., 2011; Otto et al., 1996). Other demographic variables investigated but not found to be significant were: gender, marital status, education; social class; and employment.

**Discussion**

This is the first study to systematically review the literature on predictors of relapse in anxiety-related disorders following completion of CBT. As such, this provides complementary evidence to the review by Wojnarowski et al. (2019), which specifically examined depression relapse following CBT. The pooled relapse rates found were similar between the two reviews (33.4% for depression; 21.8% for anxiety reviewed in this study), with this chapter’s relapse rate also being similar to the rate reported in Levy et al.’s (2021) meta-analysis of anxiety relapse rates after CBT (14%). The relatively higher relapse rate reported in this chapter’s review may be partly explained by this review including treatment refractory samples, and Levy et al.’s (2021) review excluding such samples. Overall, this demonstrates the high rates of relapse associated with common mental health problems, with anxiety-related relapse potentially being relatively less common than depressive relapse. Strict inclusion criteria were created and followed in this review, identifying only those studies that investigated the durability of acute-phase CBT not augmented with maintenance interventions designed to prevent relapse (e.g. booster sessions, MBCT). This criterion was followed to ensure that any predictors of relapse could be confidently associated with the delivery of CBT, and not with the role played by maintenance interventions.

The strict inclusion criteria were defined *a priori* however, and this resulted in only nine studies being identified as eligible. Only one variable investigated as a potential predictor of relapse was consistently supported as being significant in more than two studies: residual symptoms related to the primary disorder. This variable was a significant predictor of relapse in four studies, and non-significant in two studies. However, one of these two studies was the only study rated as having poor methodological quality and only involved four relapse cases in their analyses. A meta-analysis subsequently estimated a moderate positive correlation between residual symptoms and relapse, although this was not statistically significant (*p* = .08). This was potentially due to the analysis only involving five studies, and indeed this precluded more detailed sensitivity analyses to examine potential sources of heterogeneity. Wojnarowski et al.’s (2019) meta-analysis of predictors of relapse of depression also identified residual symptoms to be an important predictor of relapse (Wojnarowski et al., 2019). In fact, the pooled effect size was highly similar across both reviews (*r* = 0.34 for depression; *r* = 0.35 for anxiety). This indicates that the presence of residual symptoms is currently emerging as a risk factor of relapse across common mental health problems. Paykel (2008) suggested that residual symptoms were a predictor of relapse as they represent the persistence of the original disorder, albeit in a milder presentation. However, understanding as to why residual symptoms appear to predict relapse remains limited.

Despite this review only identifying one replicated predictor of relapse, other potential predictors were identified. For example, the presence of a personality disorder was only investigated by two studies, but was found to be the strongest predictor in both. This further highlights the limited research into the impact of co-morbid personality disorders on the outcomes of CBT, and psychological therapies more generally. Indeed, none of the studies reviewed by Wojnarowski et al. (2019) investigated the presence of personality disorders as a potential predictor. Furthermore, a recent scoping review investigating the effectiveness of psychological therapies at treating patients with depression and/or anxiety who have co-morbid personality disorders concluded that there is a dearth of research in this area, and that no firm conclusions can be drawn (French et al., 2017). Clearly, more research is required to understand the influence the presence of personality disorders have on CBT outcomes for patients with common mental health problems.

Another potential predictor identified by this review is the degree of improvement patients experience over the course of treatment, with larger symptom improvement appearing to be protective against relapse. This was a significant predictor in one study, and although the only other study to investigate this variable did not find it to significantly predict relapse, this other study was assessed as being of “poor” quality and only explored four relapse cases. Smaller symptomatic improvement may be predictive of relapse for similar reasons as those posited by Paykel (2008) in relation to residual symptoms, with it potentially representing a relatively diminished responsiveness to treatment, and the consequent persistence of the disorder in a milder form.

Finally, a third variable that holds promise in the prediction of relapse is occurrence of a stressful life-event during the post-treatment follow-up phase. Only one study investigated this variable, and only one form of stressful event (i.e., an event characterised by “conflict” vs “loss”, “medical”, or “other”) was found to be significant. However, it is possible sample constraints limited the opportunity for the other forms of stressful events to be identified as significant predictors. The systematic review on predictors of depressive relapse similarly identified only one study that investigated the occurrence of stressful life events during follow-up as a predictor, and this study also found a significant effect (Harkness et al., 2014; Wojnarowski et al., 2019). Interestingly, Harkness et al. (2014) found that exposure to stressful events mediated the predictive relationship between residual symptoms and depressive relapse. This may potentially indicate an explanation, alternative to the one posited by Paykel (2008), as to why the presence of residual symptoms may predict relapse. In summary, although these variables should be considered with caution due to limited investigation, they hold promise as potential predictors of relapse for future research.

There were a number of limitations associated with the studies included in this review. For example, the studies may not have identified statistically significant predictors due to the small sample sizes across all of the studies. As shown in Table 2.1, sample sizes ranged from *n*=40-132, while the numbers of relapse cases ranged from *n*=4-31. These samples were highly likely to be underpowered to identify significant predictors that may have even a large effect on the risk of relapse. For example, Cohen’s sample size estimate criteria (Cohen, 1992) suggests that for a comparison of means with a continuous predictor using ANOVA at least *n*=26 relapse cases would be required to detect a large effect size with 80% power. Similarly, *n*=64 relapse cases would be needed to detect a medium effect size, and *n*=393 cases for identification of a small effect size. Out of the nine studies included in this review, only Fava et al. (2001b; *n*=31 relapse cases) had a sample large enough to detect a large effect size. However, even this study was not sufficiently powered to detect medium or small effect sizes. Furthermore, all other studies investigated a sample that included less than *n*=20 relapse cases. This is a major limitation of the studies included in this review.

This issue is potentially illustrated by the investigations of one potential predictor of relapse: initial levels of depressed mood. Fava et al. (2001b) had a sample that included *n*=31 relapse cases (i.e., large enough to detect a large effect size) and found this variable to significantly predict relapse. Meanwhile, Fava et al. (1995; i.e., the preliminary study of Fava et al., 2001b), Fava et al. (2001a), and Lincoln et al. (2005) found no significant effect of initial levels of depressed mood on relapse, when they investigated their samples of *n*=15, *n*=6, and *n*=12 relapse cases respectively (i.e., not large enough to detect a large effect size). This may indicate that Fava et al. (1995, 2001a) and Lincoln et al. (2005) found ‘false negatives’ for the effect of initial levels of depressed mood on relapse. However, it remains a possibility that Fava et al.’s (2001b) finding was a ‘false positive’. This can only be better understood with further replication using adequately powered samples.

One other potential reason why studies may not have consistently identified the same predictors to have a significant effect on the risk of relapse is the variety of relapse operationalisations used. As shown in Table 2.1, no two unique studies used the same measure for relapse, except for the studies conducted by Fava et al. (2001a, 2001b; reoccurrence of DSM disorder). This lack of consistency in definition likely had an effect on differences between studies in terms of relapse rates and identified significant predictors.

Another limitation that may explain the heterogeneity in results is that all of the reviewed studies were longitudinal cohort studies. Although these studies may allow for greater generalization of results as compared to randomised controlled trials, they also have less control of the therapeutic process and consequently of confounding factors. For example, none of the nine studies reported an assessment of CBT fidelity or therapist competency. It is thus not possible to know if the therapists in these studies adhered to CBT treatment protocols. Therefore, some studies may have had relatively poorly delivered therapy, and this may potentially have influenced the contrasts in results between studies.

An additional limitation that may be related to the prevalence of cohort studies in this review is that in some studies patients received further therapy during the follow-up period. Only two studies reported that no patients in their samples received additional therapy during follow-up (Fava et al., 2001a, 2001b). Four studies reported that some patients received additional treatment but did not account for this in their analysis, and three studies did not mention whether or not patients received further treatment. This limits the certainty with which these studies findings can be applied to acute-phase CBT, as their findings may have been influenced by the potential confounding variable of patients receiving continuation-phase interventions that likely influenced the maintenance of remission. Therefore, it is important that future studies take this factor into account, by routinely recording what other interventions are being received by patients following CBT. Furthermore, this limitation, along with lack of assessments related to CBT fidelity and therapist competency, highlights the need for more trial-based designs for the investigation of relapse of anxiety, as such designs are particularly suited *for controlling such confounding variables.*

**Limitations**

There were also limitations related to the methods of this systematic review itself. For instance, no studies were identified that investigated predictors of relapse of separation anxiety disorder, selective mutism, specific phobia, generalised anxiety disorder, or PTSD. Furthermore, meta-analyses were planned for this review, however a quantitative synthesis was only possible for one investigated predictor. The small number of published articles, many predictors not being investigated in multiple studies, and a lack of sufficient reporting of statistical information prevented more meta-analyses from being possible. A narrative review was conducted instead for the remaining investigated variables, and the subjective nature of such a review is less ideal than a quantitative synthesis of results. The small number of eligible studies also did not allow for subgroup (e.g., results by intervention, primary disorder) or moderator analyses (e.g., patients receiving continuation-phase interventions). Finally, this review was also limited by the exclusions of grey literature and studies published in languages other than English.

**Future Research**

Overall, this review has highlighted the limited research that exists regarding predictors of relapse of anxiety disorders, PTSD, and OCD. This is especially the case for separation anxiety disorder, selective mutism, specific phobia, generalised anxiety disorder, and PTSD, as no eligible studies that investigated relapse of these disorders were identified. The limited research in this area echoes the conclusions drawn by Wojnarowski et al. (2019) in their systematic review of predictors of relapse of depression. Clearly, there is currently a lack of understanding regarding relapse following CBT for both depression and anxiety. Further research that is adequately powered, alongside more trial-based designs, is urgently needed to address this gap in knowledge. Moreover, additional research is needed, particularly on the effect of residual anxiety symptoms on relapse, so that more robust meta-analyses can be conducted.

Many studies were considered ineligible for this review as they did not assess relapse as a follow-up outcome, despite being well-designed studies that assessed multiple predictors of long-term outcomes with complete and sufficiently long follow-up periods (e.g., Ogawa et al., 2010). As mentioned in Chapter 1, many studies of CBT typically assess long-term treatment response using continuous measures of symptom severity, and do not use categorical measures of relapse. This process, which involves the aggregation of mean symptom severity ratings at the group-level, obscures within-individual change and prevents the identification of relapse rates, thus making it difficult to explore individual differences in the long-term maintenance of treatment gains. Therefore, it is important for future research that investigates the long-term outcomes of CBT to do so with a measurement of relapse.

However, it is also important future studies are designed using a standardised, valid, and robust definition of relapse to ensure findings can be compared across studies. I believe there are concerns regarding the majority of the relapse definitions used in the studies in this review. For example, five of the nine studies used the CGI as a measure of relapse. Although this measure has been demonstrated to have concurrent validity and to be sensitive to change (Berk et al., 2008; Leon et al., 1993), it remains a subjective measure of a clinician’s perception of a patient’s current condition, that is based upon a comparison to other patients they have personally treated. This subjectivity raises concerns regarding the validity of this approach. Therefore, we propose an approach to assessing relapse that is similar to that used by Lincoln et al. (2005), which used validated self-report disorder-screening measures and the calculation of RCI. We recommend future research should classify a patient as having relapsed when: 1) their score on the validated measure increases above the measure’s diagnostic cut-off; and 2) the increase represents a statistically reliable and clinically significant deterioration in the patient’s condition (i.e., an increase greater than the measure’s RCI; Jacobson & Truax, 1991).

**Clinical Implications**

This review highlights the highly recurrent nature of anxiety, with an estimated 21.8% of patients relapsing after CBT. Considering this, it is therefore important that relapse prevention is a valued and fundamental component of the treatment process. Furthermore, it is important that relapse prevention strategies and interventions, such as booster sessions, continuation-phase CBT, and MBCT, are offered to patients who have received clinically successful acute-phase CBT. Although offering these interventions incurs a financial cost, the effective prevention of relapse has been argued to save money in the long-term, due to a reduction in the ‘revolving door’ phenomenon where patients continually return for further treatment (Scott et al., 2003; Wojnarowski et al., 2019). With improved knowledge of which patients are vulnerable to relapse, maintenance interventions may be targeted towards ‘at-risk’ patients. The presence of residual symptoms is emerging as a potential risk factor that could be used to target patients with increased vulnerability to relapse.

This review has also highlighted the limited research in the area of relapse after CBT, with the studies that are conducted being consistently underpowered. It is therefore important that larger datasets can be constructed so that further research can be undertaken in this area that is adequately powered. One action that can be undertaken to enable this is the increased incorporation of structured follow-up into clinical practice.

**Conclusion**

In summary, approximately one fifth of patients who completed CBT for anxiety-related disorders with remission of symptoms subsequently experienced a relapse. Yet, despite this concerning statistic, knowledge regarding what factors cause or influence relapse remains limited, with little research having been conducted in this area. Nevertheless, this review has highlighted the potential importance of residual symptoms as a prognostic indicator for the relapse of anxiety, adding to previous research that has demonstrated its value in the prediction of depressive relapse. However, further research is required before residual symptoms can be established as a robust risk factor of relapse of anxiety-related disorders. Other potential predictors, which could also be fruitful targets for future research, have also been identified, including the presence of a personality disorder, the degree of treatment improvement, and the occurrence of stressful life events. Most importantly, further research with adequately powered samples, and standardised measures and definitions of relapse is required. This will enable more risk factors to be discovered and established, and facilitate the development of evidence-based maintenance interventions targeted at those patients at greatest risk of relapse.

**CHAPTER 3**

**Dynamic Prediction of Cases at Risk of Relapse Following Low-Intensity Cognitive Behavioural Therapy using a Machine Learning Approach**

The first two chapters of this thesis have demonstrated the limited research that has been conducted in the area of relapse following cognitive behavioural therapy. The systematic review discussed in Chapter 2 extended upon the systematic review conducted by Wojnarowski et al. (2019), by further highlighting the significant limitations associated with research into relapse of depression and anxiety after CBT, and how little is currently known regarding factors that are associated with its occurrence. Despite 22 studies being identified across both reviews with over 150 variables being investigated as potential predictors, most studies were grossly underpowered to identify reliable predictors of relapse, especially in samples with a low event base-rate. Furthermore, the studies included in the two systematic reviews applied suboptimal analyses to perform variable selection in samples with multiple features, and to deal with multicollinearity when correlated features were examined as potential predictors of relapse.

Contemporary machine learning approaches could potentially help to advance our understanding of relapse risk factors, since they are explicitly designed to optimize variable selection and to enhance predictive accuracy by leveraging the prognostic signal across multiple “weak” predictors (Hastie et al., 2009). In *supervised* machine learning, patterns are identified in data with the goal of using input variables to predict the values of a target outcome measure (Hastie et al., 2009). Such models are often trained to maximize generalizability to new samples. In contrast to traditional statistical models, such as general linear model (GLM) approaches, supervised machine learning approaches focus on prediction rather than explanation (Yarkoni & Westfall, 2017). This focus provides certain advantages over GLM. For instance, the predictive accuracy of GLM equations is known to be limited when they are developed using relatively small samples and a large number of predictor variables (Iniesta et al., 2016). Furthermore, GLM approaches are prone to *overfitting* (Babyak, 2004; Yarkoni & Westfall, 2017). This refers to situations where a prediction model is overly influenced by the idiosyncrasies of the cases used to develop it (i.e., it incorporates noise that is unique to this sample), and is consequently unreliable at predicting outcomes in new samples (i.e., poor generalizability). Machine learning approaches address overfitting by implementing cross-validation (i.e., training models and then examining their predictive ability in test samples) and regularization methods (i.e., penalizing overly complex models; Hastie et al., 2009).

Machine learning approaches have recently been utilized to address a series of mental health related prediction problems. For example: the prediction of persistent depressive symptoms in older adults (Hatton et al., 2019); the diagnosis of post-traumatic stress disorder (PTSD) three months after a severe injury (Papini et al., 2018); response to pharmacological treatment of depression (Chekroud et al., 2016); and targeted prescription of alternative psychological treatments for depression (Cohen et al., 2020; Delgadillo & Gonzalez Salas Duhne, 2020).

Set against this backdrop of emerging applications of machine learning in mental health, this chapter will discuss an empirical study that aimed to apply a machine learning approach to (a) identify prognostic indicators of relapse after low-intensity cognitive behavioural therapy (LiCBT), and (b) to use this information to develop a relapse prediction tool that could be used to guide relapse prevention interventions in psychological services.

**Method**

**Design, Setting and Interventions**

This study analyzed data from the West Yorkshire Low-Intensity Outcome Watch (WYLOW) study (Delgadillo et al., 2018b), which was a naturalistic, prospective, longitudinal cohort study. *N*=439 LiCBT patients with remission of depression and anxiety symptoms after treatment were recruited into the study from a psychological therapy service in West Yorkshire, England, which was part of the Improving Access to Psychological Therapies programme (IAPT; Clark, 2011). Participants were followed-up on a monthly basis after treatment completion, for up to 24 months. The overall objective of the study was to quantify relapse and recurrence rates after routinely-delivered LiCBT. Approval for the study was obtained from the NHS Health Research Authority and an independent ethics committee (Ref: 12/YH/0095).

Participants in the WYLOW study completed low-intensity guided self-help interventions based on principles of CBT, which was consistent with national clinical guidelines and competency frameworks (National IAPT Team, 2015; National Institute for Health and Care Excellence, 2011). These interventions were delivered in one-to-one sessions, group settings, or via computerized CBT programmes with adjunct telephone support. LiCBT in this service was highly standardized, protocol-driven, and delivered by qualified psychological wellbeing practitioners (PWPs) who practiced under regular clinical supervision (weekly, or every other week).

LiCBT patients who completed treatment with sub-clinical depression and anxiety symptoms (see measures section) were eligible for inclusion and were recruited within one month of their last planned treatment session. Participants were contacted by independent researchers on a monthly basis and prompted to complete depression and anxiety questionnaires via email, telephone or postal survey (based on participants’ preferences). Participants remained in the study until they met one of three criteria: (1) they were classified as having relapsed (as defined below); (2) they had failed to respond to two consecutive monthly assessments (lost to follow-up); or (3) their responses indicated that they had remained in remission throughout the follow-up period. Participants that were classified as having relapsed were provided with self-help information and supported to re-engage with mental healthcare services. Further details about the study setting, design, recruitment process, data collection and primary findings are reported elsewhere (Ali et al., 2017; Delgadillo et al., 2018b).

Although the WYLOW study collected data for up to 24 months post-treatment, this study only analyzed data from the initial 12 month period. There were two reasons for this: 1) this study’s purpose was to predict relapse (<12 months), not recurrence (≥12 months); and 2) the inclusion of recurrence events in the analysis would have limited the number of cases that remained in remission throughout the follow-up period to a point where the development of predictive algorithms would have been unfeasible.

**Participants**

The present study sample analyzed data for a subsample of *n*=317 cases from the WYLOW study, excluding *n*=122 cases that were lost to follow-up, and for whom the actual outcome of interest (relapsed or maintained remission) was unknown. Within the first half of the follow-up period, *n*=84 participants had relapsed after one month, *n*=125 after two months, *n*=155 after three months, *n*=171 after four months, *n*=178 after five months, and *n*=189 after six months. Within the second half, *n*=196 cases had relapsed after seven months, *n*=202 after eight months, *n*=208 after nine months, *n*=214 after 10 months, *n*=218 after 11 months, and *n*=223 after one year. Therefore, a total of *n*=223 (70%) cases in the present study sample relapsed during the 12-month follow-up period, while *n*=94 remained in remission.[[5]](#footnote-5)

The three most common primary presenting problems recorded in clinical records for the above cases were mixed anxiety and depression (32%); depressive episode (22%); and generalized anxiety disorder (19%). Other mental health problems (e.g., panic disorder, obsessive compulsive disorder) were recorded as the primary problem for less than 3% of cases each (*n*<10). At time of assessment, the sample recorded a mean depression score of 13.93 (*SD* = 5.49) on the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), and a mean anxiety score of 13.35 (*SD* = 4.35) on the Generalized Anxiety Disorder scale (GAD-7; Spitzer et al., 2006) respectively (see below for measures). Meanwhile, at the end of treatment, the sample recorded mean PHQ-9 and GAD-7 scores of 3.55 (*SD* = 2.40) and 3.28 (*SD* = 2.12) respectively. In terms of demographics, the sample was characterized by a majority of female patients (59%) from a white British background (90%), with a mean age of 43 (*SD* = 14.8). Approximately 10% were unemployed at the start of treatment, 12% were unemployed at the end, 11% had a self-reported disability, and 26% had a self-reported long-term medical condition (e.g., asthma, diabetes, chronic pain, etc.). Patients accessed a mean of seven LiCBT sessions (*SD* = 2; range = 2-16).

**Measures**

*Patient Health Questionnaire (PHQ-9*; Kroenke et al., 2001): A screening tool for major depression containing nine items. Each item assesses how often a specific symptom is experienced over a two-week period and is measured on a 0-3 scale (i.e. 0 = “not at all”, 3 = “nearly every day”). Responses are summed to calculate an overall severity score (range = 0-27). A cut-off of ≥10 is recommended to detect clinically significant depression symptoms (Kroenke et al., 2001).

*Generalized Anxiety Disorder scale (GAD-7*; Spitzer et al., 2006): A screening tool for anxiety disorders containing seven items. Similar to the PHQ-9, each item assesses how often a symptom is experienced by over a two-week period and the same scale (i.e. 0-3) is used. Responses are summed to calculate an overall severity score (range = 0-21), and a cut-off of ≥8 is used for the detection of an anxiety disorder (Kroenke et al., 2007).

In order to monitor deterioration over time, a reliable change index of ≥5 for both the PHQ-9 (McMillan et al., 2010) and the GAD-7 (Richards & Borglin, 2011) was applied in this study.

*Work and Social Adjustment Scale (WSAS*; Mundt et al., 2002): A questionnaire that assesses the impact of a mental health problem on five life domains (work, home management, social life, leisure activities, family and relationships). The impact on each domain is measured on 0-8 scales (i.e. 0 = “no impairment”, 8 = “severe impairment”), and the five scores are summed to derive an overall score of functional impairment (range = 0-40).

A series of other clinical, demographic and treatment variables were also available. These are described in detail by Ali et al. (2017) and summarized in Table 3.1.

***Primary Outcome***

For a patient to be classed as having relapsed, the following two criteria must have been met: (a) at least one of the symptom measures (PHQ-9 or GAD-7) was above the clinical cut-off at a monthly follow-up review; and (b) the measure was indicative of statistically reliable deterioration relative to the score observed at the final treatment session (based on the reliable change index). Patients who scored above clinical cut-offs during the follow-up period (criterion a) but did not display statistically reliable deterioration (criterion b), were not classed as having relapsed and continued to be monitored on a monthly basis. Fourteen of the 94 participants who remained in remission after 12 months had met criterion a but not criterion b at least once during the follow-up period.

|  |  |  |
| --- | --- | --- |
| **Table 3.1**  *The Potential Predictors Input into Each XGBoost Model* | | |
| Model | N Input Predictors | Explored Predictors |
| 1 –  Baseline | 17 | Age; gender; ethnicity; unemployment at start; taking medication at start; long-term condition; neighbourhood deprivation; disability; diagnosis; chronicity of problem; family history; previous treatment episodes; expectancy of treatment; responded to treatment expectancy question; baseline PHQ-9; baseline GAD-7; baseline WSAS |
| 2 –  End of Treatment | 32 | *Model 1 variables* *plus*…  …unemployment at end; taking medication at end; PHQ-9 at end; GAD-7 at end; WSAS at end; number of treatment sessions; early treatment response PHQ-9; early treatment response GAD-7; early treatment response WSAS; linear treatment response (GoF) – PHQ-9; linear treatment response (GoF) – GAD-7; linear treatment response (GoF) - WSAS; linear treatment response (slope) – PHQ-9; linear treatment response (slope) – GAD-7; linear treatment response (slope) - WSAS |
| 3 –  FU1 | 39 | *Model 2 variables plus…*  …PHQ-9 at FU1; GAD-7 at FU1; WSAS at FU1; PHQ-9 change from end to FU1; GAD-7 change from end to FU1; WSAS change from end to FU1; responded at FU1 |
| 4 –  FU3 | 53 | *Model 3 variables plus…*  *…*PHQ-9 at FU2; GAD-7 at FU2; WSAS at FU2; PHQ-9 change from FU1 to FU2; GAD-7 change from FU1 to FU2; WSAS change from FU1 to FU2; responded at FU2; PHQ-9 at FU3; GAD-7 at FU3; WSAS at FU3; PHQ-9 change from FU2 to FU3; GAD-7 change from FU2 to FU3; WSAS change from FU2 to FU3; responded at FU3 |
| Abbreviations: FU1, first month of follow-up; FU2, second month of follow-up; FU3, third month of follow-up; GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit for a regression model; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale. See Appendix D for glossary for predictors. | | |

**Analysis**

A supervised machine learning approach was applied to develop an ensemble of four prognostic models that aimed to predict relapse in a dynamic (longitudinal) way during the treatment and follow-up phases. Each successive model was trained using additional information that became available over time, thus enabling the overall ensemble of algorithms to “learn” to “adjust” the prognosis in four steps. Model 1 used information only known at the time of pre-treatment assessments; Model 2 learned from information known at the final treatment session: Model 3 included information available one month after the end of treatment; and Model 4 included information available three months after the end of treatment. These models were trained to output a predicted classification for each patient (1 = *will relapse within 12-months of completing treatment*; 0 = *will have sustained remission for up to 12 months*), at each relevant time-point described above. Importantly, the dynamic nature of this overall modelling approach means that a case predicted at one time-point (e.g., last treatment session) to maintain remission could later be classed as at risk of relapse based on newly available information (e.g., if depression symptoms increase at 3-months follow-up).

No models were developed that used information collected after three months of follow-up as predictors of relapse, as these models would have been too underpowered due to the gradual decrease of the overall sample during the follow-up period. Indeed, due to this gradual decrease, Models 3 and 4 (i.e., predictions occurring during follow-up) were developed using smaller overall samples than Models 1 and 2 (i.e., predictions occurring before follow-u). The variables tested as potential predictors in each model are summarized in Table 3.1, and further described in Appendix D (glossary of terminology). The potential predictors ranged from baseline characteristics to factors related to treatment response.

***Machine Learning Approach***

We applied extreme gradient boosting (XGBoost; Chen & Guestrin, 2016), as implemented in the R package *xgboost* (Chen et al., 2019; version 0.82.1), to develop the four prognostic models. The XGBoost package implements gradient-boosted decision trees (Friedman, 2001). Gradient boosting is an ensemble machine learning approach, meaning that it combines several simple decision trees, that are each individually poor at predicting an outcome, to develop one collective, stronger prognostic model. Specifically, when each successive decision tree is trained, subsequent trees learn from the ‘residuals’ or ‘errors’ from the previous modelling attempt. Therefore, each gradient boosting model is developed iteratively in multiple steps, with each step placing more priority in making accurate predictions for those individuals for whom prior iterations made poor predictions.

Each single decision tree consists of ‘branches’, which represent logical structures. Each study participant has a ‘path’ through a decision tree that is determined by their data; for dichotomous features (e.g., employed/unemployed), paths diverge for yes/present or no/absent responses, while for continuous features (e.g., baseline depression severity), paths diverge based on empirically derived cut-offs. Different cut-offs are examined by the algorithm until a final cut-off is selected based on the optimization of predictive accuracy. Each path terminates at a ‘leaf’, which represents a predicted probability weight for the outcome of interest. The final predicted classification (relapse/sustained remission) for each case is derived by summing their individual probability weights across all decision trees in the ensemble. An example decision tree, extracted from a model developed in this study, is illustrated in Figure 3.1.

XGBoost has several advantages over conventional regression modelling (Chen & Guestrin, 2016). For instance, missing data do not need to be imputed or excluded, as the algorithm can use missing data as an informative splitting criterion. Moreover, continuous variables do not need to be transformed, as path divergences are established based on cut-offs determined by the algorithm. Gradient boosting allows for complex, non-linear interactions among categorical and continuous variables to be incorporated into predictive models. In addition, XGBoost can also minimize overfitting through the implementation of L1 (LASSO) and L2 (Ridge) regularization procedures, and cross-validation loops.

**Parameters.** For the development of each of the four models, hyperparameters were set for the implementation of XGBoost. One parameter, *eta*, controls the learning rate of the algorithm, and can minimize overfitting by shrinking the weights of variables after each boosting step, thus making the boosting process more conservative. *Eta* can range from 0-1, and was manually set in this study at 0.1. Another parameter manually set was the *maximum depth* of a tree, which was set at 3. This meant that during the development of each individual decision tree, there can only be a maximum of two further ‘branch splits’ along a path after the initial split. Trees with larger depths are more complex and allow models to learn relations that are highly specific to a particular sample. Therefore, having a smaller maximum depth for trees minimizes overfitting. We used the XGBoost default values for regularization,where *alpha* (L1) is set to 0 and *lambda* (L2) is set to 1. Therefore, no LASSO regularization was applied in the development of predictive models, while a Ridge penalty was applied.

**Figure 3.1**

*Example Decision Tree*

Unemployed at Start

Age

Family History

Probability +0.11

Probability +0.07

Probability +0.001

Probability -0.02

Baseline WSAS

Probability -0.14

*Note.* This decision tree, extracted from Model 1, has a maximum depth of three variables. Rectangles represent ‘branch splits’ (i.e. variables), ovals represent path decisions, and hexagons represent ‘leaves’ (i.e. outcome probability weights).

The evaluation metric used to assess the predictive ability of models was the area under the receiver operating characteristic curve (AUC). This measure assesses how capable models are at distinguishing between binary outcome classes (Bradley, 1997). The metric ranges from 0-1, with 1 representing a perfect predictive model and 0.5 indicating that predictions are no better than chance. The AUC statistic has been recommended as an appropriate evaluation metric for healthcare research, due to the clinical utility of model predictions being more important than statistical power in these contexts (Moons et al., 2015).

**Internal Cross-Validation.** Internal cross-validation loops were used to develop an XGBoost ensemble model. To achieve this, the overall sample was partitioned into five subsamples using stratified randomization so that each subsample had approximately the same rate of relapse cases. Each loop would use 4/5 of the data as a training sample and 1/5 as a test sample. Each subsequent decision tree was trained using this five-fold cross-validation process, which continued until a ‘stopping rule’ was triggered if the addition of new trees resulted in overfitting with reference to the AUC observed in the test sample. The cross-validation process therefore identified the highest number of decision trees an XGBoost model can incorporate before overfitting to the dataset.

**Variable Selection.** Implementation of XGBoost produces a model with: (1) a final AUC statistic that assesses its ‘out-of-sample’ performance; and (2) information regarding the variables that the model has selected for use in making its predictions. It is desirable to train a model with the highest possible AUC index, and with the lowest number of variables selected (i.e., a more parsimonious model). If superfluous variables are made available to the algorithm, then given the large number of decisions the model must take, it will occasionally happen across a subset of data for which these extra variables appear to be predictive but do not extend to the test set. Thus, we can improve the model by iteratively testing whether removing a variable from the dataset improves the AUC index.

We therefore performed variable selection through the additional incorporation of ‘leave-one-variable-out loops’. One loop involved additional XGBoost models to be developed, with the same number of models being developed as there were variables input into the base model (*k*). For each of the models produced in this step of model development, one different variable was excluded from analysis and therefore not incorporated into the model (i.e., *k*-1). The model with the highest associated AUC statistic was considered the best model in this step of the process, and kept for further development. Following this, another leave-one-out loop was implemented using the variables included in the best model from the previous loop, and this process continued until there was only one variable remaining. The model with the highest AUC across all of the loops was considered the best overall model developed. This incorporation of leave-one-out loops ensured that the final selected model had the highest AUC possible with the lowest number of variables involved.

**Model Evaluation.** The analytical process described thus far was followed to develop each of the four prognostic models (Models 1-4). Using each of these models, predicted probabilities of relapse for each patient were calculated at four time-points during their pathway through treatment and follow-up. Probabilities ranged from 0-1 with higher scores indicating greater probability of relapse, and a patient was predicted to relapse when their calculated probability was greater than 0.5. Conventional performance metrics were calculated to assess the predictive value of each model when tested out-of-sample (in the 1/5 test set): accuracy; sensitivity; specificity; positive predictive value (PPV); and negative predictive value (NPV).

**Exploring Individual Predictors.** Before individually exploring the predictors identified by each model, a sensitivity analysis was conducted in which the four final models were reanalyzed with the application of L1 regularization (*alpha* = 0.5). In contrast to L2 regularization, L1 regularization is capable of shrinking the coefficients of variables to be exactly zero, thus eliminating likely irrelevant variables from the model (Hastie et al., 2009). Therefore, this sensitivity analysis was conducted to explore if the addition of a LASSO penalty identified any variables in the models as being spurious or potential false positives, and thus ensure that only the most important individual predictors were explored.

Following this, the variables that were (1) determined to be important by at least one of the four final XGBoost models, and (2) not deemed spurious by the addition of L1 regularization, were explored in terms of their associations with relapse using relative importance metrics, partial dependence plots, and descriptive statistics. Relative importance is a metric that indicates how useful a predictor was in the construction of boosted decision trees within a model, relative to the other predictors used by the model. It is calculated for each predictor by assessing the degree to which a model’s predictive ability improves when key decisions within decision trees use that specific variable (Hastie et al., 2009). Meanwhile, partial dependence plots demonstrate the relationship between a predictor and the outcome by aggregating all of the decisions in all of the decision trees in the model which select on that predictor (Friedman, 2001). Therefore, they illustrate the relationship that the XGBoost model judges is directly attributable to that predictor, after partialling out the effects of other predictor variables, even if they might be correlated with the predictor of interest.

**Results**

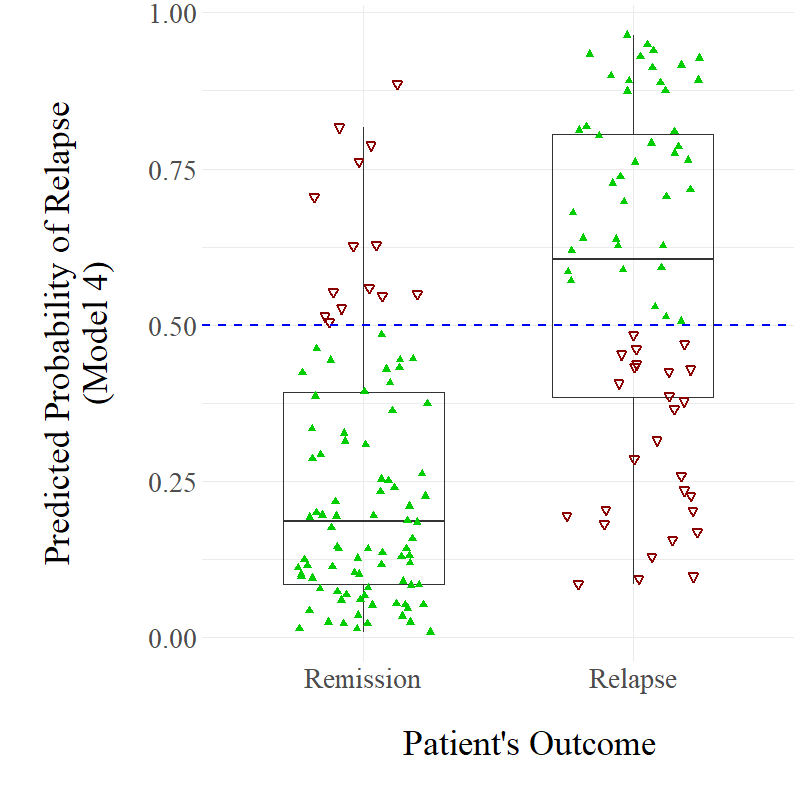
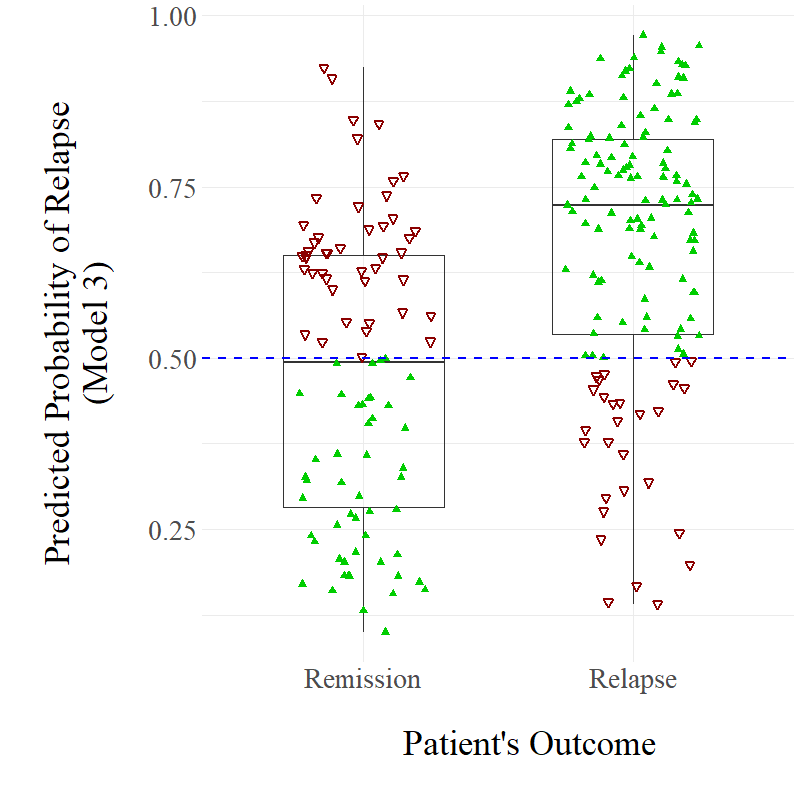
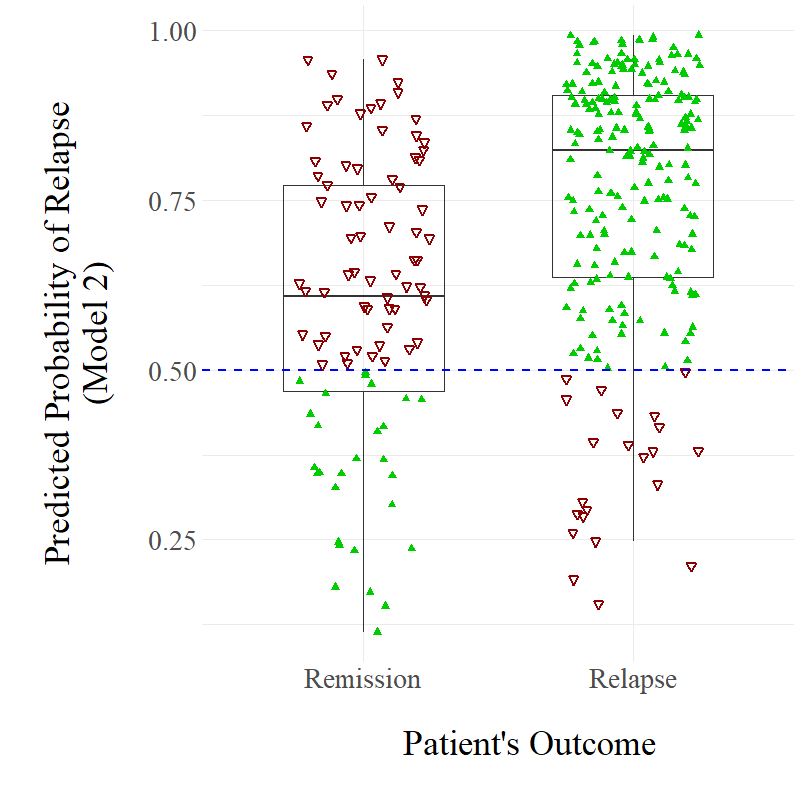
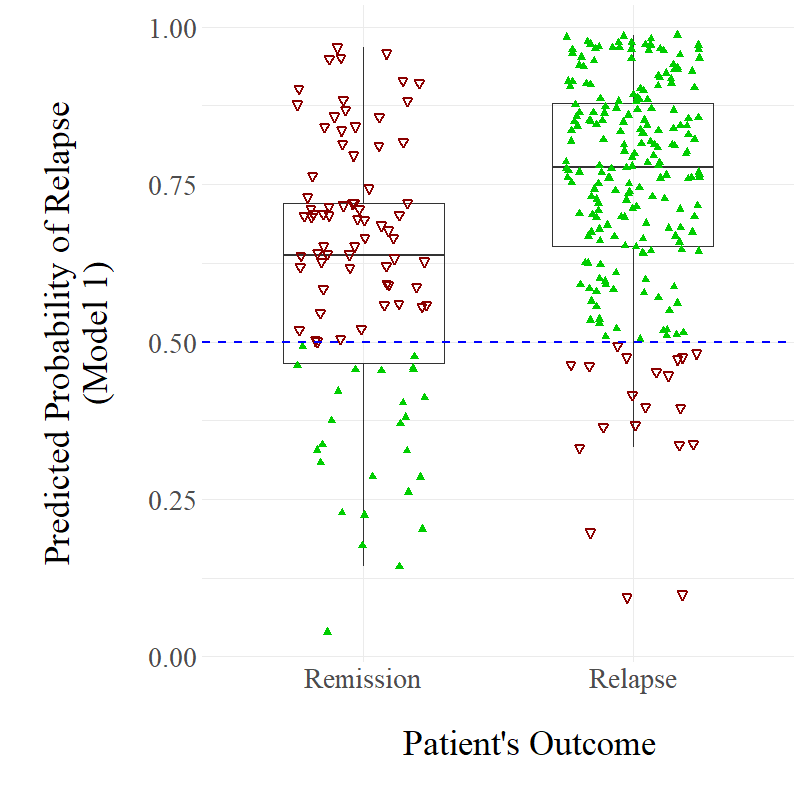
**Performance of XGBoost Models**

The predictive value of the four models when evaluated on the test set is summarized in Table 3.2[[6]](#footnote-6). The AUC became larger as more information was available to the models over time. Model 1, which only used pre-treatment assessment information, had the lowest AUC (0.70), while Model 4, which used information available up to the third month of post-treatment follow-up, had the highest AUC (0.83). Other metrics of predictive value displayed in Table 3.2 indicated that Models 1 and 2 had highly similar performances. Both models had high sensitivity (91% and 90.1% respectively), PPV (74.9% and 75.3% respectively), and accuracy (both 72.2%). However, both models also had low specificity (27.7% and 29.8% respectively) and NPV (56.5% and 56% respectively). Model 3 was found to have greater specificity (52.1%) and NPV (63.6%), however it also had the lowest accuracy (68.7%) and PPV (71.2%) of all four models, and the second lowest sensitivity (79.9%). Overall, Model 4 was found to have the best relative balance in performance indices across all four models. The distributions of the probabilities of relapse predicted by each model for each patient can be seen in Figure 3.2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3.2**  *The Ability of Each Model to Predict Relapse When Evaluated on the ‘Test’ Set* | | | | |
|  | Model 1 –  Baseline | Model 2 –  End of Treatment | Model 3 –  FU1 | Model 4 –  FU3 |
| Overall Sample (N)\* | 317 | 317 | 233 | 162 |
| Relapse Cases (n) | 223 | 223 | 139 | 68 |
| Number of Decision Trees in Model | 74 | 84 | 42 | 70 |
| AUC | 0.70 | 0.73 | 0.76 | 0.83 |
| Accuracy | 72.2% | 72.2% | 68.7% | 74.7% |
| Sensitivity | 91.0% | 90.1% | 79.9% | 60.3% |
| Specificity | 27.7% | 29.8% | 52.1% | 85.1% |
| PPV | 74.9% | 75.3% | 71.2% | 74.6% |
| NPV | 56.5% | 56.0% | 63.6% | 74.8% |
| Abbreviations: AUC, area under the receiver operating characteristic curve; FU1, first month of follow-up; FU3, third month of follow-up; PPV, positive predictive value; NPV, negative predictive value  \*Overall sample decreases for Models 3 and 4 due to some participants having already relapsed before their respective FU month. | | | | |

**Important Predictors**

Of the 141 variables input as potential predictors of relapse across all four models, a total of 42 predictors were deemed important. Model 2 was the most complex, selecting 13/32 input variables as important predictors. Models 3 and 4 each identified 11 predictors as important, with 39 and 53 variables initially being entered into each model respectively. Finally, Model 1 was the simplest, selecting only seven of the 17 variables initially entered into the model. The predictors identified as being important by each model, and the relative importance of each variable for the prediction of relapse, are displayed in Table 3.3.



a)

b)

c)

d)

**Figure 3.2**

*Distributions of Predicted Patient Relapse Probabilities made by a) Model 1, b) Model 2, c) Model 3, and d) Model 4*

*Note.* Filled, green triangles represent correct predictions for specific cases; unfilled, red upside-down triangles represent incorrect predictions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3.3**  *The Predictors Identified as Important by Each Developed Model (Relative Importance %)* | | | |
| Model 1 –  Baseline | Model 2 –  End of Treatment | Model 3 –  FU1 | Model 4 –  FU3 |
| Age (40%) | Linear treatment response (GoF) –  WSAS (20%) | Age (25%) | GAD-7 at FU3 (12%) |
| Baseline WSAS (23%) | Age (17%) | WSAS at FU1 (16%) | Linear treatment response (GoF) –  GAD-7 (12%) |
| Unemployment at start (15%) | Linear treatment response (GoF) –  GAD-7 (11%) | GAD-7 at FU1 (14%) | Linear treatment response (GoF) –  WSAS (11%) |
| Disability (8%) | Early treatment response PHQ-9 (9%) | WSAS change from end to FU1 (9%) | PHQ-9 change from FU1 to FU2 (11%) |
| Family history (7%) | WSAS at end (8%) | PHQ-9 change from end to FU1 (9%) | Age (11%) |
| Taking medication at start (5%) | Unemployment at start (7%) | Unemployment at start (7%) | WSAS change from end to FU1 (10%) |
| *Responded to treatment expectancy question (1%)\** | PHQ-9 at end (7%) | Taking medication at start (7%) | Chronicity of problem (9%) |
|  | Baseline GAD-7 (5%) | Family history (5%) | PHQ-9 change from end to FU1 (8%) |
|  | Unemployment at end (5%) | Gender (3%) | WSAS change from FU2 to FU3 (6%) |
|  | Previous treatment episodes (4%) | Disability (2%) | Early treatment response PHQ-9 (6%) |
|  | Taking medication at start (3%) | Diagnosis (2%) | Taking medication at start (5%) |
|  | Expectancy of treatment (3%) |  |  |
|  | Gender (<1%) |  |  |
| Abbreviations: FU1, first month of follow-up; FU2, second month of follow-up; FU3, third month of follow-up; GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit for a regression model; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale. See Appendix D for glossary for predictors.  \*Sensitivity analysis (reanalysis with L1 regularization) did not select this variable. | | | |

The sensitivity analyses (i.e., re-analysis of the final models with L1 regularization applied) yielded models with similar but slightly lower cross-validated AUC indices. They also retained the same number of important predictors for Models 2, 3, and 4. However, the sensitivity analysis for Model 1 did not select the variable denoting missing vs. complete response to the *treatment expectancy* variable as an important predictor, indicating that this variable is likely to be spurious. Therefore, this variable was not explored further. For the sake of parsimony, the important predictors can be grouped according to four categories: demographics; baseline clinical features; treatment process features; and residual symptoms. Partial dependence plots can be found in Appendix E.

***Demographics***

Demographic predictors leveraged a combined importance (i.e. the sum of relative importance metrics of each predictor of the same type) of 63% in Model 1, however this decreased when subsequent models had access to further information. Demographic predictors only had 30% combined importance for Model 2, 37% for Model 3, and 11% for Model 4. Younger age, unemployment (pre-treatment and post-treatment), disability, and female gender featured as important predictors across at least two models (the latter two features had relatively little importance). Unemployment was a particularly important risk factor, with all 31 participants who were unemployed before starting treatment relapsing within 12 months, and 37/39 participants who were unemployed at the end of treatment relapsing.

***Baseline Clinical Features***

Pre-treatment clinical features had a combined importance of 37% for Model 1, 15% for Model 2, and 14% for both Models 3 and 4. The following risk factors were selected as important by at least two models: use of psychotropic medication at the start of treatment, and family history of mental health problems. Both features had comparatively small importance indices. Six other baseline clinical features were deemed important by only one model each: higher baseline WSAS severity, higher baseline GAD-7 severity, previous treatment episodes, low expectancy of treatment, diagnosis of an affective or an anxiety disorder (*vs* a ‘mixed’ or ‘other’ diagnosis), and a chronic mental health problem.

***Treatment Process Features***

Treatment process features were only relevant to Models 2-4, since Model 1 only included pre-treatment information. The combined importance of these features was 40% in Model 2, 0% in Model 3 and 11% in Model 4. The following risk factors were selected as important in two models: early treatment response (higher improvement by session 3) in the PHQ-9 measure, a more linear (versus nonlinear) trajectory of improvement in the WSAS measure during treatment, and a more nonlinear (versus linear) trajectory of improvement in GAD-7. However, early treatment response had less importance compared to the latter two features, while the findings for the GAD-7 measure were relatively inconsistent and less clear from the partial dependence plots than the findings for the WSAS measure.

***Residual Symptoms***

The combined importance of residual symptoms (sub-threshold symptom scores close to the diagnostic cut-offs) post-treatment was only 15% for Model 2, however this increased to 48% and 47% for Models 3 and 4 respectively. Higher post-treatment scores and increases in PHQ-9, GAD-7 and WSAS scores between the last treatment session and follow-up assessments were selected as important predictors across at least two models.

**Discussion**

This study demonstrates the utility of a data-driven prognostic method capable of identifying cases at high risk of relapse after the completion of LiCBT. Previous studies have used traditional statistical models to predict relapse after routinely delivered CBT interventions (see review by Wojnarowski et al., 2019). The present study represents a novel and considerable advance in this line of research through the application of a machine learning ensemble of models that dynamically predict and adjust expected prognoses at multiple time-points during the course of a patient’s treatment journey. The AUC statistics for these models ranged from 0.70-0.83, with predictive accuracy tending to improve over time, when models were trained using more information. The best performance (AUC = 0.83) was observed for Model 4, which used information collected before treatment, at the end of treatment, and up to the third month of post-treatment follow-up. This is likely to be explained by the use of the maximum number of available predictors, but also due to the temporal proximity of the prediction time-point to the target outcome. A close examination of performance metrics across these models indicated that positive predictive values were superior to negative predictive values. Put simply, these models performed well at identifying cases at high risk of relapse, but they were less capable of accurately identifying patients that would maintain remission of symptoms during the 12-month follow-up period. However, this was not the case for Model 4, which had both high PPV (74.6%) and high NPV (74.8%).

In terms of predictor categories, demographic and baseline clinical features exhibited high importance in the early models. However, this decreased in later models when additional variables related to the therapeutic process and residual symptoms were also included. Residual symptoms had by far the highest combined importance of the four categories in Models 3 and 4, having the greatest temporal proximity to the target outcome. In contrast, although therapeutic process features had high combined importance in Model 2, they had no importance in Model 3, and little importance in Model 4. This inconsistent pattern may indicate limited robustness of the findings related to these features, and thus further research is required.

Specific patient features that were most informative for the prediction of relapse included younger age, unemployment, treatment response trends (linear versus nonlinear), and the presence of residual symptoms. Young age was identified as a risk factor, contrary to prior studies in this area (Evans et al., 1992; Lincoln et al., 2005; Heldt et al., 2011). However, previous studies had less than *n*=20 relapse cases in their samples, indicating that they were grossly underpowered to detect even a large effect size. Indeed, a relatively larger study (*n*=31 recurrence cases) conducted by Fava et al. (2001b), found a significant effect of young age on recurrence of panic disorder with agoraphobia after exposure-based treatment.

Unemployment was a highly important risk indicator in the current sample, with all 31 participants who were unemployed at the start of treatment being classed as having relapsed within 12 months. In addition, the majority of these relapse cases occurred in the first month of follow-up, indicating that the time-to-relapse was accelerated for unemployed patients. Interestingly, cases that were unemployed before treatment did not entirely overlap with cases unemployed at the end of treatment, yet unemployment at both time-points was identified as a risk factor. This indicates that any experience of unemployment during the therapy process increases a patient’s risk of relapse. This adds further credibility to the wealth of evidence concerning the harmful effects of unemployment on mental health (Waddell & Burton, 2006).

Patterns of change during treatment were also informative for relapse prediction. Patients who displayed a rapid reduction of depression symptoms during the first 3 sessions had higher risk of relapse. Although early response is a well-established predictor of end-of-treatment outcomes in psychotherapy, a recent systematic review suggested that this could be at least partly influenced by a quasi-placebo effect, since rapid response is also evident in placebo control groups (Beard & Delgadillo, 2019). It is, therefore, possible that patients with early response may mostly be improving due to a *remoralisation effect* (Howard et al., 1993), but remain at risk of relapse because their learning of coping strategies has not been sufficiently developed or consolidated. A linear improvement in functional impairment over the course of treatment also appeared to predict relapse. Previous studies on the relationship between treatment duration and clinical outcomes suggest that domains such as interpersonal and social functioning tend to improve at later stages of treatment (see review by Robinson et al., 2019). Therefore, a gradual and linear improvement in functioning in relatively brief interventions (less than 8 sessions) possibly indicates that such changes may possibly be due to the temporary resolution of life problems and current stressors rather than therapy, leaving patients at risk of relapse when future stressors arise if they have not had an opportunity to consolidate coping skills. Findings concerning trajectories of anxiety symptoms were mixed and therefore an interpretation of these patterns would be premature at this stage and warrants further investigation.

The identification of residual symptoms after treatment as an important predictor of relapse is consistent with previous research (Bockting et al., 2015; Wojnarowski et al., 2019). As discussed in the previous chapter, one proposed explanation is that residual symptoms represent the persistence of the underlying disorder, albeit in a milder form (Paykel, 2008). Meanwhile, another possibility is that the presence of residual symptoms is associated with a greater vulnerability to experiencing future life events as intolerably stressful and de-stabilizing, thus increasing the likelihood of relapse (Harkness et al., 2014).

**Limitations and Future Research**

A limitation concerning the study sample concerns *class-imbalance* (i.e., the number of observations belonging to each outcome class was not the same), since relapse cases in this sample outnumbered remission cases (223 versus 94). Class-imbalance can undermine the accuracy of machine learning algorithms, with poorer predictive accuracy for patients in the minority class (Lopez et al., 2013). Indeed, Models 1 and 2, which were both developed using the total class-imbalanced sample, were highly effective at classifying relapse cases correctly (high PPV), but less capable of classifying remission cases accurately (low NPV). In contrast, Models 3 and 4, which were developed using smaller, more balanced samples, had an improved ability to accurately predict remission cases (higher NPVs). This pattern (Figure 3.2) may be influenced by class-imbalance bias. Therefore, Model 4 being superior at prediction compared to the other models may not be because prediction of relapse is most accurate during follow-up, but rather an artifact of class-imbalance.

One potential solution for attenuating class-imbalance bias is to adopt resampling techniques. One such technique is over-sampling, in which the skewed distribution is addressed by generating new minority class observations (e.g. synthetic minority over-sampling technique [SMOTE]; Chawla et al., 2002). A different technique is under-sampling, in which observations from the majority class are discarded instead (e.g. Random Undersampling; Tahir et al., 2009). The technique that is most appropriate may depend on the form of class-imbalance exhibited by a sample (Haixiang et al., 2017). Future replications of this study that also use class-imbalanced samples should consider using resampling techniques to limit this bias and improve model development. In the current study, we chose not to apply such methods to minimize modelling complexity since the primary goal was to test the accuracy of dynamic prediction models, but also because it is instructive to show the effects of class-imbalance so as to inform future research.

Another important limitation is that this study only applied an internal cross-validation loop, with the models being evaluated on subsamples that were also involved in the training of the models. Although internal validation can provide an indication of how models may generalize to new samples, a more robust method to determine out-of-sample generalizability is to apply an additional external validation in a statistically independent sample that was not involved in model development (Steyerberg et al., 2003). This can be done using a ‘holdout’ subsample of the overall original sample, which was not possible in this study due to the limited sample size, and/or using a sample of participants from a different population altogether (but for whom the same variables are available). Applications of machine learning algorithms using this rigorous external cross-validation process in mental healthcare research have shown that out-of-sample predictions tend to be valid and clinically useful, but less accurate in new samples (Chekroud et al., 2016; Delgadillo & Gonzalez Salas Duhne, 2020; Delgadillo et al., 2017; Leighton et al., 2019).

There were additional limitations related to the population investigated and data collection methods used. For instance, loss-to-follow-up in the primary study sample meant that data from 122 participants were not taken into account in the development of the present models. Furthermore, participants only received LiCBT interventions, so these models may not apply to patients who receive high-intensity CBT. In contrast to high-intensity CBT, LiCBT interventions are brief, highly manualized, non-specialist, and driven by the use of highly structured psychoeducational workbooks (Bennett-Levy et al., 2010). Due to these differences, it is plausible that there are differences between the two therapeutic formats in terms of the processes that underlie relapse. Therefore, further research needs to explore any potential differences between LiCBT and high-intensity CBT in terms of predicting relapse, and in terms of the individual risk factors associated with relapse. Another limitation of this study was the lack of an assessment of treatment fidelity or competency, thus making it impossible to know if PWPs had adhered to LiCBT treatment protocols. A recent study has developed and validated two measures for the purposes of assessing LiCBT competencies (Kellett et al., 2021). These measures could be applied in future research.

Future research into the prediction of relapse should also investigate predictors that were not considered in this study. For example, previous research has indicated that the experience of stressful life events during follow-up increases the risk of relapse for both depression and anxiety (Harkness et al., 2014; Heldt et al., 2011). Use of experience sampling methods using mobile-phone and passive sensing technology could be fruitful to track life events and/or physiological and subjective responses to such events.

**Clinical Implications**

Similar to the review discussed in Chapter 2, this study highlights the need for relapse prevention to be a core component of low-intensity CBT interventions. Although relapse prevention is considered an important aspect of LiCBT (Rodgers et al., 2012), the lack of treatment fidelity measures being applied means we do not know if this essential component of treatment is being delivered. Furthermore, the development of prognostic algorithms using machine learning approaches potentially allows for patients at risk of relapse to be identified and targeted with maintenance interventions. If such algorithms were to be implemented in psychological services, it is important that models can effectively identify both patients who are vulnerable to relapse, and those who are not. This is to ensure that the prescription of these interventions is provided to patients that are in need of them, but also performed in a cost-effective manner (i.e., not offered to every patient who successfully responds to treatment). Therefore, Models 1 and 2 may be inappropriate for application in clinical contexts due to these models possessing low NPVs, although these models could potentially be improved with further research that utilizes larger, class-balanced samples. In contrast, Model 4 would be the most appropriate model for use in clinical contexts, due to it possessing a good balance between high PPV and NPVs. However, Model 4 is limited by not being able to predict relapses that occur in the first three months of follow-up. Model 4 possessing the greatest predictive power may indicate that follow-up reviews improve our ability to understand and predict relapse. Incorporating structured follow-up into clinical practice would allow for prognostic algorithms such as Model 4 to be applied by clinicians, while also enabling larger datasets to be constructed and thus allow more improved predictive models to be developed.

**Conclusions**

This study demonstrates that it is possible to identify cases at high risk of relapse using routinely available data, with considerable accuracy even before the start of treatment, and predictive accuracy improving as new information becomes available over time. Dynamic prognostic systems such as the one presented in this chapter could be used in routine care to identify ‘high risk’ cases that could be offered evidence-based relapse prevention interventions in a targeted way, thus making best use of limited resources in publicly funded mental health services.

**CHAPTER 4**

**Predicting Cases at Risk of Relapse Following High-Intensity Psychotherapy using a Machine Learning Approach**

The previous chapter demonstrated the potential in applying a machine learning approach to dynamically predict relapse of depression and anxiety following low-intensity interventions based on cognitive behavioural therapy principles (LiCBT). The developed prognostic models displayed promising predictive accuracy, and identified a number of potential predictors of relapse that could be further explored in future research. However, one limitation of the previous chapter’s study was its exclusive focus on exploring relapse following LiCBT. Consequently, the developed models and associated potential predictors may not apply to patients who receive high-intensity psychological interventions (cognitive behavioural therapy (CBT) or other high-intensity interventions). For instance, LiCBT interventions are brief, highly structured, self-help oriented, and driven by the use of standard psychoeducational workbooks (Bennett-Levy et al., 2010). In contrast, high-intensity interventions involve primarily weekly, longer-lasting treatment sessions with a suitably trained therapist, which are often delivered on a one-to-one basis (Clark, 2018). Due to these differences, it is plausible that there are differences between the two therapeutic formats in terms of the processes that underlie relapse. Further research is therefore needed that explores any potential differences between low-intensity and high-intensity psychological interventions in terms of predicting relapse, and in terms of the individual risk factors associated with relapse. In addition, there may be differences between high-intensity treatment modalities (e.g., CBT vs non-CBT-based interventions) in terms of relapse rates. As mentioned in Chapter 1, although recent research indicates that relapse may be relatively common following non-CBT psychological interventions such as interpersonal therapy (IPT; Lemmens et al., 2019), research into relapse following non-CBT interventions has been relatively limited compared to research into relapse following CBT. Consequently, further research is therefore required to explore relapse following non-CBT interventions.

Considering the above limitations, this chapter will discuss a third empirical study in which a machine learning approach was applied to predict relapse following high-intensity psychological interventions. The overall aim of this study was to investigate the occurrence of relapse after high-intensity interventions, with there being four specific research objectives: (1) to estimate the rate of relapse following high-intensity interventions generally; (2) to compare relapse rates following high-intensity CBT and non-CBT high-interventions specifically; (3) to develop a predictive model capable of identifying cases at risk of relapse following high-intensity psychotherapy; and (4) to explore specific risk factors that may be potentially associated with relapse.

**Method**

**Design, Setting and Interventions**

This study was a retrospective observational analysis of quantitative outcomes for patients who completed a high-intensity psychotherapy intervention for depression and/or anxiety with remission of symptoms. The data were collected between April 2019 and May 2021 by two psychological therapy services within the Improving Access to Psychological Therapies programme (IAPT; Clark, 2011), which are managed by South West Yorkshire Partnership NHS Foundation Trust. Approval for the study was obtained from the NHS Health Research Authority and an independent ethics committee (Ref: 20/NE/0275). The services in this study offered follow-up review appointments to some patients who were discharged after receiving a psychological intervention[[7]](#footnote-7). These reviews were offered every three months for one year after treatment completion. Reviews were conducted either by phone or in-person, and involved the therapist providing therapeutic support to the patient.

As outlined in Chapter 1, IAPT uses a ‘stepped care’ service delivery system that implements National Institute for Health and Care Excellence (NICE; 2011) guidelines for the treatment of depression and anxiety (Clark, 2011). In stepped care, patients are initially offered low-intensity guided self-help interventions (i.e., those investigated in Chapter 3), and if patients do not respond to this initial step or their risk status changes then they are subsequently offered high-intensity psychological interventions. The first-line treatments within IAPT are CBT-based low- and high-intensity interventions, however other high-intensity interventions (e.g. Counselling for Depression, IPT etc.) are also offered in some services (Baker, 2018; Health Education England, 2020). High-intensity CBT is delivered by qualified and accredited cognitive behavioural therapists following disorder-specific and protocol-driven CBT models recommended in the CBT for anxiety and depression competency framework (Roth & Pilling, 2008). CBT therapists work under regular clinical supervision (equivalent of one hour per full-time week) and their competency is regularly assessed using the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001).

**Data Management and Sample Characteristics**

A total of *N*=19138 individuals were referred to the IAPT services between April 2019 and May 2021, with *n*=7514 receiving an episode (i.e., at least two attended appointments) of low- and/or high-intensity treatment during this time. However, this study only focused on investigating post-treatment outcomes of patients who completed an episode of high-intensity psychotherapy. This was determined to have occurred if patients attended at least two appointments of therapy with the same high-intensity therapist, and they were discharged from the service with their episode of treatment concluding with appointments delivered by this same therapist. Based on these criteria, *n*=2430 patients received an episode of high-intensity treatment.

However, not all patients were offered or attended a follow-up review appointment. Furthermore, the recording of follow-up reviews being delivered was inconsistent, with some reviews taking place but then not being explicitly recorded by the therapist as being a follow-up review specifically. Consequently, in order to identify from the data whether a patient attended a follow-up review or not, a number of criteria needed to be developed and applied. These criteria were developed through discussions with the IAPT service manager and data manager, and using a data triangulation approach.

To summarise the applied criteria, patients were assumed to have attended a follow-up review appointment when: (1) after completion of their high-intensity treatment episode and discharge from the service, they attended at least one further appointment with the same high-intensity therapist who delivered their episode of treatment; (2) this new appointment with the same therapist was labelled as being a new episode of care within the service; (3) this new appointment was the first appointment that the patient received in the new episode of care (i.e., they did not have an assessment appointment with a different practitioner); and (4) this new appointment occurred at least 30 days after their discharge from the service following their treatment episode (to ensure that there was reasonable time between end of treatment and follow-up review). After applying these criteria, a total of *n*=399 patients were identified as having attended at least one follow-up review appointment following treatment. This represents 16.4% of the *n*=2430 patients who completed an episode of high-intensity psychotherapy.

The subsample of *n*=399 followed-up patients was further reduced for the purposes of addressing the study objective of exploring relapse following completion of high-intensity psychotherapy with remission of symptoms. Patients were excluded if they did not present with a clinically significant depression and/or anxiety problem as indicated by scoring above clinical thresholds on related symptom outcome measures (described below) at the beginning of treatment, had missing symptom outcome scores at the beginning of treatment, or did not have scores on these outcome measures that were below clinical thresholds at the end of treatment (*n*=134 excluded based on scores[[8]](#footnote-8)). In addition, *n*=1 patient was excluded for being below the age of 18 upon intake into the service, *n*=1 patient was excluded for having dropped out of their treatment episode, and *n*=1 patient was excluded as they were referred to a non-IAPT service for ongoing mental healthcare following treatment discharge. This resulted in a final sample of *n*=262 adult patients who completed an episode of high-intensity psychotherapy with remission of symptoms and subsequently attended at least one follow-up review appointment. Figure 4.1 summarises the data management and curation process that was followed to identify the study sample.

Of the *n*=262 patients in the final sample, *n*=228 had at least one follow-up review that was explicitly recorded by the therapist as being a follow-up review. Meanwhile, *n*=34 patients did not have an explicitly recorded follow-up review, but were assumed to have one based on the criteria discussed previously. The majority of patients (*n*=171) attended only one follow-up review within the study period, with *n*=91 (34.7%) attending more than one review. Specifically, *n*=17 of the *n*=91 patients attended three follow-up reviews, with *n*=1 patient attending four reviews. It is important to note that there was variability within the dataset of when patients completed their episodes of treatment.

**Figure 4.1**

*Flow Chart Representing Data Management and Curation Process*

N = 19138 referred to services

N = 7514 received treatment

HI treatment episode received if:

* ≥2 appointments with the same HI therapist
* Discharged with episode concluding with appointments with same HI therapist

N = 2430 received HI treatment

N = 399 attended ≥1 follow-up review

N = 262 eligible for relapse investigation

N = 91 followed-up more than once

Abbreviations: GAD-7, Generalized Anxiety Disorder scale; HI, high-intensity; PHQ-9, Patient Health Questionnaire.

Follow-up review attended if:

* After HI treatment episode, ≥1 further appointment with same HI therapist was attended
* Appointment labelled as new episode of care
* Appointment was first session in new episode
* Appointment occurred ≥30 days after HI episode

Ineligible for relapse investigation if:

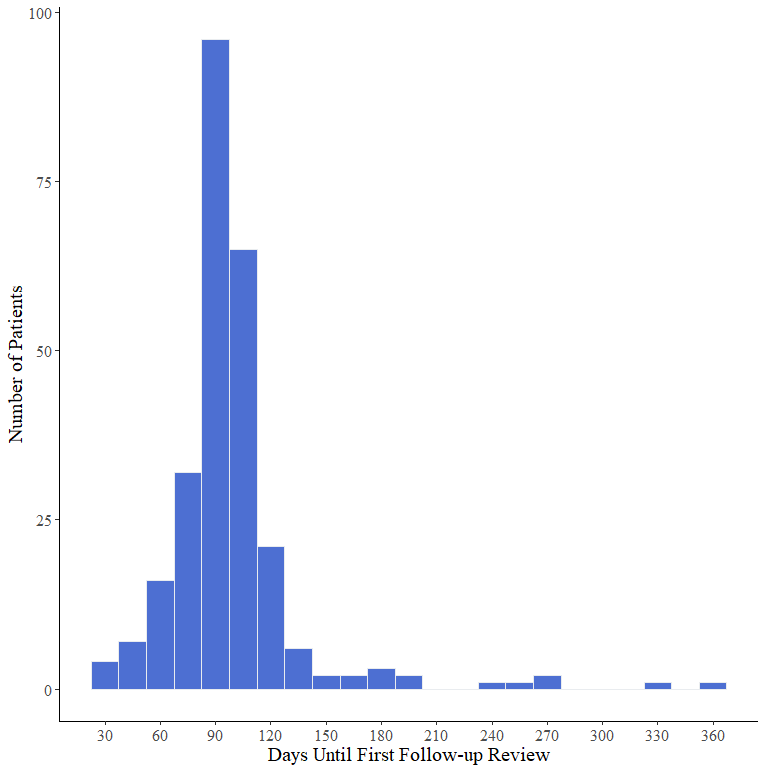
* Both PHQ-9 and GAD-7 scores below clinical thresholds at start of treatment
* Either score above threshold at end of treatment
* Missing symptom scores at start of treatment

For example, the earliest discharge from treatment in this occurred in August 2019, while the latest occurred in April 2021. Considering that the study observation period ended in May 2021, some patients had greater opportunity to attend further follow-up review appointments than others[[9]](#footnote-9).

The first follow-up review that patients attended occurred on average 99.0 days (*SD* = 38.2; range = 30 – 357 days) following treatment completion. Specifically, *n*=234 patients (89.3%) had their first review within four months (120 days), while *n*=252 patients (96.2%) had their first within six months (180 days). Figure 4.2 illustrates the variance in follow-up timings for the first attended follow-up review. A Mann-Whitney U test found no significant relationship between follow-up time and relapse (*p* = .985).

**Figure 4.2.**

*Distribution of lengths of time between patients’ discharge from treatment and first follow-up review appointment.*



The most common primary presenting problem recorded in clinical records for the treatment episode received by the *n*=262 patients was an affective disorder (depressive episode or recurrent depressive disorder), with 40.1% presenting with this problem. Meanwhile, 30.9% of patients were recorded as primarily being treated for an anxiety disorder (e.g., generalized anxiety disorder, specific phobia), 15.3% for mixed anxiety and depression, and 9.5% for a range of other problems (e.g., post-traumatic stress disorder, obsessive compulsive disorder, bereavement etc.). The remaining 4.2% of patients had missing information related to their primary presenting problems. The majority of patients received CBT (61.5%), with 38.5% receiving non-CBT-based psychotherapy (e.g. Counselling for Depression, interpersonal psychotherapy). Unfortunately, identifying the specific form of non-CBT psychotherapy was not possible with the available dataset. Before beginning the high-intensity treatment episode investigated in this study, *n*=61 (23.3%) of patients received an episode of low-intensity or group intervention (>= two sessions), while *n*=11 (4.2%) received a separate episode of high-intensity psychotherapy delivered by a different therapist (*n*=1 patient received both types of intervention before their investigated episode of treatment). At the beginning of their high-intensity treatment, patients had mean scores of 14.18 (*SD* = 4.87) and 13.79 (*SD* = 4.14) on the PHQ-9 and GAD-7 respectively, with these mean scores decreasing to 4.60 (*SD* = 2.68) and 4.17 (*SD* = 2.07) by the end of treatment respectively. The majority of patients (*n*=232; 88.5%) both completed their treatment episode and attended their first follow-up review appointment after the United Kingdom government’s first implementation of a ‘lockdown period’ from March 23rd 2020, which was introduced to both contain and prevent the spread of the coronavirus during the Covid-19 pandemic. In contrast, *n*=10 patients (3.8%) both completed their treatment and attended their follow-up review before the implementation of the first lockdown period, while *n*=20 (7.6%) completed treatment before March 23rd 2020 but had not yet attended their follow-up review.

In terms of demographics[[10]](#footnote-10), the subsample was characterized by a majority of female patients (72.9%) from a white British background (88.3%), with a mean age of 40.32 (*SD* = 14.8). Approximately 34.4% of patients were unemployed upon intake to the service, while 32.1% were unemployed at the end of treatment. There was no information within the dataset related to long-term medical conditions, and the vast majority of data related to reported disabilities were missing (85.9%). Patients accessed a mean of 9.0 high-intensity sessions with their therapist (SD = 3.6; range = 2-21 sessions).

**Measures**

Similar to Chapter 3, this study investigated patient-reported scores on three clinical outcome measures: the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), the Generalized Anxiety Disorder scale (GAD-7; Spitzer et al., 2006), and the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002). However, in contrast to Chapter 3, which applied a reliable change index of ≥5 for both the PHQ-9 and the GAD-7 to monitor deterioration over time, this study applied a reliable change index of ≥6 for the PHQ-9 and ≥5 for the GAD-7. The reliable change indices adopted in Chapter 3 were to ensure consistency with the original WYLOW studies (Ali et al., 2017; Delgadillo et al., 2018b), from which the data investigated by that chapter’s study were collected. Meanwhile, the reliable change indices applied in this chapter, were recommended more recently by Richards and Borglin (2011).

This study had access to each patient’s scores on the above measures from every treatment session and follow-up review that patients attended. If patients had missing scores on any of the three measures from any treatment episode appointment, missing data-points were imputed by applying the last-observation-carried-forward (LOCF) method, in order to analyse data in a way that is consistent with other IAPT studies (Clark et al., 2009). A series of other clinical, demographic and treatment variables were also available, and these are summarized in Table 4.1.

***Primary Outcome***

Similar to Chapter 3, for a patient to be classed as having relapsed, they must have met the following two criteria: (a) at least one of the symptom measures (PHQ-9 or GAD-7) was above the clinical cut-off at a follow-up review; and (b) the measure represented statistical reliable deterioration relative to the score observed at the final treatment session (based on the reliable change index).

|  |
| --- |
| **Table 4.1**  *The Variables Explored as Predictors of Relapse Through the Developed XGBoost Models* |
| Age; gender; ethnicity; unemployment at start; unemployment at end; taking medication at start; taking medication at end; neighbourhood deprivationa; diagnosis (affective); diagnosis (anxiety); diagnosis (mixed anxiety and depression); diagnosis (other); any previous treatment within episode; baseline PHQ-9; baseline GAD-7; baseline WSAS; early treatment response PHQ-9; early treatment response GAD-7; early treatment response WSAS; PHQ-9 at end; GAD-7 at end; WSAS at end; number of treatment sessions attended; linear treatment response (GoF) – PHQ-9; linear treatment response (GoF) – GAD-7; linear treatment response (GoF) – WSAS; computed noise variable (continuous); computed noise variable (categorical) |
| aEach patient’s home post code was linked to the English Index of Multiple Deprivation (IMD; Department for Levelling Up, Housing and Communities, 2019), which is a measure of relative deprivation for small geographical areas. The IMD ranks each area from the most to least deprived, based on seven domains (e.g., education level, crime, quality of local environment, barriers to housing and services, etc.). The indices are aggregated into local scores and decile groups, and this study further aggregated these groups into quintiles (where 1 = most deprived, 5 = least deprived).  Abbreviations: GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit for a regression model; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale. See Appendix F for glossary for predictors. |

**Analysis**

***Potential Selection Biases***

To begin, potential sources of selection bias were investigated by comparing characteristics between cases where a follow-up review was (*n*=399) or was not (*n*=2031[[11]](#footnote-11)) scheduled/attended. The specific characteristics compared were age, gender, number of attended sessions with the high-intensity therapist, PHQ-9 scores at end of treatment, GAD-7 scores at end of treatment, and WSAS scores at end of treatment. Pearson’s chi-square tests were used to compare categorical data, while Welch’s t-tests or Mann-Whitney U tests were used to compare continuous variables, depending on distributions. These analyses were repeated to compare only those patients in both subgroups (*n*=262 followed-up *vs* *n*=875 not followed-up) who met the inclusion criteria outlined previously (i.e., started treatment with symptoms above clinical cut-off(s), ended treatment with symptoms below clinical cut-offs, aged 18 years old or above, and did not drop out of treatment or be referred to non-IAPT services for continued mental health support).

***Machine Learning Approach***

The initial plan was to develop an ensemble of four predictive models that predicted relapse in a dynamic manner over the follow-up phase. Each successive model would have been trained using additional information that becomes available over time. Model 1 would have used information only known at the end of treatment; Model 2 would have learned from information known at the three-month follow-up review; Model 3 would have included information available at the six-month review; and Model 4 would have included information available at the nine-month review. Each model would have been developed to predict the occurrence of a relapse at the next follow-up review. For example, Model 1 would aim to predict relapses that occurred at the three-month follow-up review, while Model 2 would aim to predict relapses that occurred at the six-month review etc. However, the sample of eligible patients who were followed-up by the end of the study period was smaller than anticipated. Indeed, only *n*=262 patients were eligible for inclusion. Consequently, only one predictive model was developed, with this model being trained using only information known at the final treatment session to predict whether a patient will relapse at their first attended follow-up review (regardless of when this occurred during the 12-month period post-treatment). The 29 variables tested as potential predictors are summarized in Table 4.1, and further described in Appendix F (glossary of predictors). The potential predictors ranged from baseline characteristics to factors related to treatment response.

**Model Development.** The supervised machine learning approach of extreme gradient boosting (XGBoost: Chen & Guestrin, 2016), as implemented in the R package *xgboost* (Chen et al., 2021; version 1.4.1.1), was applied to develop the model. The specific approach implemented in this study was highly similar to the approach adopted in Chapter 3. For example, the parameters *eta*, *maximum depth*, *alpha* and *lambda* were set to the same values (0.1, 3, 0, and 1 respectively). In addition, other similarities include the use of the area under the receiver operating characteristic curve (AUC) as the evaluation metric to assess the predictive ability of the model, the use of stratified five-fold internal cross-validation alongside an “early stopping rule” of 10 rounds, and the incorporation of ‘leave-one-variable-out loops’ for the purposes of variable selection.

On the other hand, there were also two additions to the approach carried out in Chapter 3. Firstly, during data pre-processing, two ‘noise variables’ were randomly computed and input into the XGBoost model, alongside the various demographic, clinical, and treatment variables that were investigated as potential predictors. The inclusion of these ‘noise variables’ provides one method of exploring whether a developed model has overfit to the data (i.e., if a model includes a ‘noise variable’ after performing variable selection, then the model is overfitting). The two variables consisted of: one random continuous variable computed based on the sample’s final treatment PHQ-9 scores (i.e., it had the same mean and standard deviation); and one random categorical variable computed based on the sample’s level of unemployment at the beginning of treatment (i.e., it had the same base rate).[[12]](#footnote-12)

Secondly, this study also incorporated LASSO regularization in the development of the final model. Specifically, the predictive variables involved in the development of the best model from the leave-one-out loops process were identified and input into three final XGBoost models; each incorporating varying levels of LASSO regularization. These models had the exact same hyperparameters that had been set previously, except for *alpha*, which varied across the three models: one model had *alpha* set at 0.1; one model had *alpha* set at 0.25; and one model had *alpha* set at 0.5. In other words, all three models included the same selected variables and applied the same values for *eta*, maximum *depth*, and *lambda*, but applied varying levels of LASSO regularization. The model with the highest AUC observed in the internally cross-validated test sample was determined to be the final developed XGBoost model. The incorporation of a LASSO penalty at this stage of model development may have further helped prevent the final developed model from being overfit to the data. Furthermore, it may also have eliminated potentially spurious variables that the model thus far still considered as being important.

**Model Evaluation.**The final developed model was evaluated in the same manner as the models developed in Chapter 3. Specifically, the predictive value of the model will be assessed when internally cross-validated out-of-sample (in the 1/5 test set). During this process, predicted probabilities of relapse for each patient were calculated, ranging from 0-1 with higher scores indicating greater probability of relapse. Conventional performance metrics were calculated to assess the predictive performance of the model: AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). It has recently been recommended to compare multiple risk thresholds when evaluating a developed predictive model, as usually there is no universally optimal threshold, but that the most plausible threshold will depend on the clinical context (Wynants et al., 2019). Therefore, the latter five performance metrics (all excluding AUC) were calculated twice, using two different risk thresholds: 0.50, and 0.19 (i.e., the base rate of relapse in the sample). For each threshold, if a patient’s calculated probability was greater than the respective threshold, they were predicted to experience a relapse.

**Exploring Individual Predictors.** The variables determined to be important by the developed model were explored in terms of their associations with relapse using relative importance metrics, partial dependence plots, and descriptive statistics.

**Sensitivity Analysis.**The sample was found to be class-imbalanced (i.e., the number of observations belonging to each outcome class – relapse vs. remission – were not similar), which has been argued to undermine the accuracy of machine learning algorithms (Lopez et al., 2013). The resampling technique SMOTE (synthetic minority over-sampling technique; Chawla et al., 2002) attempts to address the skewed distribution by generating new minority class observations, and thus balancing the outcome classes within the sample. Therefore, the XGBoost model development approach discussed above was repeated separately with the implementation of SMOTE, to explore whether the addition of this resampling technique improved the predictive ability of the developed model.

However, there were a few changes to the previous approach in order to allow SMOTE to be implemented and the developed XGBoost model to be evaluated. For instance, before implementing SMOTE, missing data needed to be imputed. Therefore, nonparametric missing value imputation was performed on the sample using the R package *missForest* (Stekhoven, 2013; version 1.4). *MissForest* has been demonstrated to be effective at handling missing values in variables that have up to 30% missing information (Stekhoven & Bühlmann, 2012). Following this, SMOTE was implemented using the R package *smotefamily* (Wacharasak, 2019; version 1.3.1). For the synthetically produced minority class observations, the generated variable values were not integers, but had multiple decimal values. This would be an issue for binary predictor variables, as the algorithm could select values between 0 and 1 as being the important values for the prediction of relapse, rather than the integer value that represents the category of the respective variable. Therefore, the synthetically produced values for the binary predictor variables were rounded to ensure the variables remained binary in nature. In addition, new noise variables needed to be computed for the entire dataset (real and synthetic cases) following the implementation of SMOTE. As with the class imbalanced model, the two noise variables consisted of one spurious continuous variable computed based on the sample’s final treatment PHQ-9 scores, and one spurious categorical variable computed based on the sample’s level of unemployment at the beginning of treatment. Finally, before evaluating the developed model’s predicted performance, the predicted probabilities for the synthetically generated minority class cases were removed, and the conventional performance metrics were only calculated using the *n*=262 real cases.

**Results**

**Analyses of Potential Selection Biases**

Comparisons of the characteristics of patients who did (*n*=399) and who did not (*n*=2031[[13]](#footnote-13)) attend at least one follow-up review following an episode of high-intensity psychotherapy indicated that older patients (*U* = 377181, *p* = .029), those who attended more contacts (*U* = 275812, *p* < .001), and those that had lower post-treatment PHQ-9 (*U* = 522091, *p* < .001), GAD-7 (*U* = 517866, *p* < .001), and WSAS (U = 505502, *p* < .001) scores were more likely to attend at least one follow-up review. There was no significant gender difference (*p* = .283) between those patients who did and who did not attend at least one follow-up review.

Comparisons of patients who did (*n*=262) and who did not (*n*=875) attend at least one follow-up review following the completion of an episode of high-intensity psychotherapy *with remission of symptoms* indicated that those who attended more sessions (*U* = 97733, *p* < .001) and those who had lower post-treatment WSAS scores (*U* = 129973, *p* < .001) were more likely to be followed-up. There were no significant differences in terms of age (*p* = .205), gender (*p* = .422), post-treatment PHQ-9 scores (*p* = .065), or post-treatment GAD-7 scores (*p* = .128).

**Relapse Rates**

Of the *n*=262 patients who attended at least one follow-up review within 12 months of completing high-intensity psychotherapy with remission of symptoms, *n*=50 were classed as having experienced a relapse by their first follow-up review. This represents a relapse rate of 19.1% following high-intensity psychotherapy[[14]](#footnote-14). In terms of treatment modality, the relapse rate for the *n*=161 patients who received CBT was 18.6% (*n*=30 relapse cases), while the relapse rate for the *n*=101 patients who received non-CBT-based psychotherapy was 19.8% (*n*=20 relapse cases). A Pearson’s chi-square test found no significant difference between the two therapeutic modalities in terms of relapse (*p* = .815).

Of the *n*=91 patients who attended at least two follow-up reviews, *n*=64 did not experience a relapse in either of the first two reviews, *n*=9 were classed as having relapsed in both reviews, *n*=7 had a relapse in their first review only, while *n*=11 had a relapse in their second review only. Meanwhile, of the *n*=17 patients who attended at least three follow-up reviews, *n*=10 did not relapse in any of the three reviews, *n*=1 was classed as having relapsed in each of the three reviews, *n*=3 experienced a relapse in their first review only, *n*=1 in their second review only, and *n*=1 in their third review only. Finally, the *n*=1 patient who attended four follow-up reviews within the 12-month period post-treatment did not experience a relapse in their first three reviews, but was classified as having relapsed in their final review. In summary, of the *n*=212 patients who did not relapse in their first follow-up review following treatment completion, *n*=75 received at least one additional follow-up, and *n*=14 of these patients experienced a relapse at a further follow-up review.

**Performance of XGBoost Model**

The primary XGBoost model was developed using 45 decision trees, and the model was developed using an alpha (L2 regularization) of 0.1. Ten of the 29 input variables were selected as predictive features by the model, with both of the computed noise variables being selected (indicative of overfitting). The model had an internal cross-validation AUC of 0.72. When adopting a predictive risk threshold of 0.50 (i.e., 50% expected probability of relapse), the model only predicted *n*=10 patients to experience a relapse, and was estimated as having an accuracy of 80.9%, sensitivity of 10.0%, specificity of 97.6%, PPV of 50%, and NPV of 82.1%. In contrast, when adopting a risk threshold of 0.19 (i.e., 19% expected probability of relapse, consistent with the observed base rate in the sample), the model predicted *n*=101 patients to experience a relapse, and was estimated to have an accuracy of 67.6%, sensitivity of 66.0%, specificity of 67.9%, PPV of 32.7%, and NPV of 89.4%.

***Sensitivity Analysis***

The secondary XGBoost model (same as primary model but developed with an additional *n*=150 synthetic minority class observations generated using SMOTE) was developed using 160 decision trees, and an alpha level of 0.25. Eighteen of the 29 input variables were selected as predictive features by the model, and neither of the computed noise variables were selected (indicating the model was less likely to be undermined by overfitting). This model had an internal cross-validation AUC of 0.81. When using a risk threshold of 0.50, the model was estimated as having an accuracy of 78.2%, sensitivity of 54.0%, specificity of 84.0%, PPV of 44.3%, and NPV of 88.6%. In contrast, when adopting a risk threshold of 0.19, the model was estimated as having an accuracy of 64.5%, sensitivity of 84.0%, specificity of 59.9%, PPV of 33.1%, and NPV of 94.1%. The predictive value of the two developed models when evaluated on the test set is summarized in Table 4.2, and the distributions of the probabilities of relapse predicted by each model for each patient can be seen in Figure 4.3.

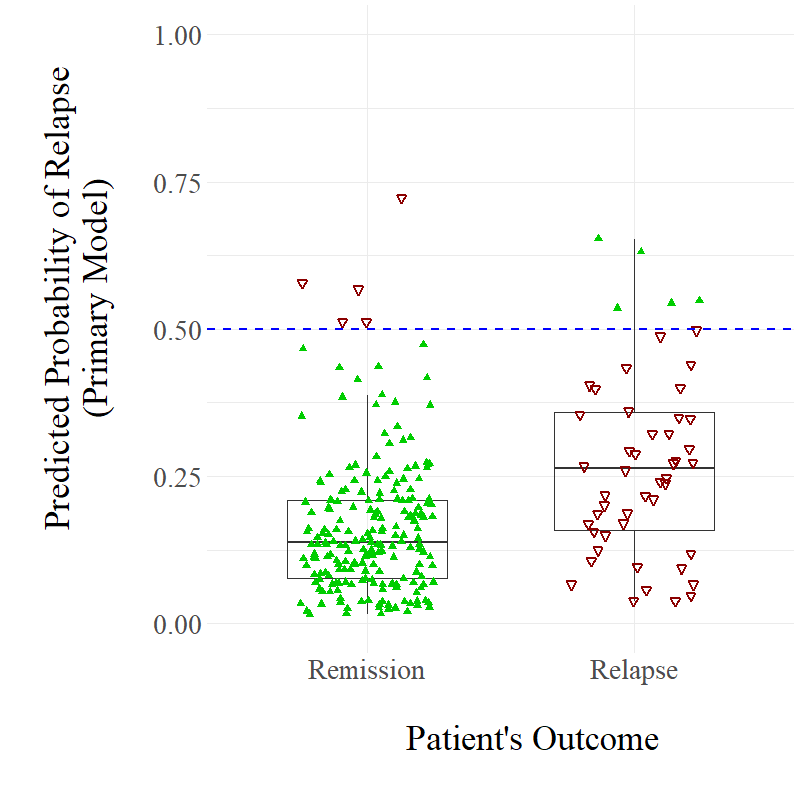
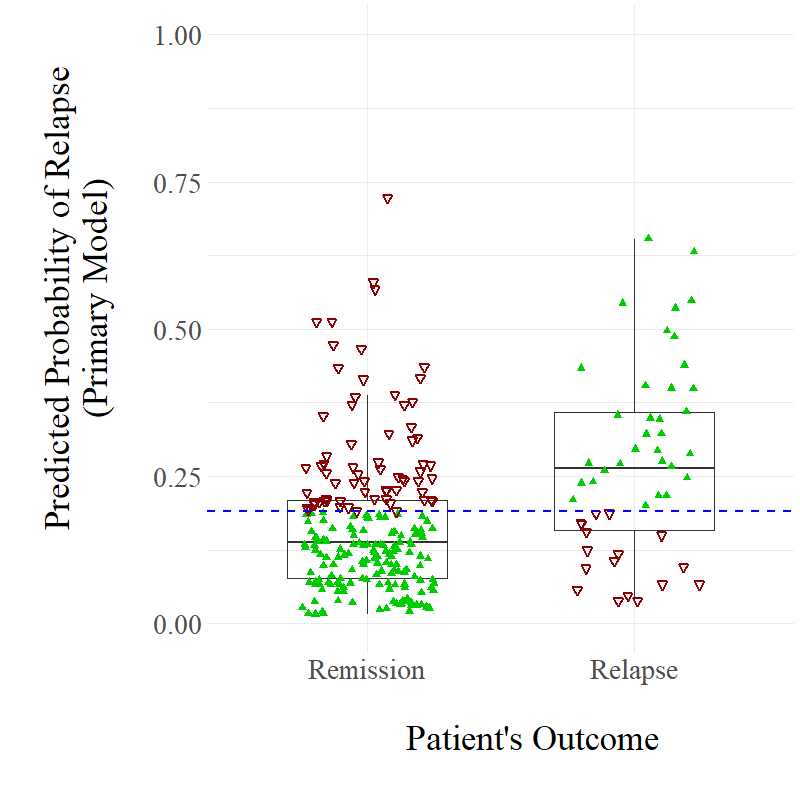
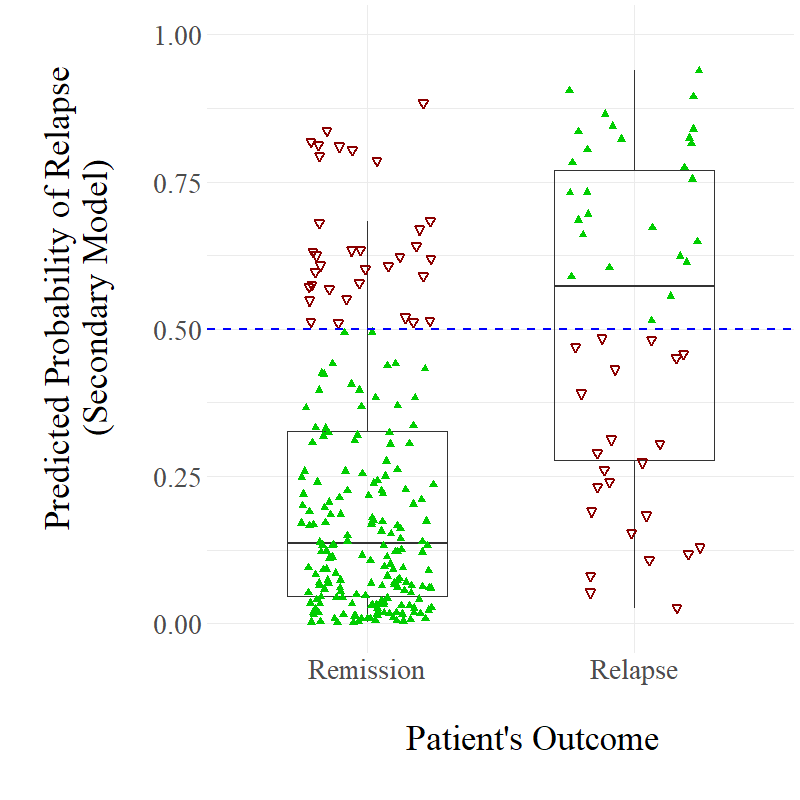
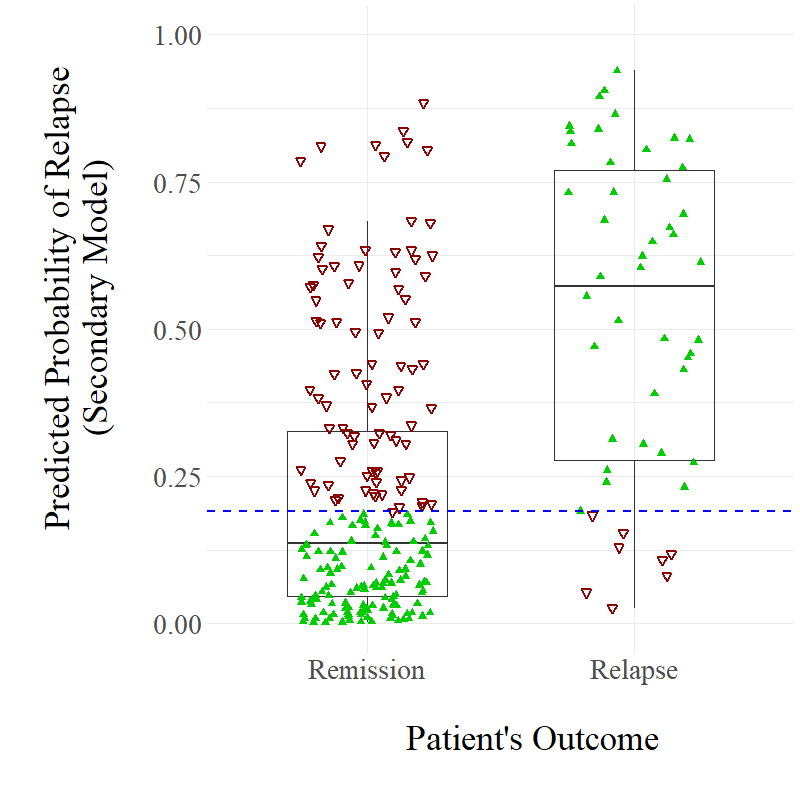
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 4.2**  *The Ability of Each Model to Predict Relapse When Evaluated on the ‘Test’ Set* | | | | |
|  | Primary Model  (Class-Imbalanced) | | Secondary Model  (Class-Balanced - SMOTE) | |
| L2 Regularization | 0.1 | | 0.25 | |
| Number of Decision Trees | 45 | | 160 | |
| AUC | 0.72 | | 0.81 | |
| *Diagnostic Risk Threshold* | *0.50* | *0.19* | *0.50* | *0.19* |
| Accuracy | 80.9% | 67.6% | 78.2% | 64.5% |
| Sensitivity | 10.0% | 66.0% | 54.0% | 84.0% |
| Specificity | 97.6% | 67.9% | 84.0% | 59.9% |
| PPV | 50.0% | 32.7% | 44.3% | 33.1% |
| NPV | 82.1% | 89.4% | 88.6% | 94.1% |
| Abbreviations: AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value | | | | |

**Important Predictors**

Of the 29 variables input as potential predictors of relapse for both models, a total of 22 predictors were deemed important (to varying degrees) by at least one of the two developed models. There was little consistency between the two models in terms of the selected predictors, with only five of the 10 predictors selected by the primary model being included in the 18 predictors selected by the secondary (SMOTE-balanced) model. The predictors identified as being important by each model, and the relative importance of each variable for the prediction of relapse, are displayed in Table 4.3. For the sake of parsimony and consistency, the selected important predictors (excluding the two noise variables) can be grouped according to the four categories discussed in Chapter 3: demographics; baseline clinical features; treatment process features; and residual symptoms. Partial dependence plots can be found in Appendix G.

**Figure 4.3**

*Distributions of Predicted Patient Relapse Probabilities made by the Primary (Panels a and b) and Secondary Models (Panels c and d)*



a)

b)

c)

d)

*Note.* Panels a and c represent predictions when the risk threshold is 0.5, while panels b and d represent predictions when the threshold is 0.19. Filled, green triangles represent correct predictions for specific cases; unfilled, red upside-down triangles represent incorrect predictions.

|  |  |
| --- | --- |
| **Table 4.3**  *The Predictors Identified as Important by Each Developed Model (Relative Importance %)* | |
| Primary Model (Class-Imbalanced) | Secondary Model (Class-Balanced) |
| WSAS at End (26.7%) | Neighbourhood Deprivation (19.2%) |
| *Continuous noise variable (22.1%)* | Baseline WSAS (9.4%) |
| Baseline PHQ-9 (19.0%) | Baseline GAD-7 (9.2%) |
| Diagnosis – Mixed Anxiety/Depression (9.8%) | PHQ-9 at End (8.3%) |
| GAD-7 at End (9.7%) | WSAS at End (8.2%) |
| Treatment Modality (4.9%) | Age (8.2%) |
| *Categorical noise variable (2.8%)* | Baseline PHQ-9 (8.0%) |
| Neighbourhood Deprivation (2.2%) | Number of Attended Treatment Sessions (6.4%) |
| Unemployed at Start (1.7%) | Diagnosis – Mixed Anxiety/Depression (6.3%) |
| Diagnosis – Anxiety (1.1%) | Linear Treatment Response WSAS (GoF; 6.0%) |
|  | Early Treatment Response PHQ-9 (3.0%) |
|  | Diagnosis - Other (2.5%) |
|  | Unemployed at Start (2.4%) |
|  | Taking Medication at Start (1.8%) |
|  | Gender (<1%) |
|  | Previous Treatment in Episode of Care (<1%) |
|  | Treatment Modality (<1%) |
|  | Unemployed at End (<1%) |
| Abbreviations: GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit for a regression model; PHQ-9, Patient Health Questionnaire; SMOTE, synthetic minority over-sampling technique; WSAS, Work and Social Adjustment Scale. See Appendix F for glossary for predictors. | |

***Demographics***

Demographic predictors leveraged little importance in the primary model, with variables of this type only having a combined importance of 3.9%. However, they had substantially more importance for the secondary model, with a combined importance of 30.3%. This comparative increase is primarily due to neighbourhood deprivation being by far the most important predictor in the secondary model. This variable was also only one of two demographic predictors selected by both models. Exploration of this feature indicated that, although patients in the least deprived areas were the least likely to relapse, patients in the most deprived areas were also less likely to relapse than patients with more moderate levels of neighbourhood deprivation – thus indicating a potentially non-linear relationship. The other demographic variable to be selected by both models was pre-treatment unemployment, with *n*=23/90 patients (25.6%) who were unemployed before starting treatment relapsing, and *n*=27/172 employed patients (15.7%) relapsing. However, this variable had relatively little importance in both models. Three other demographic features were selected as important by the secondary model: middle-age (30-55 years of age), being female, and unemployment post-treatment. However, the latter two features had negligible importance.

***Baseline Clinical Features***

Pre-treatment clinical factors had a combined importance of 29.9% for the primary model, and 40.3% for the secondary model. Two features of this type were selected by both models: higher baseline PHQ-9, and not having a recorded primary problem of mixed anxiety and depression. Interestingly, only *n*=2/40 patients treated primarily for mixed anxiety and depression experienced a relapse. Pre-treatment PHQ-9 score was the only baseline clinical measure selected as a predictor in the primary model; however, all three baseline clinical measures (PHQ-9, GAD-7, and WSAS) were selected as important in the secondary model, and each had similar, relatively high levels of importance. Specifically, patients with higher WSAS and GAD-7 scores were at greatest risk of relapse. However, according to partial dependence plots, patients with the highest WSAS scores (>30) were in fact at least risk of relapse. In terms of diagnostic features, having an anxiety disorder as the primary presenting problem was selected by the primary model as being associated with a decreased risk of relapse, with *n*=16/81 patients with this primary problem experiencing a relapse. Meanwhile, having a primary problem other than an anxiety disorder, an affective disorder, or mixed anxiety and depression (e.g., post-traumatic stress disorder, obsessive compulsive disorder, bereavement etc.) was selected by the secondary model as being predictive of relapse. Indeed, *n*=15/25 patients with this type of primary problem experienced a relapse. However, both of these diagnostic features had relatively little importance in their respective models. Finally, two other baseline clinical features were selected by the secondary model as being predictive of relapse, however both had little importance: taking medication at the beginning of treatment, and not having previous treatment within the same episode of care.

***Treatment Process Features***

Treatment process features had relatively little importance in both models. The combined importance of these features was 4.9% in the primary model and 15.6% in the secondary model. Treatment modality was the only variable of this type selected by both models, however it had negligible importance in the secondary model, and the direction of association was different according to both models. Specifically, the primary model indicated that patients who received CBT were less likely to relapse, while the secondary model indicated that these patients were more likely to relapse. As demonstrated previously, there were little differences between CBT and non-CBT therapies in terms of relapse rates. Three other features of this type were selected by the secondary model as being predictive of relapse: attendance of fewer treatment sessions, non-linear change in GAD-7 over the course of treatment, and a moderate early treatment response in the PHQ-9 measure (i.e., decrease of 3-7 points). However, for the latter feature, partial dependence plots also indicated that a large early treatment response (i.e., decrease of >7 points by session 3) was associated with the lowest risk of relapse.

***Residual Symptoms***

The combined importance of residual symptoms (sub-threshold symptom scores close to the diagnostic cut-offs) post-treatment was 36.4% for the primary model, and 16.5% for the secondary model. Post-treatment WSAS score was the only feature of this type to be selected by both models, and in fact, was the only selected predictor with higher relative importance than the computed continuous noise variable in the primary model. Generally, patients with higher post-treatment WSAS scores were at greater risk of relapse according to both models. However, patients with the lowest WSAS scores (i.e. 1-5) were also at slightly increased risk according to partial dependence plots. Meanwhile, higher post-treatment GAD-7 scores and higher post-treatment PHQ-9 scores were only selected by the primary model and secondary model respectively, with both features having relatively high levels of importance for their respective models.

**Discussion**

This study investigated relapse following high-intensity psychotherapy for depression and anxiety. Routinely collected data from patients who were followed-up at least once following the completion of psychotherapy at English IAPT services was examined. It was estimated that approximately 19.1% of patients experienced a relapse by their first follow-up review, and that there was little difference between CBT and non-CBT-based psychotherapy in terms of relapse rates. This study also attempted to extend upon the study discussed in Chapter 3, by implementing a similar machine learning approach to develop predictive models capable of identifying patients who are at greatest risk of relapse following completion of high-intensity psychotherapy. The developed models displayed promising ability to discriminate between relapse and remission cases (AUC = 0.72-0.81), further indicating the potential utility of applying machine learning approaches to predict cases at ‘high risk’ of relapse following psychological interventions.

The identified relapse rate following high-intensity psychotherapy of 19.1% is similar to some studies previously conducted. For example, the systematic review discussed in Chapter 2 and the recent meta-analysis conducted by Levy et al. (2021) reported pooled relapse rates of 21.8% and 14% respectively for the treatment of anxiety disorders with CBT. However, this study’s relapse rate is also slightly smaller than relapse rates reported by other previous research. For instance, meta-analyses conducted by Vittengl et al. (2007) and Wojnarowski et al. (2019) estimated relapse rates of 29% and 33.4% respectively for the treatment of depression with CBT. However, the latter reviews investigated studies that involved longer follow-up durations than the present study, with the included studies having a minimum of 12 months’ follow-up. This may explain the difference in reported relapse rates, with these studies having more time to observe relapse events than the present study. Indeed, Vittengl et al. (2007) demonstrated that studies with longer follow-up durations report significantly higher relapse rates. Furthermore, the present study investigated a larger sample of participants who responded to treatment and were followed-up than any study included in any of the four mentioned reviews (range of *n*=5-172 across all four reviews).

The present study’s identified relapse rate following high-intensity psychotherapy is substantially smaller than the relapse rate following LiCBT of 53% reported in the original WYLOW study (Ali et al., 2017). However, the 53% statistic represents a 12-month relapse rate and therefore this is not a fair comparison, considering that the current chapter’s study involved a less consistent, and overall shorter duration of follow-up. Indeed, a large majority of patients in this study (89.3%) were followed-up within 1-4 months following the completion of treatment. Nevertheless, the WYLOW study did report a relapse rate following a similar length of follow-up time (4 months), with approximately 37% of the investigated sample relapsing within this period. Therefore, this indicates that relapse following high-intensity psychotherapy is less common than following LiCBT.

One potential reason for this is that LiCBT is a briefer intervention that is highly manualized, self-help oriented, and highly guided by the use of structured psychoeducational workbooks (Bennett-Levy et al., 2010). Consequently, high-intensity interventions may enable the provision of more intense and more bespoke support to patients, which may enable patients greater opportunity to learn and implement skills gained through therapy. However, there may also be a methodological explanation for the difference in relapse rates between the two studies. For instance, when estimating relapse occurrence, the present study only investigated one follow-up assessment at one point of time, while the WYLOW study investigated multiple follow-up assessments that occurred on a monthly basis. Consequently, the WYLOW study had a greater chance of observing relapse occurrence, as its intense and structured follow-up procedure left less gaps in follow-up time. Indeed, the meta-analysis conducted by Vittengl et al. (2007) found that studies that had larger gaps in follow-up assessment time reported significantly lower relapse rates.

This study also found no significant difference between CBT and non-CBT-based psychotherapies in terms of relapse rates (18.6% vs 19.8% respectively), which supports previous research in this area. For example, Vittengl et al’s (2007) meta-analysis found no significant difference between CBT and other psychotherapeutic approaches in terms of depressive relapse rates. More recently, the long-term follow-up of a randomized controlled trial comparing CBT and interpersonal psychotherapy (IPT) for the treatment of depression found no significant difference in terms of relapse rates (Lemmens et al., 2019). This indicates that psychotherapies other than CBT have similar enduring effects. However, it should be noted that Vittengl et al.’s meta-analysis only included four studies, demonstrating the limited research that has compared relapse rates between CBT and other psychotherapeutic interventions. Therefore, further research is needed that compares CBT with other specific modalities of treatment in terms of relapse.

The XGBoost models developed in this study demonstrated promising ability to discriminate between relapse and remission cases (AUC = 0.72-0.81), further indicating the potential utility of applying machine learning approaches to predict the occurrence of relapse following psychotherapy (as previously demonstrated in Chapter 3). However, it should be noted that the primary (class-imbalanced) developed model included both of the computed noise variables, thus demonstrating that the model was unstable and likely overfitting to the data. Consequently, considering that the secondary (SMOTE-imputed) developed model did not select either of the noise variables, this model should be considered the more credible model of the two. However, considering the small sample size, the low event base rate of relapse, and the high number of features selected as important by this model (*k* = 18), it is likely that this model also overfitted to the dataset and that its AUC statistic of 0.81 is inflated. Therefore, the models and their selected predictive features should be interpreted with caution, and regarded as exploratory.

Compared to Chapter 3, the important variables were harder to interpret within this study, due to fewer models being developed, and there being less consistency between models in terms of the selected features. Nevertheless, a number of potential predictors of relapse that could be investigated further in future research were identified. For example, similar to Chapter 3, residual symptoms were identified as being important for the prediction of relapse. Indeed, higher post-treatment WSAS scores was the most important predictor in the primary developed model, and the fifth most important predictor in the secondary model. Furthermore, this was the only variable that had greater relative importance than a computed noise variable for both models, potentially indicating that this variable was the most stable identified feature. However, it should be noted that further exploration of this feature demonstrated that, although a very high score of WSAS post-treatment was associated with the greatest risk of relapse, a very low score was also associated with a relative increase in risk. This limits interpretability of this feature, and therefore further research is needed to understand the relationship between post-treatment functioning and relapse. Nevertheless, residual depressive and anxious symptoms were selected as important by one of the two developed models each, with each having relatively high levels of importance for their respective model. Combined, this further indicates the potential importance of residual symptoms for the prediction of relapse (as previously discussed in Chapters 2 and 3).

Another variable selected as potentially important for the prediction of relapse that was also identified in Chapter 3 was unemployment; however, this feature had less relative importance in this study. This may indicate that although unemployment is an important risk factor for relapse following both low- and high-intensity psychological interventions, it is less important for the latter. Further research is needed to prospectively investigate the relationship between unemployment and relapse, and to compare any differential effects between low- and high-intensity treatments.

Another demographic, psychosocial variable, which was not identified as important in any model presented in Chapter 3 but identified as important in this study, was neighbourhood deprivation. Indeed, both developed models selected this feature as important, with it being the most important feature in the secondary model. Previous research has documented associations between socioeconomic deprivation and poor mental health (Fryers et al., 2003; Silva et al., 2016), and also indicated that patients living in deprived areas have poorer treatment outcomes for depression and anxiety (Delgadillo et al., 2016; Finegan et al., 2020). Therefore, this study’s finding that those living in the most deprived areas were at greatest risk of relapse is in line with previous research. However, patients living in the least deprived areas were also suggested to be at relatively more risk of relapse than those living is moderately deprived areas, which conflicts with previous research. Further research is needed to understand how low levels of deprivation could be associated with increased risk of relapse following psychotherapy. Combined, the identification of unemployment and neighbourhood deprivation as potentially important predictors of relapse indicate that the investigation of psychosocial variables should be priorities for future research in this area.

Finally, higher baseline severity in depression, and to a lesser extent baseline anxiety and functioning, was also identified as being potentially important for the prediction of relapse. Specifically, patients with severe depression at the beginning of treatment were identified as being at greater risk of relapse by both models. Previous research has been mixed in terms of the relationship between baseline depression severity and treatment outcomes for CBT, but a recent individual-participant data meta-analysis indicated that baseline severity has little influence on the efficacy of CBT (Furukawa et al., 2017). Therefore, this may suggest that patients with severe depression at baseline may be as likely as patients with less severe depression to recover following acute-phase treatment, but may be more likely to not maintain their treatment gains long-term and thus experience a relapse.

As with the potential individual predictors of relapse following LiCBT identified in Chapter 3, the variables discussed above need to be investigated prospectively in future research to test if any relationships with relapse actually exist, and to fully understand the extent and directions of any such relationships.

**Limitations**

One primary limitation of this study was that not every patient within the service who completed high-intensity psychotherapy was offered or attended a follow-up review appointment. Specifically, it was estimated that only 16.4% of patients (*n*=399/2430) who completed an episode of high-intensity psychotherapy attended at least one follow-up review following treatment. Multiple selection biases related to which patients attended a follow-up review or not were identified, with older patients who attended more treatment sessions and finished treatment with better outcomes being more likely to be followed-up. However, when specifically exploring those patients who completed psychotherapy *with remission of symptoms* (i.e., the subsample specifically explored in this study for the investigation of relapse), fewer selection biases were identified. Nevertheless, two differences between patients who attended follow-up reviews and those who did not were still identified within this subsample, with patients who attended more treatment sessions and finished treatment with higher levels of functioning (i.e., lower WSAS scores) being more likely to be followed-up. Furthermore, considering higher post-treatment WSAS scores was suggested by the two predictive models as being important for the prediction of relapse, it is possible that the potential selection biases present within this study may have resulted in an underestimated rate of relapse. In summary, the provision of follow-up within the investigated IAPT services may not have been systematic, and may have been influenced by certain selection biases, which consequently could influence the results found in this study.

Another limitation of the dataset explored in this study was that there were issues related to data management, with there being inconsistencies in terms of the recording of whether follow-up review appointments occurred or not. Due to this, a number of criteria and assumptions needed to be developed to determine whether a patient received a follow-up review appointment or not. Although these criteria were developed systematically in co-operation with the associated IAPT services, it is not possible to assess their reliability or accuracy. Therefore, it is possible that a number of followed-up patients were not included within the dataset, and that some patients who were not followed-up but instead initiated a new episode of treatment were included. Future research that explores relapse following high-intensity psychotherapy should ensure that the recording of follow-up reviews is reliable and consistent, and should consider following a similar highly-structured methodology to that adopted in the WYLOW study (Ali et al., 2017), which was explored in Chapter 3.

An additional significant limitation of this study was the small sample size and the low event base rate of relapse. Specifically, when exploring only the first follow-up review that patients attended, only *n*=50 relapse cases were identified and could be used for the application of the machine learning approach. Consequently, the application of machine learning in this study was underpowered, and therefore the developed models are highly likely to be unstable and overfitting to the dataset. Indeed, this is demonstrated by the inconsistency between the two models in terms of the features selected as important for the prediction of relapse, and the primary model selecting two computed noise variables as being important. Considering this, the machine learning implementation in this study should be interpreted with caution, and regarded as an exploratory investigation that provides potential areas for future research to investigate further. In addition, another limitation associated with the dataset is that the timing of follow-up review appointments was inconsistent within the services, with some patients having their first review within one month of completing treatment, and other patients having their first review more than six months after finishing treatment. This inconsistency may have affected the relapse rate and identified predictors that were identified in this study.

Finally, there were also two additional limitations of this study that were shared with the study discussed in Chapter 3. First, the developed XGBoost models were not evaluated in an external dataset. Second, the XGBoost models were only developed using routinely collected data, and consequently other potentially important variables could not be explored (e.g., experience of stressful life events, physiological measures etc.).

**Clinical Implications**

Similar to Chapters 2 and 3, this study further highlights the importance of relapse prevention as a core component of psychotherapeutic interventions offered in routine services. Approximately one in five patients were estimated to relapse following high-intensity psychotherapy, with this potentially being an underestimation due to methodological limitations. Consequently, it is important for services and therapists to ensure that relapse prevention and the maintenance of gains long-term is given sufficient attention when discharging patients from care. However, as demonstrated in this study, on the rare occasion when follow-up reviews are provided by IAPT services, a minority of patients are offered or attend such reviews. Follow-up needs to be commissioned as part of the intervention package and offered in a systematic manner. Furthermore, it appears that patients who have better outcomes following acute-phase treatment (i.e., the patients who may be at lowest risk of relapse) are more likely to receive such reviews. Those patients most in need therefore may be being systematically excluded from follow-up support. This limited provision may be a reflection of limited resources or time within services to provide such follow-up reviews. Considering that the follow-up review appointments delivered by the services investigated in this study were equivalent to ‘booster sessions’, where patients receive further therapeutic support, it is important that the patients who are offered these reviews are those who are at greatest risk of relapse. Consequently, as discussed in Chapter 3, the development of prognostic models that are capable of identifying those patients who are greatest risk of relapse following treatment could be highly beneficial to services, as they would allow these patients to be identified and thus provided with such post-treatment support.

Within this study, the secondary (SMOTE-imputed) model, with the application of a diagnostic threshold of 0.5, would be the most credible model to use in clinical contexts, due to the model not selecting any computed noise variables and it having the best balance between the various evaluation metrics. However, this model would not be suitable for clinical use, due to its lack of external validation, and the fact that it was developed with an underpowered sample. Therefore, other prognostic models would be needed to be developed in the future, and this would require more follow-up reviews to be provided by services in order for samples of sufficient sizes to be curated.

**Conclusion**

This study estimated that approximately one in five patients who completed high-intensity psychotherapy with remission of symptoms experience a relapse within months of completing treatment. In addition, similar rates of relapse were found following CBT and non-CBT-based psychotherapy. Furthermore, this study further demonstrates the potential utility of applying machine learning approaches for the purposes of predicting relapse following psychotherapy, with the developed models displaying promising ability to discriminate between relapse and remission cases. However, the models were developed using a small sample and had apparent instability. Consequently, additional models are needed to be developed with more structured and sufficiently sized samples before predictive models such as those developed in this study can be employed within mental health services. Such models could enable services to provide follow-up reviews to patients in a cost-effective manner, whilst also ensuring that they are being provided to patients in greatest need.

**CHAPTER 5**

**A Qualitative Exploration of Patients’ Perspectives of Symptom Deterioration Following Low-Intensity Cognitive Behavioural Therapy**

The two previous chapters in this thesis have demonstrated the potential of applying advanced quantitative techniques, such as machine learning approaches, for the purposes of predicting relapse after cognitive behavioural therapy (CBT). However, there are limitations associated with only using quantitative approaches for the investigation of relapse. For instance, although quantitative approaches may be able to *identify* potential associations between risk factors and relapse, they are less capable at providing an understanding as to *how* and *why* associations exist (Bryman, 2012). Qualitative research is a potential avenue for helping to address this limitation, as it is able to contextualize observed findings within the experiences of participants’ lives (McLeod, 2011). This allows for potential reasons behind findings to be understood and for insights regarding potential clinical implications to be established (Midgley et al., 2014). Qualitative research has therefore been argued to be an important enrichment to quantitative outcome research (Dattilio et al., 2010; Midgley et al., 2014).

Considering this, qualitative research can enrich quantitative relapse research in two ways. Firstly, it enables patients’ perspectives on the processes underlying change to be explored, with such perspectives not being considered in purely quantitative outcomes research (McLeod, 2011). As patients are “the site of change” (Greenberg, 1991, p. 10) during treatment, their perspectives are therefore a valuable source for researchers to obtain new information regarding change processes and protective/risk factors (Elliott, 2010). Therefore, qualitative investigations of relapse may enable new important risk and protective factors to be identified.

Secondly, qualitative research can also enrich quantitative relapse research by exploring outcome measurement. For instance, although construct validity is a fundamental aspect of measurement (Cronbach & Meehl, 1955), it has been argued that quantitative researchers often take for granted the validity of instrument-based approaches to therapy outcome research (McLeod, 2011). This may be the case for quantitative relapse research, with a wide variety of different relapse measures being used (Chapter 2; Wojnarowski et al., 2019). This lack of established operationalization suggests some uncertainty regarding how best to define and assess relapse, raising questions regarding the construct validity of relapse measures. Therefore, adopting a qualitative approach may provide one avenue through which the validity of relapse measurements can be examined.

To date, no qualitative research has explored how and why relapse occurs following CBT, or how relapse is perceived by patients. Therefore, this study aimed to address this gap in the literature by adopting a qualitative approach to investigate symptom deterioration following low-intensity cognitive behavioural therapy (LiCBT). Specifically, this study explored responses from participants of the WYLOW study (see Chapter 3 for a summary) who had been classified as having relapsed, and thus aimed to extend upon and enrich the quantitative findings of Ali et al. (2017). There were three primary aims of the study. First, to examine the validity of the statistical relapse measure used by the WYLOW study, by exploring whether participants agreed or disagreed with their relapse classification. Second, to investigate what factors participants personally attributed to their increases in symptoms. Third, to explore what actions participants were planning on taking to help alleviate their distress.

**Method**

**Design**

This study analyzed data collected in the WYLOW study (Ali et al., 2017; Delgadillo et al., 2018b; REC reference: 12/YH/0095). Data from this naturalistic, prospective, longitudinal cohort study were also investigated in Chapter 3, and fuller study details can be found there. To summarise, *N*=439 patients who had completed LiCBT with remission of depression and anxiety symptoms were followed-up on a monthly basis for up to 24 months. Participants classed as having relapsed/experienced a recurrence during the follow-up period were provided with self-help information and a qualitative feedback questionnaire (see measures section), and supported to re-engage with mental health services. Although the WYLOW study collected data for up to 24 months’ post-treatment, this study only analyzed data from the initial 12 months’ period, as the primary aim was to explore the experiences of relapse (<12 months), not recurrence (≥12 months).

**Participants**

Of the *N*=439 participants in the WYLOW study, *n*=223 were classified as having relapsed during the 12-month follow-up period, *n*=94 were classified as having remained in-remission throughout follow-up, and *n*=122 were lost to follow-up. Of the *n*=233 relapse cases, *n*=45 returned a completed feedback questionnaire that could be linked to the study data, and this was this study’s primary sample. In terms of primary presenting problems, the most common recorded was mixed anxiety and depression (42%), followed by depressive episode (24%), generalized anxiety disorder (7%), and finally recurrent depression (2%). The primary presenting problem was unspecified for the remaining 24% of participants. At the beginning of treatment, participants recorded mean scores of 14.23 (*SD* = 4.89) and 13.47 (*SD* = 4.63) on the PHQ-9 and the GAD-7, while, at the end, they recorded average scores of 3.82 (*SD* = 2.05) and 3.56 (*SD* = 1.88). The sample was characterized by a majority of female patients (69%) from a white British background (91%) with a mean age of 43 (*SD* = 14.8). Five participants were unemployed at the start of treatment, while four were unemployed at the end (of whom, three were also unemployed at the start). Four participants had a self-reported disability, and 13 had a self-reported long-term medical condition (e.g., asthma, diabetes, etc.). Participants accessed a mean of seven treatment sessions (*SD* = 2; range = 2-11). Demographic and clinical information for each participant is summarized in Table 5.1.

**Measures**

***Relapse Definition***

As reported in Chapter 3, two self-administered outcome measures for depression and anxiety were used to determine the occurrence of relapse: the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001) and the Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006). A reliable change index of ≥5 for both the PHQ-9 (McMillan et al., 2010) and the GAD-7 (Richards & Borglin, 2011) was applied to monitor deterioration over time. To be classed as having relapsed, patients must have met two criteria: (a) at least one of their PHQ-9 or GAD-7 scores was above the clinical cut-off at a monthly follow-up review; *and* (b) the measure above the clinical cut-off also represented statistically reliable deterioration relative to the score observed at the final treatment session (i.e., an increase greater or equal to the reliable change index).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 5.1**  *Selected Demographic and Clinical Information of Participants* | | | | | |
| Participant | No. of LiCBT sessions | Follow-up month in which relapse occurred | Age | Unemployed? | Residual depressive symptoms (PHQ-9 ≥ 5) |
| P1 | 7 | 3 | 23 | No | Yes (5) |
| P2 | 6 | 4 | 43 | No | No (4) |
| P3 | 5 | 9 | 63 | Start | No (3) |
| P4 | 4 | 1 | 56 | No | No (4) |
| P5 | 8 | 1 | 59 | Start and end | No (4) |
| P6 | 6 | 2 | 52 | No | Yes (5) |
| P7 | 11 | 4 | 47 | No | No (3) |
| P8 | 5 | 1 | 45 | No | No (1) |
| P9 | 7 | 1 | 19 | Start and end | Yes (6) |
| P10 | 7 | 1 | 47 | No | No (2) |
| P11 | 9 | 2 | 64 | No | No (3) |
| P12 | 8 | 10 | 52 | No | No (1) |
| P13 | 9 | 4 | 45 | No | No (4) |
| P14 | 5 | 3 | 34 | No | Yes (5) |
| P15 | 6 | 1 | 64 | No | Yes (7) |
| P16 | 9 | 3 | 28 | No | Yes (5) |
| P17 | 10 | 3 | 25 | No | No (3) |
| P18 | 7 | 1 | 20 | No | Yes (9) |
| P19 | 10 | 9 | 54 | No | Yes (6) |
| P20 | 4 | 2 | 58 | No | Yes (6) |
| P21 | 8 | 8 | 39 | No | No (3) |
| P22 | 8 | 7 | 23 | Start | Yes (5) |
| P23 | 10 | 11 | 50 | End | Yes (5) |
| P24 | 5 | 2 | 63 | No | No (0) |
| P25 | 7 | 2 | 35 | No | Yes (6) |
| P26 | 7 | 3 | 49 | No | No (2) |
| P27 | 6 | 1 | 18 | Start and end | Yes (5) |
| P28 | 10 | 3 | 47 | No | No (3) |
| P29 | 9 | 2 | 20 | No | Yes (7) |
| P30 | 7 | 1 | 38 | No | Yes (5) |
| P31 | 7 | 5 | 60 | No | No (2) |
| P32 | 8 | 7 | 42 | No | No (4) |
| P33 | 11 | 3 | 24 | No | No (3) |
| P34 | 5 | 2 | 33 | No | No (4) |
| P35 | 2 | 2 | 17 | No | Yes (7) |
| P36 | 8 | 8 | 68 | No | No (4) |
| P37 | 7 | 3 | 33 | No | No (0) |
| P38 | 7 | 1 | 55 | No | No (4) |
| P39 | 8 | 12 | 43 | No | No (0) |
| P40 | 8 | 11 | 29 | No | No (0) |
| P41 | 8 | 2 | 34 | No | Yes (6) |
| P42 | 9 | 1 | 41 | No | No (3) |
| P43 | 7 | 3 | 64 | No | No (4) |
| P44 | 7 | 1 | 49 | No | No (3) |
| P45 | 8 | 6 | 43 | No | No (2) |

***Feedback Questionnaire***

The feedback questionnaire sent via post contained four questions. The first two questions were designed to examine whether participants agreed with their relapse classification. The first question (Q1) was worded as follows: “*The questionnaire scores for this month suggest a worsening in your symptoms of depression and/or anxiety. Do you consider that you have relapsed back into depression and/or anxiety?*” This question was multiple choice, with the following three options: “Yes”; “No”; or “Not sure/maybe”. The second question (Q2) sought to understand participants’ responses to Q1, and was worded as follows: “*Please describe the reasons for the response you gave above*”. This question was open-ended, and the number of words written in response ranged from 4-89 (*M* = 33). The third question (Q3) was also open-ended, with this question exploring what factors participants attributed to their increase in symptoms. The question was worded as follows: “*What do you think are the reason(s) for the apparent increase in your depression and/or anxiety symptoms?*” The number of words written in response to Q3 ranged from 0-166 (*M* = 22). Finally, the last question (Q4) aimed to understand what participants were doing to manage their increase in symptoms, and was worded as follows: “*What will do you to cope with the apparent increase in your depression and/or anxiety symptoms?*” This question was also open-ended, and the number of words written in response ranged from 0-137 (*M* = 19).

**Theoretical Framework**

The philosophical assumptions underpinning this study are rooted in a critical realist paradigm (Bhaskar, 1975). Similar to positivism, critical realism argues that there is a real, objective world that operates independently of our knowledge. In contrast, however, critical realists propose that the world can be distinguished between an ‘observable level’ and a ‘real level’ (Fletcher, 2017). Within the former level, events are experienced and these events can be observed empirically, however the interpretation of these events is influenced by human perceptions and social ideas. Meanwhile, the ‘real level’ comprises of unobservable causal structures or mechanisms, which cause events at the ‘observable level’ to occur. Consequently, it is the goal of critical realism research to understand the social world by understanding the unobservable causal mechanisms that produce observable events. Based on this, we assumed that participants experienced an observable period of increased distress, and aimed to understand the unobserved mechanisms that generated this experience, and what this event meant for participants.

**Analysis**

***Potential Selection Biases***

To begin, potential sources of selection bias were investigated by comparing characteristics between those patients who returned a completed feedback questionnaire following their classification of relapse (i.e., those patients who constituted the primary sample for this chapter’s study; *n*=45) and those patients of the wider WYLOW study who did not return a completed feedback questionnaire following their relapse classification (*n*=178). The specific characteristics compared were age, gender, ethnicity, unemployment at the beginning of treatment, unemployment at the end of treatment, number of attended sessions, PHQ-9 scores at end of treatment, GAD-7 scores at end of treatment, and WSAS scores at end of treatment. Pearson’s chi-square tests were used to compare categorical data, while Welch’s t-tests or Mann-Whitney U tests were used to compare continuous variables, depending on distributions.

***Primary Research Aims***

This study followed a framework analysis approach to analyze the data (Ritchie & Spencer, 2002). The advantages of framework analysis are that it offers a flexible, systematic, and comprehensive approach to data analysis, which also enables comparisons both between and within individual cases.

There are five stages to framework analysis. The first stage is *familiarization*, where the researcher immerses themselves in the qualitative data by reading through the transcripts, and becoming aware of key ideas and recurring themes. The second stage is *identifying a thematic framework*, where emerging themes are used to form a thematic framework. The framework that is developed at this stage is provisional, with there being further opportunities to refine it at subsequent analytical stages – demonstrating the dynamic, flexible nature of the approach. Following this stage is *indexing,* where the thematic framework is applied to all transcripts, identifying specific portions of the data that correspond to a particular theme. The fourth stage is *charting*, in which the indexed portions of data are ‘charted’ into a framework matrix, which consists of rows (cases), columns (thematic codes), and cells of qualitative data. The development of this structure allows researchers to systematically reduce and organize the data, making it more manageable ahead of the final stage of the framework analysis approach, *mapping and interpretation*. In this stage, the researcher, guided by the study’s research questions, attempts to understand the dataset by analyzing the key characteristics of the charted data. This five-stage process was followed three times to address each of the study’s aims. To provide additional information and clarity regarding the magnitude of each theme, discussions of each theme and sub-theme were counted across participants. This is similar to the process that is often undertaken in qualitative content analysis (Bengtsson, 2016; Morgan, 1993).

Data analysis was independently conducted by two researchers; one of whom was an academic researcher, and the other a senior psychological wellbeing practitioner (PWP) and educator. Both analysts independently applied the first two stages of framework analysis (familiarization and identifying a thematic framework), before meeting to compare thematic frameworks, and agree upon a shared framework to adopt for later stages. Following this, both analysts independently applied the final three stages of framework analysis (indexing, charting, and mapping and interpretation). Once this was completed, they analysts met again to compare their produced charts, and agree upon finalized charts and interpretations. Any disagreements or uncertainties that arose from discussions between the two analysts were taken to a supervising team of two different researchers for further discussion in order for consensus to be reached.

***Secondary Analyses***

The final framework matrices were explored further with reference to three specific features identified in Chapter 3 using the same primary study dataset as being predictive of relapse. The predictive models developed in Chapter 3 suggested that patients who were younger than 35 years old, unemployed, and/or exhibited residual depressive symptoms at the final treatment session (PHQ-9 scores ≥ 5)[[15]](#footnote-15), were at relatively greater risk of relapse than patients who did not possess those characteristics. Therefore, each of the developed frameworks were divided in a stratified manner according to these characteristics, and then investigated to explore if there were any thematic differences between these groups.

**Results**

**Analysis of Potential Selection Biases**

Comparisons of the characteristics of patients in the wider WYLOW study who did (*n*=45) and did not (*n*=178) return a completed feedback questionnaire following their classification of relapse indicated that there were no significant differences between the two groups in terms of the investigated characteristics (all p>0.10).

**Qualitative Analysis**

The themes identified from each participant’s explanation of their agreement/ disagreement with their relapse classification, as well as the risk factors of symptom increases and coping strategies discussed by each participant, are summarized in Appendix H.

***Relapse Classification Agreement***

Of the study’s 45 participants, 21 (47%) agreed with their relapse classification. Fifteen (33%) were unsure, while 9 (20%) did not believe they had relapsed. The identified themes related to the reasons behind patients’ perspectives of their relapse classifications varied between the three groups, and these are illustrated with example quotes in Table 5.2.

The majority (*n*=16) of ‘agreeing participants’ discussed *worsening or persistent symptoms* as a reason behind their agreement, with participants mentioning a wide variety of specific symptoms (e.g., trouble concentrating, worrying, etc.). Additionally, multiple ‘agreeing participants’ (*n*=13) discussed *specific life factors* that they attributed to their increases in symptoms, while four also believed they had relapsed due to an *impairment on everyday functioning*.

|  |  |  |
| --- | --- | --- |
| **Table 5.2**  *Qualitative Themes Related to the Reasons Behind Patients’ Perspectives of their Relapse Classifications* | | |
| Relapse Perspective | Theme | Example Quote |
| Agree | Worsening or Persistent Symptoms | *I felt all of the physical, psychological + emotional symptoms I felt last time I suffered from anxiety . . . I felt these for about 4 weeks straight + I have been feeling them on and off for a further 4 weeks.* (P40) |
| Specific Life Factors | *Drink, anniversaries of deaths, work* (P23) |
| Functioning Impaired | *Struggling to cope with day to day tasks.* (P21) |
| Temporary Deterioration | *she decided to have an abortion which actually made me feel a lot better … ...I'm now feeling a lot better, almost on a high actually* (P22) |
| Unsure | Worsening or Persistent Symptoms | *sometimes when things are difficult I do still relapse into depression. And if I have difficult decisions to make it can confuse me which then leads me into a phase of anxiety. (P27)* |
| Specific Life Factors | *My workload is fluctuating at the moment and I have been busy* (P2) |
| Temporary Deterioration | *I always experience ups and downs in my mood in such a way. Feelings come + go and can be very intense + rapid. But I feel better now.* (P18) |
| Partial Relapse | *I would have thought your survey would have included everybody, not just those who don't relapse a little* (P10) |
| Potential Medical Issue | *I hope it may be a medical issue ... I am awaiting test results for thyroid + hormone imbalance.* (P16) |
| Disagree | Temporary Deterioration | *I think I was just having a bad couple of weeks like most people. Feel fine again now*. (P37) |
| Other Specific Life Factors | *Circumstances at work were causing the stress and anxiety symptoms.* (P19) |

The 15 ‘unsure participants’ and nine ‘disagreeing participants’ also commonly discussed *worsening or persistent symptoms* and *specific life factors*. However, they also often reported that their symptom increases were *temporary*, and therefore did not represent a permanent relapse (*n*=7 ‘unsure participants’; *n*=7 ‘disagreeing participants’). Instead, symptom increases were perceived as being a fluctuating part of life, part of the recovery process, or simply reflective of the participant having a ‘bad week’. Indeed, most of the participants who described their symptom increases as being temporary explicitly stated that they were feeling better at the time of response (*n*=4 ‘unsure participants’; *n*=6 ‘disagreeing participants’). Additionally, three of the six ‘disagreeing participants’ also discussed that receiving therapeutic support (i.e., further therapy, a mindfulness course, or a wellbeing course) during follow-up had helped to alleviate their distress.

Three ‘agreeing participants’ also expressed comments related to the idea that symptom increases are temporary. For example, one mentioned that they were hoping that their recent stressful period was “*just a bad patch*”. Meanwhile, another discussed that they had experienced a highly stressful situation that had caused a relapse, but they felt substantially better once the situation had been resolved. The third participant explicitly stated their relapse was temporary, as their symptoms had been relieved after receiving additional pharmacological treatment.

There were two other reasons why some participants were unsure about their relapse classification; three participants believed they had only experienced a partial relapse, while two participants suggested potential medical issues may have been behind their symptom increases.

***Perceived Risk Factors***

Three themes were identified within participants’ discussions of potential reasons behind their symptom increases: psychosocial factors; stressful events; and cognitive and behavioural factors. These three themes, alongside their 10 subthemes, are presented in Table 5.3. Five participants did not discuss any risk factors.

**Psychosocial Factors.**Psychosocial risk factors were discussed by *n*=32 participants. Work-related stress was the most common subtheme, with *n*=24 discussing a variety of work-related issues. For example, seven participants discussed changes at work, such as fluctuating workloads or retraining of roles, while seven participants also discussed having high job demands. Other work-related issues included: poor work atmosphere (*n*=3); returning to work after time off (*n*=2); uncertainty after leaving job (*n*=2); lack of clarity about priorities (*n*=1); lack of career progression (*n*=1); poor work performance (*n*=1); failing at promotions (n=1); caring for family members while working (*n*=1); and unemployment (n=1).

The second most common psychosocial risk factor expressed by participants was social relationship issues, with *n*=13 participants discussing this subtheme. Family issues (*n*=9) were more commonly discussed than romantic/sexual relationship issues (*n*=5)[[16]](#footnote-16), while one participant discussed friendship issues. A commonly discussed issue (*n*=7) was caring for family members who had physical or mental health problems, and/or taking care of children.

Two other psychosocial sub-themes were discussed by patients as being risk factors: education-related issues (e.g., exam stress, heavy workload, completing PhD thesis etc.; *n*=5); and issues related to living environment or security (e.g., housing issues, homesickness, immigration bureaucracy, etc.; *n*=4).

**Stressful Events.**The second most common theme (*n*=22) regarding risk factors was the occurrence of stressful events during follow-up. These events could be categorized as being either ‘unanticipated’ (*n*=13), or ‘anticipated and/or organized’ (*n*=11). Only two of the 22 participants reported they had experienced both types of event.

In terms of unanticipated stressful events, (potential) periods of poor physical health were the most commonly discussed (*n*=7), with participants reporting ill health, surgical operations, and a perceived chemical imbalance. The second most commonly discussed unanticipated event was the death of a relative (*n*=3). Meanwhile, in terms of anticipated and/or organized events, the most commonly reported was a return to work/college (*n*=3), with the Christmas holidays being the second most commonly discussed event (*n*=2).

|  |  |  |
| --- | --- | --- |
| **Table 5.3**  *Qualitative Themes Related to the Risk Factors that Participants Attributed to their Symptom Increases* | | |
| Theme | Subtheme | Example Quote |
| Psychosocial Factors | Work-Related Stress | *Work stress increase → longer hours, more pressure to hit higher targets.* (P16; unsure) |
| Social Relationship Issues | *Issues with family - brother in law - alcohol issue (worried about my sister & nieces) - father unwell.* (P45; disagreed) |
| Education | *returning to college, thinking about university… Problems thinking about the future. (P35; agreed)* |
| Living Environment/Security | *tenancy ending through no fault of my own and trying to find a new home (P17; agreed)* |
| Stressful Events | Unanticipated | *Within last 2 weeks family member died* (P13; unsure) |
| Anticipated/Organised | *I had just terminated my previous employment (my choice) and I had a degree of uncertainty ….this was just a natural reaction to a very significant and life changing event.* (P7; disagreed) |
| Cognitive and Behavioural Factors | Poor Self-Efficacy | *Not believing in myself, lack of confidence and having so much going on in my head.*  (P17; agreed) |
| Stopped Using LiCBT Strategies | *At this point I had not been using any of the CBT tips + stress relievers on a daily/weekly basis as a preventative as I thought I was "better" and that my anxiety was unlikely to come back.* (P40; agreed) |
| Limited Activites | *Because I am not working and not being able to do the things I used to do.* (P3; agreed) |
| Substance Use | *Cannabis use spending a lot of time alone having paranoid, degrading thoughts/voices in head*  (P35; agreed) |

**Cognitive and Behavioural Factors.**Some participants (*n*=9) also discussed cognitive and behavioural factors as being behind their symptom increases. Interestingly, every participant who reported a cognitive and behavioural factor agreed with their relapse classification. The most commonly discussed risk factor related to this theme was poor self-efficacy, with *n*=3 participants reporting low self-esteem or confidence. Other risk factors related to this theme included: the inability to do certain activities; no longer using CBT resources; and substance use (each *n*=2).

***Coping Strategies***

Two primary themes were identified in relation to coping strategies: *external strategies*; and *internal strategies*. These two themes, alongside their nine subthemes and example quotes, are presented in Table 5.4. Six participants discussed no strategies.

**External Strategies.**The majority of participants (*n*=34) discussed at least one coping strategy that would involve them seeking external support. The most commonly reported coping strategy was seeking further healthcare support or information (*n*=24; 53%). This included: returning to a General Practitioner (GP) for medical guidance (*n*=15); receiving further psychotherapy (*n*=13); starting/continuing/increasing pharmacological support (*n*=11); and seeking other therapeutic support (i.e., a wellbeing or mindfulness course; *n*=2). Participants who agreed they had relapsed were more likely to discuss seeking further healthcare support/information (*n*=14/21) than participants who were unsure (*n*=7/15) or disagreed (*n*=3/9). Three other external strategies were reported: seeking social support (e.g., talking to family/friends; support groups; *n*=7); seeking support from work (*n*=2); and simply waiting for a stressful event to end (*n*=7).

**Internal Strategies.**Nearly half of participants (*n*=22) discussed implementing coping strategies that only require individual action from themselves. For instance, *n*=9 participants reported that they would attempt to relax more. This included actions such as practicing mindfulness/meditation, taking time off work, and ensuring they had time for themselves. Other common internal strategies reported by participants included: using techniques or resources that were attained during LiCBT (*n*=8); exercising more and/or being healthier (*n*=7); and continuing as they were previously doing (*n*=4). Additionally, two participants discussed other non-health related distraction behaviours, including doing activities that occupy the mind (*n*=1), and gambling (*n*=1).

|  |  |  |
| --- | --- | --- |
| **Table 5.4**  *Qualitative Themes Related to the Coping Strategies that Participants Discussed Employing to Address their Symptom Increases* | | |
| Theme | Subtheme | Example Quote |
| External Strategies | Seeking Further Healthcare Support | *I am seeing my GP regularly & am probably going to start medication this week. I am also contacting the Mental Health Team.* (P44; agreed) |
| Seeking Social Support | *Keep speeking to people that are aware of whats going on.* (P29; unsure) |
| Seeking Support from Work | *Spoken to work for support* (P21; agreed) |
| Await End of Stressful Event | *Those things were transient ↑, exams are finished. I am at home now + off work* (P18; unsure) |
| Internal Strategies | Relax More | *I have started meditating for 15 mins everyday which really helps relieve stress and my regular tension headaches… I always have 1 full day off per week, no matter how much work I have on.* (P40; agreed) |
| Use LiCBT Strategies | *Use the tools and techniques I've picked up as part of the support I've been given this year.* (P7; disagreed) |
| Exercise and Be Healthier | *Trying to relax/eat healthy/exercise (P21; agreed)* |
| Continue as Previously Doing | *None at all…just keep working to the light at the end of the tunnel!* (P4; disagreed) |
| Other Non-Health Related Distraction | *I've also started gambling I'm feeling much better.* (P22; agreed) |

‘Disagreeing participants’ were more likely to report the use of at least one internal coping strategy (*n*=6/9) than an external strategy (*n*=5/9), while “agreeing participants” and “unsure participants” were both more likely to report external strategies (*n*=17/21, *n*=12/15, respectively) than internal strategies (*n*=11/21, *n*=5/15).

**Thematic Differences: Vulnerable Groups**

***Young Age***

Younger participants (*n*=16 were ≤35 years old) were more likely to believe that they had relapsed, with only *n*=1 younger participant disagreeing with their relapse classification, while *n*=10 agreed and *n*=5 were unsure. In contrast, older participants were more divided, with *n*=11 agreeing, *n*=10 unsure, and *n*=8 disagreeing.

Psychosocial factors remained the most common form of risk factor reported by both groups. Similarly, work-related stress was also the most frequently discussed factor for both groups, with there being no discernible differences in terms of subthemes. In contrast, there were noticeable differences in terms of social relationship issues discussed by both groups (*n*=5 younger participants; *n*=8 older participants). Specifically, six of the seven participants who discussed caring for family members as a risk factor were older, with the social relationship issues discussed by younger participants more likely to be related to romantic/sexual relationships. Indeed, only one younger participant, but every older participant, discussed family issues. There were also differences in terms of the remaining two psychosocial factors subthemes of education, and living environment and security. Specifically, all five reports of education-related issues, and four of the five reports of issues related to living environment and security, were reported by younger participants.

Stressful events remained the second most common theme reported by both groups, with the only slight difference being that younger participants were more likely to discuss anticipated or organized stressful events than older participants (*n*=6 vs *n*=5). Younger participants were also more likely to report cognitive and behavioural risk factors (*n*=5 vs *n*=4), with all three reports of poor self-efficacy being made by younger participants. There were no discernible differences between age groups in terms of coping strategies.

***Unemployment***

Only six participants in this study’s sample were unemployed at the beginning and/or end of their treatment. One participant who was unemployed at the end of treatment discussed work-related stress as a risk factor, thus indicating that they had since gained employment, while another participant who was unemployed at the beginning of treatment, but employed at the end, stated they were no longer working.

None of the participants disagreed with their relapse classification, with *n*=5 agreeing and *n*=1 unsure. Due to the small sample, only a few risk factors were discussed. The inability to do certain activities was the only specific risk factor reported by more than one participant (*n*=2). In fact, no other participant across the whole sample discussed this risk factor. The only coping strategy discussed by multiple participants who had experienced unemployment was seeking further healthcare support/information (*n*=5).

***Residual Depressive Symptoms***

Within the sample, *n*=17 had residual depressive symptoms at their final treatment session (PHQ-9 ≥ 5), while *n*=28 did not. There were no differences between the groups in terms of the perceptions of relapse classifications (participants with residual symptoms: *n*=8 agreed, *n*=6 unsure, *n*=3 disagreed; participants without residual symptoms: *n*=13 agreed, *n*=9 unsure, *n*=6 disagreed), or the reasons underlying these perceptions. In terms of risk factors, there were only two noticeable differences, with participants with residual symptoms being less likely to discuss social relationship issues (*n*=3 vs *n*=10), or stressful events (*n*=6 vs n=16), than participants without residual symptoms.

Participants with residual symptoms were substantially more likely to report the use of external coping strategies (*n*=16/17) than internal strategies (*n*=6). In contrast, participants without residual symptoms discussed using internal strategies (n=16/28) and external strategies (n=18) at a similar level. Indeed, no participants with residual symptoms discussing the use of LiCBT techniques/resources, while n=8 participants without residual symptoms discussed this strategy. One similarity between the groups however, was that seeking further healthcare support/information was the most frequently reported coping strategy for both groups (*n*=10 participants with residual symptoms, *n*=14 participants without residual symptoms).

**Discussion**

This is the first study to explore patients’ experiences of symptom deterioration following a course of low-intensity CBT for depression and anxiety. Fewer than half of the participants (47%) explicitly agreed that they had experienced a relapse, with one-fifth disagreeing. Participants who were unsure or disagreed that they had experienced a relapse frequently viewed their symptom increases as being temporary, representing a normal part of life, a part of the recovery process, or simply ‘a bad week’. These findings raise questions regarding the validity of the statistical criteria applied in this study, and in other studies, to determine the occurrence of relapse (i.e., a combination of clinical cut-offs in the PHQ-9 and GAD-7 questionnaires, and reliable change indices).

Discussions of symptom increases as being temporary potentially indicate that the applied outcome measure assesses the occurrence of a *lapse*, rather than a *relapse*. Chellingsworth et al. (2013) defined a *lapse* as being a brief return to feeling low or anxious, and/or problematic behaviours, and a normal part of life. In contrast, they defined a *relapse* as occurring when negative thinking and problematic behaviours return, and are maintained, over a longer period of time. The key difference between the two definitions is therefore the temporal aspect. The relapse measurement criteria applied in the WYLOW study only examined symptom scores at one point in time, and therefore lacked this temporal component. Consequently, rather than specifically assessing the occurrence of a relapse, this approach may also be evaluating the occurrence of a lapse.

An important element of LiCBT is that patients learn how to prevent symptom increases from escalating into relapses, by learning how to regularly monitor and understand their mental health, predict and recognize when problems arise, and have strategies in place to help deal with these problems (Bennett-Levy et al., 2010; Lovell & Richards, 2012). As many participants discussed their symptom increases as only being temporary, this would suggest that these individuals were able to effectively manage and ameliorate their symptom increases. This ability may therefore represent an outcome of *effective* therapy, as the participant has demonstrated they have developed valuable strategies for managing their mental health. This would contradict these participants’ initial relapse classifications, which indicate that participants gained little long-term benefit from therapy. Therefore, if relapse classifications are overestimating the rate of relapse, as their criteria may also be including participants who have only experienced *lapses*, this may suggest that the long-term effectiveness of interventions is potentially being underestimated.

Another issue with having an imprecise relapse measure is that it limits our ability to accurately predict its occurrence. This ability is also limited by a current lack of understanding regarding relapse risk factors (see Chapter 2; Wojnarowski et al., 2019). This study aimed to help address this gap in knowledge by exploring the factors that participants personally attributed to their symptom increases. A variety of risk factors was reported; however, most were psychosocial in nature. This highlights the importance of clinicians taking into consideration the psychosocial aspects of patients’ lives during treatment, and particularly when discussing relapse prevention. The most common risk factor reported was work-related stress, with over half of participants discussing this factor. Work-related stress being such a frequently discussed risk factor is interesting, considering that unemployment was identified as being an important predictor of relapse in Chapter 3, and to a lesser extent in Chapter 4 as well. This may indicate that although work is generally beneficial to mental health (Waddell & Burton, 2006), especially in relation to unemployment, it may be that certain specific factors of work can also be detrimental to mental wellbeing. Indeed, multiple systematic reviews have identified several work stressors (e.g., high job demands, low job control, etc.) that have significant adverse effects on mental health (Bonde, 2008; Law et al., 2020; Netterstrøm et al., 2008; Nieuwenhuijsen et al., 2010; Stansfield & Candy, 2006). Overall, extreme aspects of work appear to be associated with a risk of relapse: unemployment on one end, and highly stressful work conditions at the other end.

Another psychosocial risk factor commonly discussed by participants was social relationship issues, with nearly a third of participants discussing this factor. The issues discussed ranged from being related to family members (e.g., parents, siblings, children) to romantic/sexual relationships and friendships. In particular, caring for a family member was an issue raised by multiple participants. Understandably, the type of relationship issue discussed by participants varied according to age. Younger participants were more likely to discuss having romantic/sexual relationship issues, while older participants were more likely to discuss having family issues, and these issues were frequently related to caring for family members. These findings support previous research that has highlighted the important role that social relationships have in terms of mental health and well-being. For example, recent reviews have demonstrated that high levels of perceived social support and large, diverse social networks have protective effects against depression (Santini et al., 2015; Wang et al., 2018). Participants also discussed the occurrence of stressful events as being a reason for their symptom increases, with these also being predominantly psychosocial in nature. This finding is consistent with previous research that has demonstrated that the occurrence of stressful events is predictive of both the onset and relapse of depression and anxiety (Harkness et al., 2014; Heldt et al., 2011; Kendler et al., 1999; Kessler, 1997).

This study also explored the coping strategies that participants adopted to help manage their symptom increases, and a range of actions were reported. Participants were more likely to discuss external coping strategies (i.e., where they seek support externally from themselves), than internal strategies (i.e., strategies that only require individual action from themselves). Indeed, the most common strategy discussed was seeking further healthcare support, with over half of participants reporting this action. Participants who agreed they had relapsed were more likely to discuss seeking healthcare support than those who were unsure or disagreed with their relapse classifications. This highlights the potential health economic problems that the occurrence of relapses can produce. Specifically, symptomatic deterioration often results in patients considering to seek further healthcare support, thus creating a ‘revolving door’ cycle in mental healthcare services (Roscoe, 2019). However, it may be that participants discussed this strategy as, for ethical reasons, they were supported to re-engage with mental healthcare after being classified as having relapsed. Nevertheless, many patients, and especially those who believed they had relapsed, still considered receiving further healthcare support following an increase in symptoms.

However, although external coping strategies were more commonly discussed, nearly half of participants also discussed using internal strategies. Indeed, the second most common coping strategy reported was the use of LiCBT techniques or resources. This indicates that these participants found these skills and resources helpful to begin with and also perceive them to be potentially useful in managing their current symptoms. Furthermore, patients are encouraged to reuse LiCBT techniques when their mood becomes low (Chellingsworth et al., 2013; Lovell & Richards, 2012), and it is therefore promising that these participants were engaging in this aspect of relapse prevention. Yet, it is also concerning that only eight participants of 45 explicitly discussed this coping strategy, and that no participants who displayed residual depressive symptoms at the end of treatment reported it. This would suggest that many patients were not engaging in an important aspect of relapse prevention. However, it may be that participants did in fact discuss specific aspects of their relapse prevention plan, without explicitly stating they were using specific techniques of LiCBT. For example, some participants discussed seeking social support, exercising, or seeking support from work or taking time off as coping strategies, and these actions may have been discussed during LiCBT and been an important element of their relapse prevention plan. Indeed, these actions are often emphasized to participants as being important in order to stay well (Lovell & Richards, 2012).

Frequent reports of external coping strategies by patients may reflect that such patients possess an external locus of control (i.e., the belief that life events result from external factors beyond their control), as they mostly react to symptom increases by seeking external support. In contrast, discussions of internal coping strategies may reflect some degree of internal locus of control (i.e., the belief that they have control over the outcome of life events). Previous research has demonstrated that an external locus of control is associated with depression and anxiety (Cheng et al., 2013), and one potential explanation for this is that internal locus of control helps motivate people to actively engage in strategic coping behavior (Ryan & Deci, 2000). This may be reflected in this study, as patients who believed they had relapsed were more likely to discuss using external strategies, while patients who did not were more likely to discuss using internal strategies. Moreover, patients with residual depressive symptoms at the end of treatment were more likely to discuss using external coping strategies, while patients without residual symptoms discussed both types of strategy at a similar frequency. Residual symptoms may therefore reflect the unsuccessful acquisition of an internal locus of control, resulting in patients not feeling confident in applying internal coping strategies. This may potentially explain why residual symptoms is an apparent risk factor of relapse after CBT, and further research could explore this potential mechanism.

**Reflexivity Considerations**

Inevitably, our analysis and interpretation of the questionnaire data will have been influenced by our personal experiences and philosophical orientations. Our theoretical reflections are rooted in our particular schools of thoughts, including cognitive behavioural therapy, and acceptance and commitment therapy. Furthermore, we primarily adopt quantitative research methods and paradigms when conducting research, and this grounding has likely influenced our approach and interpretations within this study. Therefore, the theoretical reflections that we have discussed thus far must be considered with our particular theoretical orientations, and the following study limitations, in mind.

**Limitations**

The primary limitation of this study was the use of questionnaires for data collection, rather than interviews. Although the use of questionnaires decreases participant and researcher burden, and thus enables more participants to be included and therefore more perspectives to be taken into account, these benefits come at the expense of depth and focus. Indeed, many of the responses that participants provided were vague or general, such as when work-related or social relationship stressors were discussed. The questionnaire approach prevented more information related to a participant’s response from being identified, such as information regarding what specific issues participants had experienced, and how such issues were involved in increasing symptoms. Additionally, this approach also prevented potential misunderstandings of questions to be addressed. For example, many participants discussed specific factors in their lives when asked to explain why they believed or did not believe they had relapsed, potentially not realizing that another question would be asking about risk factors. Such misunderstandings may have prevented participants from providing more focused explanations. Moreover, the questionnaire used in this study was also limited by its focus on risk factors, as an additional exploration of protective factors may have provided greater insight. For example, participants who reported negative aspects of work may also have had several positive aspects of their jobs to report, but they were not asked about these.

Moreover, it could also be argued that the question asking participants whether they believed they had relapsed or not was potentially leading, as it states that their previous clinical score indicated that they had experienced “a worsening in [their] symptoms”. Some participants may have perceived this as meaning that they had experienced a relapse, and therefore they may have been more likely to report a belief that they had relapsed. However, this would mean that although participants may have been potentially led to believe they had relapsed, less than half reported such a belief. Therefore, this does not change our conclusion that the adopted statistical measure of relapse may be over-estimating occurrence.

Additional limitations were concerned with the population investigated. For instance, there may have been potential selection biases associated with the sample. This study only included participants who were not lost to follow-up during the WYLOW study (Ali et al., 2017), and who also completed and returned the feedback questionnaire. Therefore, the perspectives of this study’s participants may not represent those who relapsed but were lost to follow-up, or those who relapsed but did not return feedback questionnaires. Indeed, although analyses of potential selection biases indicated no significant differences between patients classified as having relapsed in WYLOW study who returned a feedback questionnaire and those patients who did not, there may have been important variables that determined the constitution of these groups that were not investigated. Furthermore, another limitation is that perspectives of participants who remained in remission over the 12-month follow-up were not explored. This would have allowed a valuable comparison in perspectives between the two groups. For instance, it would have been interesting to explore if any participants classified as remaining in-remission agreed or disagreed that they had remained in remission, and if they also would have reported similar risk factors for symptom increases (e.g., work and relationships) or coping strategies.

Finally, participants only received LiCBT interventions, so these findings may not apply to patients who receive high-intensity CBT or other forms of psychotherapy. Furthermore, the treatments that participants received were not assessed for treatment fidelity or competency, and therefore it is impossible to know if PWPs had adhered to LiCBT treatment protocols. Two measures have recently been developed and validated for the purposes of assessing LiCBT competencies, and these measures could be applied in further research (Kellett et al., 2021).

**Further Research**

Future studies investigating relapse after psychological interventions should incorporate strategies to assess the temporal stability of symptom increases, in order to differentiate between lapses and relapses. An example of how this can be achieved is the approach used in White et al.’s (2013) investigation into relapse of panic disorder following CBT. In this study, patients were assessed on a monthly basis during the follow-up period, with a self-report validated screening tool. When patients met criteria for a potential relapse during one of their monthly assessments, they were monitored more intensively with weekly assessments. This continued until patients met criteria for relapse for two consecutive weeks (and consequently classified as having relapsed), or until there was a two-week period in which they did not meet relapse criteria in either week, in which case they returned to monthly assessments. This method may provide a more robust measure of relapse as it includes a temporal component in its measurement of symptom increases. In study designs or clinical contexts where such frequent follow-up may be difficult or unacceptable, another alternative is to assess relapse using the Longitudinal Interval Follow-up Evaluation (LIFE; Keller et al., 1987). The LIFE is a semi-structured interview that assesses the longitudinal course of psychopathology using a retrospective rating system, and it can be administered in less frequent intervals, such as three or six months. The LIFE has been demonstrated to be a reliable and valid method of understanding the week-by-week course of psychopathology over a period of one year (Warshaw et al., 1994; Warshaw et al., 2001).

Further research should also explore the effects of relapse on healthcare utilisation. This study demonstrated that over half of participants classified as having relapsed considered seeking further healthcare. However, it is unknown if all of these participants did in fact eventually seek such support. Therefore, a priority of future research should be to evaluate the extent of the ‘revolving door’ problem, by investigating how many patients who complete LiCBT return for additional support, and the logistical and economic impacts that this has on healthcare services. Additionally, investigating the extent to which routinely-delivered LiCBT incorporates competent relapse prevention is an important area for future research.

Meanwhile, looking beyond relapse, future research should also consider whether there are similar issues with how patients’ mental well-being is evaluated in assessment appointments. This study indicates that measuring symptoms at only one point in time may provide an inaccurate or unclear account of what is actually being experienced. Therefore, considering that IAPT assesses whether a referred patient has severe enough symptoms to be regarded as a clinical case (i.e., meeting ‘caseness’) by only measuring symptoms at one point in time, it may be that caseness is being over-estimated.

Finally, the aims of the present study need to be replicated with a study design that uses semi-structured interview techniques. This would address the limitations of using a questionnaire approach discussed previously, by enabling an interviewer to encourage more depth and focus to participants’ responses, and preventing misunderstandings of questions. Replications of this study could also consider assessing participants who have been classified as remaining in-remission as well as those who have been classified as having relapsed, investigate protective factors of symptom increases as well as risk factors, and consider differences in perspectives between patients who receive LiCBT and patients who receive high-intensity CBT.

**Clinical Implications**

Similar to previous chapters, this study highlights the need for relapse prevention to be a valued and core element of LiCBT interventions. In particular, patients may benefit from psychoeducation during relapse prevention planning of the key differences between lapses and relapses, which will also normalise that lapses are a part of life. Patients in this study who reported viewing their symptom increases as temporary were less likely to think they had relapsed, and also less likely to discuss returning to care. Furthermore, this study also highlights the importance that the delivery of relapse prevention is ideographically tailored to each patient, considering that participants discussed a wide range and variety of personal risk factors and coping strategies. Finally, the importance of providing booster sessions and relapse prevention interventions after treatment has also been underlined, considering that over one half of participants discussed potentially seeking further healthcare support for their increased symptoms. More robust measurements of relapse and an improved understanding of associated risk factors would allow for patients at highest risk to be identified, and consequently targeted with relapse prevention interventions. As discussed previously, this would allow the interventions to be delivered to those patients with the greatest need for them, but also that the prescription of these interventions would be performed in a cost-effective manner (i.e., not offered to every patient who complete treatment with remission of symptoms).

**Conclusion**

Although all participants in this study were classed as having relapsed, around half were either unsure or disagreed that they had experienced a relapse. Therefore, the outcome measurement criteria applied overestimated the occurrence of relapse events, by treating temporary lapses as more stable patterns of symptomatic deterioration. Consequently, we recommend the inclusion of a temporal component in the assessment of symptom increases in future studies to enhance the robustness of relapse measures. Prominent self-reported risk factors for symptomatic deterioration after LiCBT include employment-related stressors, relationship-related stressors and other stressful life events. This demonstrates the importance of idiographic approaches when investigating relapse following psychotherapy, and the personalisation of effective relapse prevention. Although this study is limited by its questionnaire approach, it represents a first-step in the qualitative exploration of symptom deterioration following LiCBT.

**CHAPTER 6**

**An Investigation of Treatment Return in an Improving Access to Psychological Therapies (IAPT) service**

The previous chapters in this thesis have demonstrated that relapse is relatively common following psychological interventions. In addition, Chapter 5 highlighted that patients who are classed as having relapsed following treatment frequently report seeking further healthcare support in order to address their symptom deterioration. As discussed in Chapter 5 and in earlier chapters, experiences of relapse may therefore contribute to a ‘revolving door’ cycle in mental healthcare services, whereby patients continually return to services for additional treatment (Roscoe, 2019). Not only does treatment return indicate that patients are experiencing persistent distress, but this process also increases waiting times for treatment and therefore hinders access for individuals who have yet to engage in services (Kazdin, 2018). Therefore, a ‘revolving door’ cycle within mental healthcare services would likely exacerbate currently existing barriers that prevent access to mental health treatment, such as the supply of mental health practitioners often not meeting population demand (Andrilla et al., 2018; Health Resources and Services Administration, 2016), and waiting times for treatment often being excessively long (Goldner et al., 2011; Triggle, 2019). Consequently, it is important for the extent of treatment return to be understood, as well as the reasons behind its occurrence, as this may allow its prevalence to be reduced and thus enhance both the long-term outcomes for patients and service efficiency.

However, research investigating treatment return within mental healthcare has been sparse, limiting our ability to understand its occurrence. Nevertheless, a few studies have recently been conducted in this area across different clinical contexts. For example, in the context of child and adolescent psychotherapy, a recent study demonstrated that 28% of youth patients who received a course of psychotherapy subsequently received further mental healthcare support within two years of the initial treatment episode (Reeder et al., 2020). Similarly, Kilcullen et al. (2021) demonstrated that 30% of students who received psychotherapy at a university counselling centre return for additional treatment within three or four years. Meanwhile, in terms of adult mental health in routine practice, Boerema et al. (2016) found that 14% of patients who received an initial outpatient treatment for depression (pharmacotherapy and/or psychotherapy) in the Netherlands returned for treatment within three or four years. Collectively, these studies evidence the existence of a ‘revolving door’ phenomenon within mental health services, across different clinical populations and age groups.

Nevertheless, further research is necessary to fully understand the extent of this phenomenon across different service contexts, disorders, and interventions. For instance, although Boerema et al. (2016) investigated the ‘revolving door’ within routine care for depression in the Netherlands, no research has thus far extended upon this research to explore this phenomenon in other countries, or for patients who receive treatment for a wider range of common mental health disorders (i.e., depression and anxiety-related problems). In addition, although Kilcullen et al. (2021) used machine learning techniques to explore predictors of treatment return in university settings, such an approach has not been applied within the context of routine practice in the general population.

This chapter will discuss a final empirical study that aimed to address the above gaps in the literature, by investigating the extent of the ‘revolving door’ within an English stepped-care mental health service that provides psychotherapy for the treatment of common mental health disorders, and by implementing a machine learning approach to develop a predictive model capable of identifying treatment returners. There were four primary objectives of this study: (1) to define treatment-uptake rates, drop-out rates, recovery rates, and treatment-return rates for both initial episodes of treatment and subsequent episodes; (2) to examine differences between patients who return to treatment and patients who do not in terms of treatment pathway and clinical characteristics; (3) to develop and evaluate a predictive model capable of identifying patients at increased risk of treatment return; and (4) to explore the developed model to identify specific risk factors that may be potentially associated with treatment return.

**Method**

**Pre-Registration and Ethical Approval**

The design and analyses for this chapter’s study were pre-registered on AsPredicted.org (https://aspredicted.org/blind.php?x=nd4xa3). Ethical approval for the analysis of a fully anonymised archival clinical dataset was obtained from the North East-Newcastle & North Tyneside NHS research ethics committee, and from the Health Research Authority (REC Reference: 15/NE/0062)

**Design and Setting**

This study was a retrospective, observational cohort study, based on the analysis of data collected during routine practice at an Improving Access to Psychological Therapies (IAPT; Clark, 2011) service in the north of England between January 2010 and June 2015. Data related to treatment referrals and treatment episodes provided by the IAPT service between 4th January 2010 and 30th June 2015 were available. To ensure that patients were observed for at least one year following initial referral or treatment discharge for each case, only data for those patients who were first referred to treatment, or discharged from treatment, between 4th January 2010 and 1st July 2014 were included. All available data for these included patients from between 4th January 2010 and 30th June 2015 were analysed.

**Participants**

Between 4th January 2010 and 1st July 2014, *N*=54339 patients were referred to the IAPT service. Some patients received only one referral for treatment, while others received multiple referrals, and some patients received one or more episodes of treatment, while others were signposted to other services. This total sample was analysed to explore the rate at which referred patients receive an episode of treatment. Within the same timeframe, *n*=21086 patients received, and were discharged from, at least one episode of treatment.[[17]](#footnote-17) A small proportion of patients were excluded from further analyses (*n*=57; 0.003%), due to data inputting errors that prevented the number of episodes of care that these patients received from being ascertained. Consequently, a final subsample of *n*=21029 patients was analysed to address the remaining study objectives, including the estimation of the rate of treatment return.

The most common primary presenting problem recorded in clinical records for the initial treatment episode received by the *n*=21029 patients was mixed anxiety and depression (28.7%). Meanwhile, 26.3% of patients were recorded as primarily being treated for an affective disorder (e.g., depressive episode, recurrent depression), 16.4% for an anxiety disorder (e.g., generalized anxiety disorder, panic disorder), 2.1% for obsessive compulsive disorder, 1.9% for post-traumatic stress disorder, and 3.6% for a range of other problems (e.g., bereavement, eating disorders). The remaining 21.0% of patients had missing information related to their primary presenting problems. The majority of patients (65.1%) received interventions based on cognitive behavioural therapy (CBT), with only a small minority (2.7%) receiving non-CBT based interventions (32.2% missing relevant data). At initial assessment, the sample had mean PHQ-9 and GAD-7 scores of 15.04 (*SD* = 6.41) and 13.33 (*SD* = 5.21) respectively. Meanwhile, at the time of discharge from their initial treatment episode, participants had mean scores of 9.70 (*SD* = 7.31) and 8.50 (*SD* = 6.18) on the two respective measures. In terms of the length of time in which patients were observed following their discharge from their initial treatment episode (i.e., the time between their date of treatment discharge and the end of the study observation period), patients were observed for an average of 1079.9 days (3.0 years; *SD* = 439.7 days), with a range of 365-1986 days (1 - 5.4 years). In terms of demographics[[18]](#footnote-18), the subsample was characterized by a majority of female patients (64.4%) from a white British background (89.9%), with a mean age of 38.5 (*SD* = 13.9). Approximately 35.1% of patients were unemployed at the start of treatment, 36.9% were unemployed at the end, and 19.8% had a self-reported long-term medical condition (e.g., asthma, chronic pain, etc.).

**Measures**

Similar to Chapters 3 and 4, this study investigated patient-reported scores on three clinical outcome measures: the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), the Generalized Anxiety Disorder scale (GAD-7; Spitzer et al., 2006), and the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002). However, in contrast to Chapter 3, which applied a reliable change index of ≥5 for both the PHQ-9 and the GAD-7 to monitor deterioration over time, and also in contrast to Chapter 4, which applied indices of ≥6 for the PHQ-9 and ≥5 for the GAD-7, this study applied indices of ≥6 for the PHQ-9 and ≥4 for the GAD-7. The indices adopted in Chapter 3 were to ensure consistency with the original WYLOW studies (Ali et al., 2017; Delgadillo et al., 2018b), while, the indices applied in Chapter 4 were recommended more recently by Richards and Borglin (2011). In contrast, the indices adopted in this chapter were applied to ensure consistency with outcome measures (other than relapse) that are endorsed and utilised within IAPT services and also investigated in this study (see below).

This study had access to each patient’s scores on the above measures from their assessment appointment and their final treatment session. Similar to Chapter 4, if patients had missing scores on any of the three measures from their final treatment session, this missing data were imputed by applying the last-observation-carried-forward (LOCF) method, in order to analyse data in a way that is consistent with other IAPT studies (Clark et al., 2009). A series of other clinical, demographic, and treatment variables were also available, and these were explored as potential predictors of treatment return. The analysed dataset was limited by not having sufficient information to distinguish whether patients received low-intensity or high-intensity treatments.

***Outcome Definitions***

**Treatment Return.** The primary outcome in this study was treatment return, defined as occurring when a patient started a new treatment episode after being discharged from a prior episode of treatment (either because they dropped out, or completed their treatment). Patients were also classified as ‘non-returners’ if they received only one episode of treatment, as ‘single returners’ if they received two episodes of treatment, and as ‘frequent returners’ if they received three or more episodes of treatment.

**Secondary Outcomes.** Apatient was considered to have shown ‘*reliable improvement*’ over treatment if their PHQ-9 or GAD-7 score reliably decreased (i.e., a decrease of ≥6 points for the PHQ-9, or ≥4 points for the GAD-7), and their score for the other scale did not reliably deteriorate. ‘*Reliable deterioration’* was determined to have occurred when a patient’s PHQ-9 or GAD-7 score reliably increased over treatment, and their score on the other scale did not reliably improve.In contrast, *‘stasis’* was defined as having not responded to treatment, with symptoms remaining relatively unchanged from the initial (pre-treatment) to the last-observed outcome measure, with no evidence of either reliable improvement or reliable deterioration. Meanwhile, an individual was considered to have met criteria for ‘*recovery’* if they scored above the clinical cut-off on either PHQ-9 and/or GAD-7 at an initial (pre-treatment) assessment (i.e., met *‘caseness’*), and both their PHQ-9 and GAD-7 scores dropped below the clinical cut-offs post-treatment. A patient was classified as having met criteria for ‘*reliable recovery’* following treatment when they had met criteria for both recovery and reliable improvement. As mentioned previously, these various outcome definitions are consistent and endorsed by the IAPT Manual (National Collaborating Centre for Mental Health, 2020).

Similar to Chapters 3 and 4, a patient was classified as having relapsed when: (1) their PHQ-9 and GAD-7 scores at the end of treatment were below the clinical cut-offs; (2) at least one of their PHQ-9 or GAD-7 scores was above the cut-off at the beginning of their second treatment episode; and (3) the outcome score above the cut-off displayed reliable deterioration.

**Analyses**

Descriptive statistics were calculated to investigate the various treatment pathways and outcomes that patients experienced through the IAPT service. This included calculating rates of treatment-uptake, drop-out, recovery, and treatment-return, alongside average assessment-wait time and average treatment duration. These are summarized in a STROBE diagram. Different pathway and outcome statistics were also compared between non-returners, single returners, and frequent returners. Included within these comparisons was an investigation of whether the length of time that patients were observed for within the study period following their initial treatment discharge was associated with treatment return status. All comparisons were made through one-way between-groups ANOVAs and Pearson chi-square tests.

***Machine Learning Approach***

Extreme gradient boosting (XGBoost; Chen & Guestrin, 2016), as implemented in the R package *xgboost* (Chen et al., 2021; version 1.4.1.1), was used to develop the predictive model designed to predict which patients were most likely to return to treatment at least once. The specific XGBoost approach that was applied in this study (e.g., the parameters and cross-validation procedures) was determined *a priori* and pre-registered on AsPredicted.org (https://aspredicted.org/blind.php?x=nd4xa3). The initial dataset consisted of *n*=21029 patients who received an initial episode of treatment. However, *n*=4 were excluded from this sample for the machine learning implementation as they were missing data for over 50% of the variables input into the model as potential predictors. This resulted in an overall sample of *n*=21025 that was utilized for predictive model development and evaluation. The variables input into the model as potential predictors of treatment-return were 31 clinical, demographic, and treatment variables that were routinely collected in the IAPT service (see Appendix I for list of input variables and glossary of terminologies).

The specific approach implemented for model development was highly similar to the approach adopted in Chapter 3. For example, the parameters *eta*, *maximum depth*, *alpha* and *lambda* were set to the same values (0.1, 3, 0, and 1 respectively). In addition, other similarities include the use of the area under the receiver operating characteristic curve (AUC) as the evaluation metric to assess the predictive ability of the model, the use of stratified five-fold internal cross-validation alongside an “early stopping rule” of 10 rounds, and the incorporation of ‘leave-one-variable-out loops’ for the purposes of variable selection.

However, there were also two additions to the approach carried out in Chapter 3. Firstly, the predictive ability of the model developed in this study was evaluated using a ‘holdout’ subsample. Specifically, at the beginning of the model development process, the dataset was randomly divided in two, with 70% of the overall sample being used as a ‘training’ subsample for developing the prognostic model (*n*=14718), and the remaining 30% being used as a ‘holdout’ sample for evaluating the predictive performance of the model (*n*=6307). Secondly, as the training subsample was identified as being class-imbalanced, with a return to treatment being less common than no return to treatment, the resampling technique of random down-sampling was applied to the dataset (Estabrooks et al., 2004). This resampling technique attenuates class-imbalance by randomly excluding majority class observations. Consequently, the training subsample was reduced to a total of *n*=4030 patients, with *n*=2015 patients not returning to treatment, and *n*=2015 returning to treatment. The training subsample was resampled as class imbalance can undermine the accuracy of machine learning algorithms (Lopez et al., 2013), as discussed previously in Chapter 3.

To evaluate the predictive performance of the developed model, when evaluated in the holdout sample (*n*=6307), the same conventional performance metrics assessed in Chapters 3 and 4 were calculated: accuracy; AUC; sensitivity; specificity; positive predictive value (PPV); and negative predictive value (NPV). Finally, the individual variables determined to be important by the XGBoost model were explored in terms of their associations with treatment-return using relative importance metrics, partial dependence plots, and descriptive statistics.

**Missing Data.**XGBoost can use missing data as an informative splitting criterion, and it is thus unnecessary to impute or exclude missing data when implementing this approach (Chen & Guestrin, 2016). Therefore, the primary XGBoost model in this study was developed using data that did not have missing data imputed. However, as a secondary sensitivity analysis, an additional predictive model was developed that followed the same approach as the primary model, but instead used data in which missing data were imputed.For this model,nonparametricmissing value imputation was performed on the ‘training’ and ‘hold-out’ subsamples separately using the R package *missForest* (Stekhoven, 2013; version 1.4). MissForest has been demonstrated to be effective at handling missing values in variables that have up to 30% missing information (Stekhoven & Bühlmann, 2012). Considering this, one variable available in the initial dataset, treatment modality (CBT vs non-CBT), which had over 30% missing values was therefore not included in the development of the imputed predictive model. To maintain consistency between the two developed models, the variable of treatment modality was also not included in the development of the primary XGBoost model. This ensured that the same 31 variables were input as potential predictors in both the non-imputed and imputed models.

**Changes from Pre-Registration**

There were two main changes for this study from the pre-registered protocol. First, this study investigated patients who were first discharged from treatment between 4th January 2010 and 1st July 2014, as opposed to between January 2010 and December 2014 as stated in the protocol. This was due to this study only having access to data from the IAPT service between 4th January 2010 and 30th June 2015. Therefore, to ensure that every participant was observed for at least one year following their initial discharge from treatment, a cut-off date of 1st July 2014 for first treatment discharge was adopted.

Second, the pre-registered protocol stated that this study would have a secondary objective of estimating the health economic costs of delivering additional treatment to treatment returners. However, this objective was not addressed, as the dataset did not have sufficient information to distinguish whether patients received low-intensity or high-intensity interventions, and consequently estimations of cost of treatment per session would not have been accurate.

**Results**

**Treatment Pathways, Outcomes and Treatment Return[[19]](#footnote-19)**

Figure 6.1 illustrates the patient pathways via the STROBE summary. Within the study observation period, *n*=54339 individuals were referred for psychological treatment, but only 45.7% (*n*=24838) received treatment. Specifically, 36.7% (*n*=19926) received treatment following their first referral, while 9.0% (*n*=4912) received treatment following later referrals, after not taking up treatment following their initial referral. Of the *n*=34413 (54.3%) who did not receive any treatment from the service, *n*=13608 (25.0%) received an assessment, but did not take up the offer of subsequent treatment.

Meanwhile, *N*=21029 patients were discharged from the service following an initial course of treatment within the observation period[[20]](#footnote-20). These patients waited an average of 26.2 days (*SD* = 22.8; range = 0 - 377) between referral and their assessment appointment. Patients received an average of 7.3 sessions (*SD* = 5.9; range = 2–52), and were treated for an average of 161.4 days (*SD* = 115.6; range = 1–1343). Approximately one-third of patients (34.5%; *n*=6985) were recorded as having dropped-out.

**Figure 6.1**

*Flow Chart Displaying Different Pathways Taken by Patients Through the Psychological Service*

N = 54339 referred to service

N = 24838 (45.7%) receive treatment

N = 3809 patients not eligible for further analyses, for following reasons:

* Received initial treatment after 1st July 2014
* Still receiving treatment by end of study observation period
* Started second treatment episode within 7 days of ending first

N = 21029 patients eligible for treatment return investigation

Initial Treatment Episode (N = 21029)

CSP = 88.9%; RR = 42.8%; D/O = 33.2%

N = 18151 patients do not return

Second Treatment Episode (N = 2878)

CSP = 92.2%; RR = 34.9%; D/O = 35.3%

Third Treatment Episode (N = 311)

CSP = 93.9%; RR = 31.5%; D/O = 21.2%

Fourth Treatment Episode (N = 47)

CSP = 93.6%; RR = 22.7%; D/O =19.1 %

N = 2567 patients do not return again

N = 264 patients do not return again

N = 44 patients do not return again

Fifth Treatment Episode (N = 3)

Abbreviations: CSP = met criteria for clinically significant problem; RR = recovery rate; D/O = drop-out rate

In terms of clinical measures and outcomes, 89.4% of patients (*n*=18686) met criteria for caseness (i.e., scored above clinical cut-off on either PHQ-9 and/or GAD-7) at the time of initial (pre-treatment) assessment. Meanwhile, 60.1% of patients (*n*=12549) reliably improved over the course of treatment, 33.3% (*n*=6951) remained in stasis (i.e., no reliable change), and 6.7% (*n*=1397) reliably deteriorated. Moreover, 42.8% (*n*=7996) of the *n*=18686 patients who scored above clinical cut-offs at the beginning of treatment were determined to have recovered following treatment, with 40.9% (*n*=7637) having reliably recovered. The average changes in PHQ-9, GAD-7, and WSAS scores over the course of treatment for all patients were -5.34 (95% CI [-5.43, -5.25]), -4.83 (95% CI [-4.91, -4.75]), and -5.36 (95% CI [-5.49, -5.23]) respectively.

***Treatment Return***

A total of *n*=2878 patients (13.7%) returned for additional treatment. Specifically, *n*=2567 (12.2%) returned for only one additional episode of treatment (i.e., the single return group), whilst *n*=311 (1.5%) returned multiple times (i.e., the frequent return group). Figure 6.2 illustrates the proportions of non-returners, single returners, and frequent returners, with associated recovery rates. Whilst *n*=2759 patients (13.1%) received another referral for treatment after their initial treatment episode, they did not take up the offer of further treatment. Therefore, a total of 26.8% (*n*=5637) patients were re-referred after their first treatment episode concluded, and approximately half of these actually went on to receive another treatment episode. Across the study’s observation period, the investigated *n*=21029 patients received a combined total of 232,663 contacts. The 13.7% who sought additional treatment received a total of 62,105 contacts, representing 26.7% of the total contacts provided. Specifically, 14.5% (33,663) of the total provided contacts were delivered as part of additional treatment episodes.

For those patients who returned to treatment at least once, there was an average time of 483.4 days (*SD* = 384.3; range = 8 – 1822) between discharge from the first treatment episode and referral for the second episode. Of these patients, *n*=1659 (58.0%) were above the clinical threshold for either PHQ-9 and/or GAD-7 at the end of their initial treatment episode. Although *n*=1203 treatment returners were below both thresholds, *n*=1064 of these patients had experienced a relapse by the start of their second treatment episode. Therefore, approximately 37.3% of treatment returners had experienced a demonstrable relapse when returning for further treatment[[21]](#footnote-21).

At the beginning of their second treatment episode, *n*=2650 (93.1%) of the *n*=2878 treatment returners met caseness criteria. Of these patients, 35.0% (*n*=927) were demonstrated to have recovered following this episode, with 33.5% (*n*=887) reliably recovering. A total of *n*=2465 treatment returners met criteria for caseness at the start of both their first and second treatment episodes. Nearly half of this subsample did not recover following either episode of treatment (*n*=1177; 47.7%). Meanwhile, 18.4% (*n*=453) recovered following both episodes, 18.5% (*n*=457) recovered following the first episode but not the second, and 15.3% (*n*=378) recovered following the second episode after not recovering after the first. The average changes in PHQ-9, GAD-7, and WSAS scores over the second treatment episode for all patients were -4.96 (95% CI [-5.20, -4.71]), -4.20 (95% CI [-4.42, -4.00]), and -5.11 (95% CI [-5.45, -4.77]) respectively.

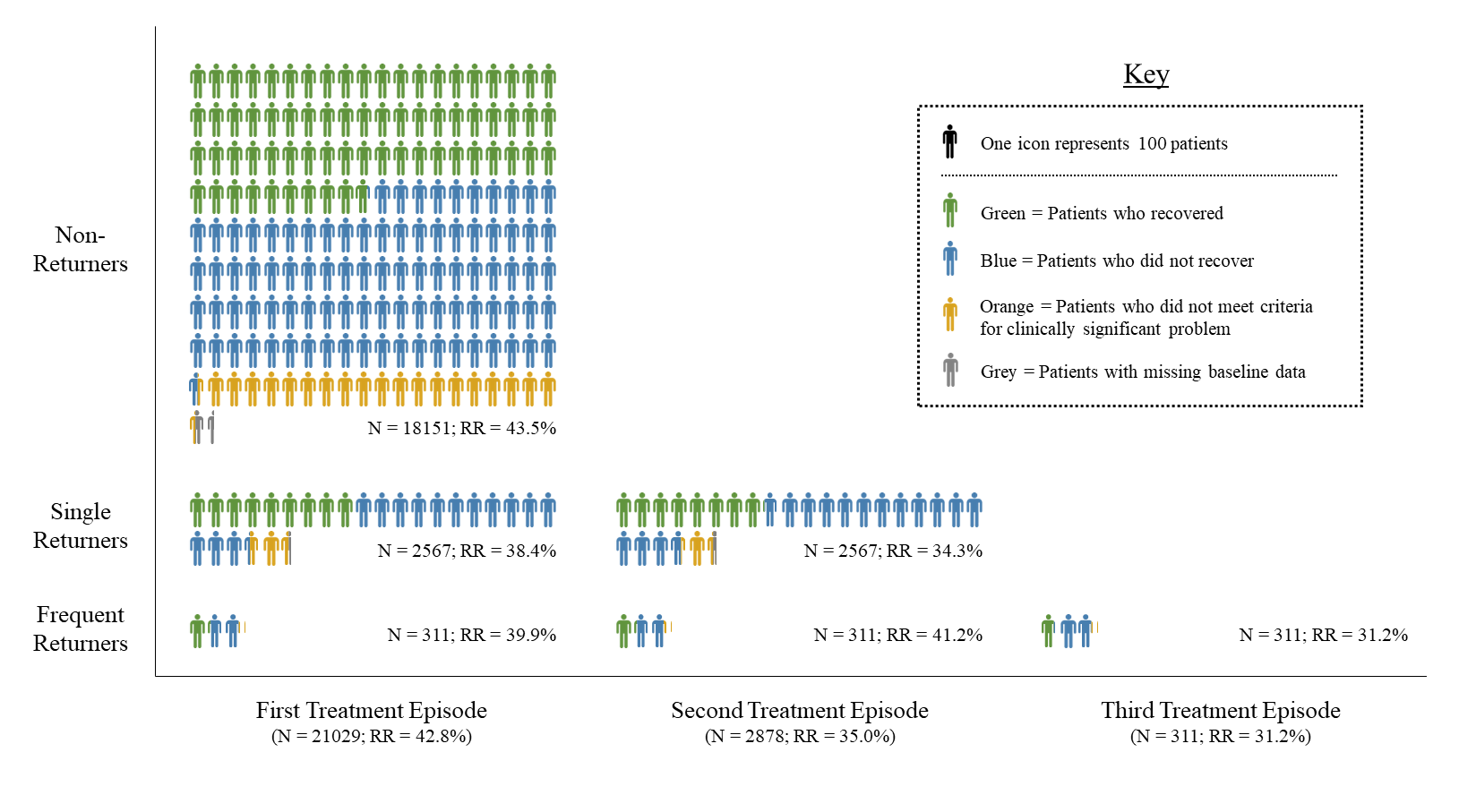
Following this, a total of *n*=311 patients returned for a third treatment episode, with 94.8% (*n*=292) meeting caseness criteria at the beginning of this episode. Of these *n*=292 patients, 31.5% (*n*=92) recovered by the end of treatment, with 31.2% (*n*=91) reliably recovering. Over the course of this third episode of treatment, patients demonstrated average change scores of -4.74 (95% CI [-5.48, -4.00]), -3.95 (95% CI [-4.58, -3.31]), and -5.89 (95% CI [-6.94, -4.83]) for the PHQ-9, GAD-7, and WSAS respectively. After a third course of treatment, *n*=47 patients returned for a fourth episode, with 93.6% (*n*=44) meeting caseness criteria. Ten of these patients (22.7%) recovered following treatment, with all 10 reliably recovering. Patients displayed average change scores of -3.70 (95% CI [-5.59, -1.82]), -3.31 (95% CI [-4.81, -1.82]), and -1.96 (95% CI [-4.38, 0.46]) for the PHQ-9, GAD-7, and WSAS respectively. Three patients who received a fourth episode of treatment returned again for a fifth episode.

Figure 6.3 illustrates the average change scores of the three outcome measures within each episode of treatment.As can be seen from Figures 6.2 and 6.3, and in the reported statistics above, recovery rates and average reductions in symptom changes either slightly decreased or remained static across subsequent treatment episodes.

**Figure 6.2**

*Proportions of Sample Who Did Not Return for Treatment, Returned for Treatment at Least Once, or Returned Multiple Times*

Abbreviations: RR = recovery rate

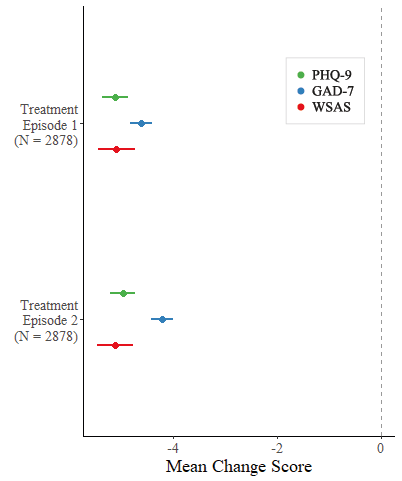
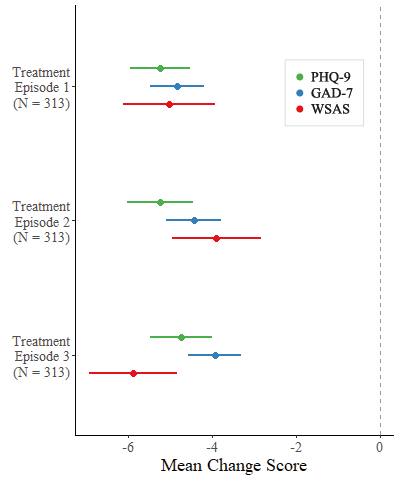
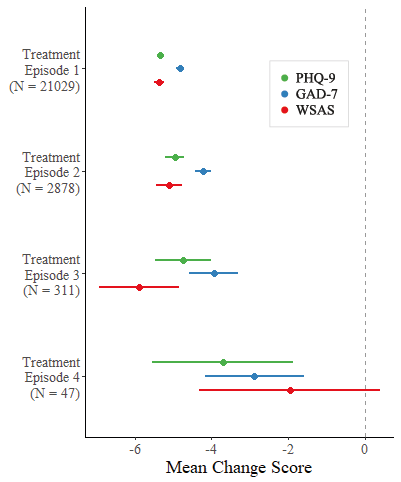


**Figure 6.3**

*Average Change Scores of PHQ-9, GAD-7, and WSAS After Each Treatment Episode Received by Patients*

*Note.* Panel A includes all patients across four treatment episodes; panel B includes all treatment returners across two treatment episodes; and panel C includes only frequent returners across three treatment episodes.

Abbreviations: GAD-7, Generalized Anxiety Disorder scale; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale.



a)

b)

c)

**Comparisons Between Treatment Return Type**

There were few differences between patients who did not return for further treatment (non-returners; NR), patients who returned once (single returners; SR), and patients who returned more than once (frequent returners; FR) in terms of treatment pathway/outcome statistics related to the first episode of treatment. For example, each sample had similar wait times for their assessment appointments (NR mean = 25.9 days, SR mean = 28.2 days, FR mean = 28.1 days). In addition, each sample received similar treatment durations, in terms of both sessions received (NR mean = 7.2 sessions, SR mean = 7.6 sessions, FR = 7.4 sessions), and time between assessment and discharge (NR mean = 161.0 days, SR mean = 164.6 days, FR = 159.7 days). Although, there was no significant difference in terms of time between assessment and discharge (*p* = .337), significant differences in terms of assessment wait times (*F*(2, 20954) = 12.67, *p* < .001, η2 = .001) and contacts received (*F*(2, 21026) = 5.40, *p* = .005, η2 < .001) were identified (specifically between non-returners and single returners). However, as can be seen above, the effect sizes for these findings suggested little if any relationship between treatment return and assessment wait time, or treatment return and contacts received.

Furthermore, approximately one-third of patients for each group were recorded as having dropped-out of their first treatment episode (NR = 34.3%, SR = 36.7%, FR = 33.1%). Although a significant difference between the groups in terms of drop-out (χ2(2) = 6.06, *p* = .048) was identified, the effect size for this finding suggested little if any association between drop-out and treatment return (Cramer’s *V* = .02). Finally, there appeared to be a small difference between the three groups in terms of recovery following treatment, with non-returners having a higher recovery rate (NR = 43.5%, SR = 38.4%, FR = 39.9%) and higher reliable recovery rate (NR = 41.5%, SR = 36.7%, FR = 36.8%) than single returners and frequent returners. However, although significant differences between the three groups in terms of recovery (χ2(2) = 22.67, *p* < .001) and reliable recovery (χ2(2) = 21.73, *p* < .001) were identified, the effect sizes for these findings suggested little if any associations between recovery and treatment return (Cramer’s *V* = .03, for both findings).

One difference that was identified between the different treatment return statuses was related to the study method; specifically, the average lengths of time that each group was observed for within the study period (i.e., time between first treatment episode’s discharge date and end date of study observation period). Frequent returners had a longer average length of observation (*M* = 1001.5 days) than both single returners (*M* = 819.4 days) and non-returners (*M* = 696.4 days). Moreover, these comparisons between the three groups were found to be significantly different, with the relationship between observation length and treatment return type representing a small effect (*F*(2, 21026) = 157, *p* < .001, η2 = .015).

**Machine Learning Model**

The XGBoost model was developed using 49 decision trees, and 15 of the 31 input variables were selected as important by the model. The model had an internal cross-validation AUC of 0.61, however, when predictive performance was externally evaluated in the hold-out sample, the AUC statistic decreased to 0.51. The model was also estimated as having an accuracy of 73.0%, sensitivity of 15.4%, specificity of 82.1%, PPV of 12.0%, and NPV of 86.0%, when evaluated in the hold-out sample.

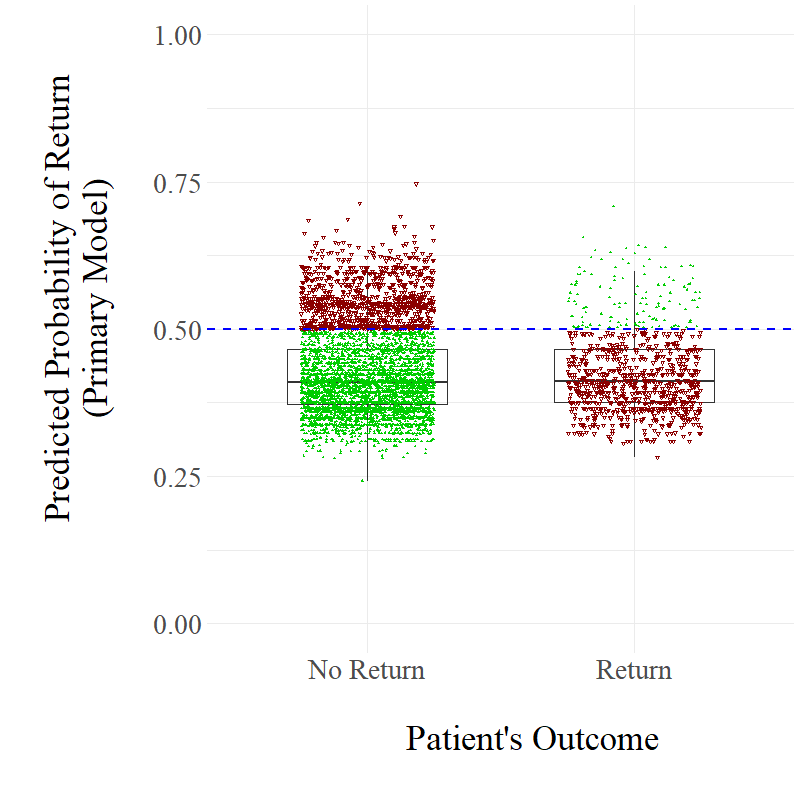
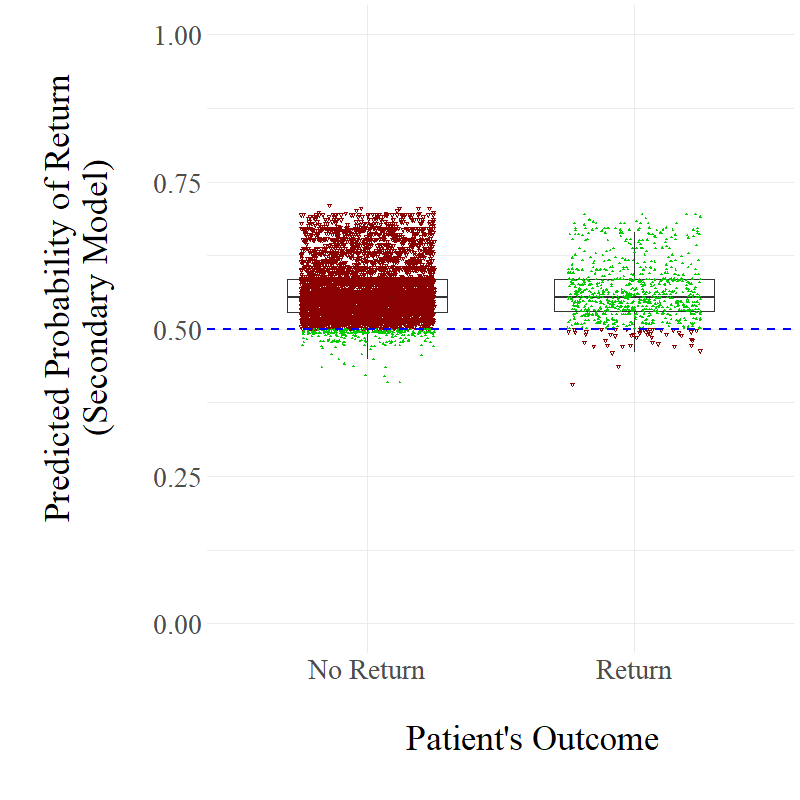
***Secondary Sensitivity Analysis***

The secondary XGBoost model (same as primary model but developed using data in which missing values were imputed)was developed using 33 decision trees, with 11 of the 31 imputed input variables being selected as predictive features by the model. This model had an internal cross-validation AUC of 0.59, with the AUC statistic decreasing to 0.50 when the model was evaluated in the imputed hold-out sample. The model was also estimated as having an accuracy of 17.5%, sensitivity of 94.4%, specificity of 5.3%, PPV of 13.7%, and NPV of 85.7%, when evaluated in the hold-out sample. The distributions of the probabilities of treatment return predicted by each model for each patient can be seen in Figure 6.4.

***Selected Predictive Features***

The features that were selected by the two developed models as being predictive of treatment return can be found in Table 6.1. Partial Dependence plots representing the predictive nature of each feature, can be found in Appendix J.

|  |  |
| --- | --- |
| **Table 6.1**  *The Predictors Identified as Important by Each Developed Model (Relative Importance %)* | |
| Primary model (non-imputed data) | Secondary model (imputed data) |
| Missing ethnicity information (20.6%) | Longer wait for assessment appointment (20.1%) |
| Having a long-term condition (16.6%) | Having a long-term condition (16.9%) |
| Higher post-treatment PHQ-9 (11.9%) | Younger age (15.5%) |
| Younger age (8.2%) | Higher pre-treatment PHQ-9 (9.1%) |
| Higher pre-treatment GAD-7 (7.5%) | More contacts attended (7.9%) |
| More contact attended (6.4%) | Higher post-treatment PHQ-9 (7.7%) |
| Large improvements or small deteriorations in PHQ-9 over treatment (5.9%) | Minority ethnic (7.1%) |
| Fewer contacts cancelled (5.9%) | Fewer contacts cancelled (6.4%) |
| Less neighbourhood deprivation (4.7%) | Less neighbourhood deprivation (4.5%) |
| Receiving low-intensity intervention at end of treatment (2.7%) | Not taking medication at start of treatment (2.5%) |
| Having primary problem of anxiety (2.7%) | Having primary problem of anxiety (2.3%) |
| Taking medication at end of treatment (2.4%) |  |
| Not having primary problem of PTSD (1.7%) |  |
| Not having primary problem of mixed anxiety and depression (1.6%) |  |
| Female gender (1.2%) |  |
| Abbreviations: GAD-7, Generalized Anxiety Disorder scale; PHQ-9, Patient Health Questionnaire; PTSD, post-traumatic stress disorder. See Appendix I for glossary for predictors, and Appendix J for partial dependence plots. | |



b)

**Figure 6.4**

*Distributions of Predicted Treatment Return Probabilities made by the Primary and Secondary Model*

*Note.* Filled, green triangles represent correct predictions for specific cases; unfilled, red upside-down triangles represent incorrect predictions.

a)

**Discussion**

This study aimed to investigate the extent to which patients who receive treatment for a common mental health problem returned for additional psychological treatment within a five-year observation period. A large sample of naturalistic, routinely collected data from an English stepped care mental health service was examined. It was found that 26.8% of patients were re-referred for additional treatment after an initial treatment episode, with approximately half of these patients then actually receiving additional treatment. Specifically, a treatment return rate of 13.7% was identified. Most of these treatment-returning patients only returned for one additional treatment episode, with frequent returners being relatively rare (1.5%). This study also attempted to develop a predictive model capable of identifying patients who are at greatest risk of returning to treatment, by implementing a machine learning approach. However, the developed model had very poor predictive performance when evaluated in a hold-out sample (AUC = 0.51), indicating that treatment return is a difficult outcome to predict with routinely available data.

The identified treatment return rate of 13.7% is highly similar to the 14% treatment return rate calculated by Boerema et al (2016), in their investigation of *N*=85754 patients who accessed outpatient, mental health clinics in the Netherlands for the treatment of depression. This is interesting considering that, alongside being set in different countries, there was also a difference between the two studies in terms of the interventions received by patients. Specifically, this study primarily investigated treatment return following CBT-based treatments, while Boerema et al investigated treatment return following a range of interventions (predominantly supportive psychotherapy and/or pharmacotherapy). This may indicate that the rate of treatment return in routine care for common mental health problems is similar across a range of interventions and service settings.

Combined, these two studies suggest that approximately one in seven patients who receive psychological treatment in routine care return for additional treatment. One interpretation of this statistic is that treatment return is relatively rare, with a large majority of patients (86%) not returning to treatment. However, 14% still represents a sizeable number of patients, and delivering additional treatment to these patients can have a substantial impact on service efficiency. This can be demonstrated by this study’s observation that the 13.7% of patients who received more than one episode of treatment accounted for 26.7% of the contacts provided by the IAPT service. This provision likely lengthens waiting times, limiting access to individuals requiring support who have not yet engaged with psychological services (Kazdin, 2018). This issue may also become more pressing as rates of referral and treatment uptake in services continue to increase (IAPT, 2019). Furthermore, it is possible that a treatment return rate of 13.7% is an underestimation. For instance, patients observed as not returning to treatment may have returned to treatment after the observation period, received treatment that occurred before the observation period, or received additional treatment at different IAPT services or in other settings (e.g., accessed pharmacotherapy through general medical practice, accessed private therapy, etc.). These potential factors could not be accounted for in this study or Boerema et al.’s study, and it is therefore probable that a treatment return rate of 13.7% is a conservative underestimation.

One potential explanation for why some patients seek additional treatment is that the initial treatment episode did not address the primary issue experienced by patients, and this explanation is partly supported by this study’s findings. Specifically, 58.0% of treatment returning patients were above the clinical threshold for depression and/or anxiety at the end of their initial treatment episode, suggesting that these patients did not fully recover following this episode. Previous research has also demonstrated that many patients do not respond to psychological interventions for common mental health problems (Clark et al., 2018; Cuijpers et al., 2014; IAPT, 2019). Therefore, it is possible that patients sought additional treatment as their initial course of treatment did not fully address the issues that they were experiencing, and they returned to the service in the hope that further treatment would produce different outcomes.

On the other hand, many patients did achieve recovery following their first episode of treatment, and yet still returned for additional treatment. Instead, many of these patients appear to have experienced a deterioration following treatment discharge, which is demonstrated by 37.3% of treatment returners experiencing a demonstrable relapse by the beginning of their second treatment episode. This is not surprising considering that depression and anxiety have been demonstrated to be highly recurrent disorders, and that relapse is common following CBT-based interventions (as demonstrated in previous chapters). Therefore, some patients may have sought additional treatment due to their initial treatments not addressing their primary issues, while others may have returned as their treatment gains were not maintained long-term. This may reflect the phenomena of enduring common mental health problems (i.e., the service profiles of those patients with treatment resistant anxiety and depression). These observations indicate that there are different pathways that underlie treatment return. This reflects previous research conducted by Siddall et al. (1988), who found that some treatment returning patients perceive their previous therapy as being helpful, while others perceive it as being not helpful, and some patients return for a new problem, while others return for the same problem. This supports the hypothesis that there are a range of reasons why participants would decide to return to psychological services for additional treatment.

The apparent range of different pathways that underlie treatment return may also potentially explain the poor predictive performance of the developed XGBoost model in this study. As the model only had clinical and demographic information from the first treatment episode available to develop its predictions, it may not have been able to distinguish between those patients who returned to treatment due to their initial problems not being adequately addressed, and those patients who did initially recover but eventually relapsed. For example, treatment returning patients who did not recover during the first episode will have had higher post-treatment clinical scores than treatment returning patients who did recover but experienced a relapse; yet both groups will have had the same outcome (i.e., returned to treatment). Differences between the two groups such as this potentially impaired the machine learning algorithm’s ability to identify consistent patterns within the data that were associated with the outcome. Nevertheless, regardless of the reasons behind the poor performance of the developed model, this study demonstrates the difficulty in predicting treatment return with the available data.

It is also important to acknowledge a more optimistic perspective of treatment return, where it is perceived as a successful outcome in which patients are socialized during initial treatment to demonstrate healthy and adaptive help-seeking behaviours. Obviously, patients who feel they have relapsed, or are suffering due to an ongoing problem, should feel encouraged to seek support to help address the issues they are experiencing. However, it is also important to consider, alongside this optimistic perspective, the detrimental effects that a ‘revolving door’ phenomenon within mental healthcare services can have on service efficiency and waiting lists, and the harmful effects this can have on individuals waiting to access the system.

Furthermore, this study’s results indicate that treatment may become less effective the more episodes a patient receives (i.e., lower recovery rates and lower average symptom change scores). Indeed, only 15.3% of treatment returners who met caseness recovered following their second episode of treatment after not recovering following their first, while 18.5% experienced the opposite effect and 47.7% did not recover following either episode. These findings potentially indicate that returning patients would receive less benefit from the additional treatment provided than those patients who would be new to treatment. One potential explanation for this is the “kindling hypothesis” (Kendler et al., 2000). As discussed in Chapter 1, this model posits that the onset of initial episodes of mental health problems is more greatly influenced by psychosocial stressors than subsequent episodes, with these later episodes being more autonomous in their onset due to a potential process of sensitization to a distressed state. If this is the case, then this may potentially explain why later episodes of psychological interventions may be less effective than initial episodes of treatment.

**Limitations**

One primary limitation is that patients were not observed within the study following their treatment discharge for similar periods of time. For example, some patients were discharged from an initial episode of treatment over five years before the end of the study observation period, while others were discharged only one year before the end of the study period. Therefore, considering that those patients with the longest periods of observation will have had greater opportunity to return for additional treatment than those with the smallest, this likely will have influenced the results. Indeed, it was demonstrated that frequent returners had significantly longer observation periods than single returners, who in turn had significantly longer observation periods than non-returners. Consequently, it is likely that this study is underestimating rates of treatment return, especially with regards to that of frequent treatment return.

There were other limitations associated with the available data. For example, it was not possible to identify whether patients received low-intensity and/or high-intensity interventions in their contacts during their treatment episodes. This prevented a comparison in treatment return rates between the two forms of intervention, and prevented an estimation of the health economic costs of delivering additional treatment to treatment returners. Another limitation of the dataset was the lack of additional background information of patients, such as whether they had a history of mental illness or had received previous treatments. Due to this, it is possible that some patients labelled as not returning for further treatment may have actually been returning to the service after a previous treatment episode that occurred before the study’s observation period. Moreover, there was also a lack of assessments of treatment fidelity or competency, thus making it impossible to know if the provided treatments were delivered appropriately. This is problematic as there may have been variability between clinicians in terms of how they ended their treatments, and discussed potential treatment return. In addition, this study only analysed data from one IAPT service, and it is therefore possible there is variability in the extent of treatment return between services. Finally, this study also initially aimed to investigate potential individual predictors of treatment return, through an exploration of a developed machine learning predictive model. However, the developed model had very poor predictive performance, and therefore it was not appropriate to explore the selected individual features in depth. The predictive modelling was therefore limited by being only able to make use of routinely collected data, and no theoretically informed variables.

**Clinical Implications**

This study’s findings indicate that treatment return can hinder service efficiency, and that treatment may become less effective with each subsequent treatment episode that is received by treatment-returning patients. Consequently, although it is important to ensure that patients who are suffering after an initial treatment episode feel comfortable with seeking further support, it is also important that during their first treatment episode, patients are sufficiently provided with the support to maintain improvements after ending treatment, and the skills to manage new issues that may arise. Two strategies that may help facilitate this are increasing recovery rates, and decreasing relapse rates. To help increase recovery rates, services should ensure that patients receive the appropriate number of treatment sessions (i.e., the appropriate ‘dose’), as per relevant clinical guidelines (e.g., National Institute for Health and Care Excellence; 2011). Meanwhile, to prevent relapse, services should ensure that booster sessions are available to patients if necessary, and that relapse prevention interventions are provided, such as mindfulness-based cognitive therapy (Segal et al., 2002) or continuation-phase CBT (Jarrett et al., 2001). As discussed in previous chapters, although delivering such interventions would incur a financial cost, the effective prevention of relapse would likely save money long-term and improve service efficiency by reducing the rate of treatment return. Therefore, services need to be able to sensitively label treatment returners (single versus multiple), and consider what interventions have been delivered previously and what interventions are currently available.

**Future Research**

Considering that there are potentially many underlying pathways associated with treatment return, and the apparent difficulty in predicting its occurrence with the available data, it may be a greater priority for future research to focus on developing prognostic models that are capable of predicting relapse and recovery. Previous research has demonstrated the potential of developing such models, with Chapters 3 and 4 demonstrating the potential utility of prognostic models of relapse. Meanwhile, for prediction of recovery, Bone et al. (2021) recently developed an algorithm capable of dynamically predicting reliable and clinically significant improvement in depression and anxiety symptoms, with the models generalizing well to external test samples. Prognostic models such as this could be used to provide feedback to therapists on patient progress during treatment, allowing therapists to adjust treatment when a patient is predicted to be ‘off-track’. Previous research has indicated that such feedback systems can help to improve treatment outcomes (de Jong et al., 2021; Delgadillo et al., 2018a). These research examples demonstrate the potential in being able to predict treatment response and relapse. Successful development of effective prognostic tools in these areas could consequently enhance recovery rates and decrease relapse rates, and in turn, potentially lower the risk of patients returning for additional treatment.

In addition, additional studies investigating prevalence of treatment return are needed that observe patients for more consistent periods of time, and aim to compare rates of treatment return between different services. This would allow for more accurate estimations of frequent treatment return, and for variability between services in terms of treatment return to be examined. Finally, future research should also prioritize adopting qualitative research approaches to explore the reasons why patients seek additional treatment(s) following an initial episode of treatment. This would provide a greater understanding of the different pathways behind treatment return, and thus potentially provide more specific targets for limiting its occurrence.

**Conclusion**

This study provides further support for the estimation that approximately one in seven patients who receive treatment for common mental health problems in routine practice return for additional treatment. This phenomenon can undermine the efficiency of mental healthcare services, by increasing waiting lists and limiting access, and potentially damages hope in change for the patient and those providing care. There may be different pathways that underlie treatment return, with some patients returning to address an unresolved initial issue, and others returning as their treatment gains have not been maintained long-term. Considering this, future research should prioritize focusing on improving recovery rates and decreasing relapse rates. This could enable patients to be less likely to require professional therapeutic support, and consequently address the ‘revolving door’ phenomenon within mental healthcare.

**CHAPTER 7**

**General Discussion**

The objective of this final chapter is to summarise the results of the five studies presented in this thesis, and to discuss the overall theoretical, clinical, and research implications. To begin, a summary of the thesis findings is provided. An interpretation of these results is then discussed in relation to: 1) the prevalence of relapse of depression and anxiety following psychological interventions; 2) the prediction of relapse occurrences; 3) potential risk factors associated with relapse; 4) the measurement of relapse; and 5) the consequences of relapse. Following this, the implications of the thesis findings for clinical practice and services are explored, followed by a review of the strengths and limitations of the thesis. Finally, recommendations for future research are provided.

**Summary of Thesis Findings**

The systematic review discussed in Chapter 2 highlighted the limited research that has investigated predictors of relapse of anxiety-related disorders following CBT. Only nine studies were identified, with all studies being highly underpowered and inconsistent measures of relapse being applied across the studies. A pooled relapse rate of 21.8% was identified, and only one consistently supported predictor of relapse was identified: residual anxiety symptoms. A meta-analysis estimated a moderate positive correlation between residual symptoms and relapse, although this was not statistically significant (*r* = 0.35, *p* = .08), potentially due to the small number of investigated studies. Other potential predictors were also identified, including the presence of a personality disorder, the degree of treatment improvement, and the occurrence of stressful life events.

The application of a machine learning ensemble of models that dynamically predict relapse after the completion of LiCBT was demonstrated to have promising predictive utility (Chapter 3). The four developed models were assessed as having AUC statistics ranging from 0.70-0.83, when evaluated through internal cross-validation on the test set. The models also displayed better predictive ability when they were developed using information that would be available at later stages of a patient’s treatment journey (i.e., during follow-up). A number of potential risk factors of relapse following LiCBT were identified, including young age, unemployment, (non-)linear patterns of change, and residual symptoms. Unemployment was identified as a particularly important predictor of relapse, with all participants who were unemployed at the beginning of treatment experiencing a relapse (*n*=31/31)

Complementing Chapter 3, which investigated relapse after LiCBT, Chapter 4 investigated relapse following high-intensity psychotherapy (including CBT and non-CBT psychotherapies). A relapse rate of 19.1% was estimated, with no significant difference in relapse rates between CBT and non-CBT-based psychotherapies being found (*p* = .815; 18.6% vs 19.8% respectively). Similar to Chapter 3, the potential utility of machine learning approaches for the purposes of predicting relapse was demonstrated (AUC = 0.72 for primary, class-imbalanced model; AUC = 0.81 for secondary, SMOTE-imputed model). However, the primary model selected two computed noise variables as important predictors of relapse, while the secondary model selected a high number of features (*k* = 18), potentially indicating that both models had overfit to the data. A number of potential risk factors of relapse following high-intensity psychotherapy were identified, including neighbourhood deprivation, unemployment, residual symptoms, and higher baseline severity in depression.

The qualitative exploration of patients’ perspectives of symptom deterioration following completion of LiCBT identified that less than half of participants agreed with their relapse classification during a longitudinal cohort study (Chapter 5). Those participants who were unsure or disagreed that they had relapsed frequently viewed their symptom deteriorations as being transient, and thus not constituting a full and sustained relapse. Participants also reported a variety of risk factors that they personally attributed to their symptom increases, with most of these being psychosocial in nature, and primarily related to work-related stress. Participants were also more likely to report implementing external coping strategies (i.e., where they seek support externally from themselves) than internal strategies (i.e., strategies that only require individual action from themselves) to manage their symptom increases, with the most common strategy being to seek further healthcare support.

The final empirical study (Chapter 6) found that treatment return is relatively common in routine practice, with 26.8% of patients being re-referred for additional treatment after an initial episode, and 13.7% actually receiving additional treatment. For those seeking further help, 58.0% remained above clinical thresholds for depression and/or anxiety at the end of their initial treatment episode, while 37.3% experienced a demonstrable relapse by the start of their second treatment episode. The impact of treatment return on service efficiency was also assessed, with the 13.7% of treatment-returning patients being identified to account for 26.7% of the contacts provided by the investigated mental health service. Furthermore, it was also identified that treatment may become less effective the more episodes a patient receives (i.e., lower recovery rates and lower average symptom change scores). Finally, a prognostic model developed to identify cases likely to return for additional treatment was demonstrated to have very poor predictive performance when evaluated in a ‘hold-out’ subsample (AUC = 0.51).

**Interpretation of Results**

Discussion of how the thesis findings can be interpreted is divided into five sections: 1) first, what the findings indicate regarding the prevalence of relapse of depression and anxiety following psychological interventions; 2) second, what the results suggest regarding our ability to predict relapse; 3) third, what risk factors have been identified as being potentially associated with relapse; 4) fourth, what the findings may mean for how relapse should be measured; and 5) fifth, what the results demonstrate in terms of the consequences of relapse on both patients and services.

**Prevalence of Relapse**

The overall findings from this thesis demonstrate that relapse is relatively common following evidence-based psychological treatment for depression and anxiety in routine care. Two studies have provided two new estimations of relapse prevalence, with both statistics being highly similar. Specifically, the systematic review outlined in Chapter 3 estimated a pooled relapse rate of 21.8% for anxiety-related disorders following CBT, while the study discussed in Chapter 4 estimated a relapse rate of 19.1% following high-intensity psychotherapy delivered in Improving Access to Psychological Therapies programme (IAPT) services. However, although these two statistics are similar, there are important differences between certain aspects of the two studies that should be highlighted. For example, Chapter 4 examined relapse of both depression and anxiety through the implementation of a short follow-up period (average of four months), while Chapter 2 synthesised studies that used longer follow-up periods (range 1-14 years) to specifically investigate anxiety relapse. Consequently, they are not directly comparable statistics, and in fact, the relapse rate in Chapter 4 is more concerning than the rate in Chapter 2 due to the substantially shorter follow-up period investigated.

The short follow-up period adopted in Chapter 4 may explain why the estimated 19.1% relapse rate is smaller than the pooled relapse rates that have previously been estimated in systematic reviews of randomised controlled trials investigating depressive relapse following CBT (29-33.4%; Vittengl et al., 2007; Wojnarowski et al., 2021). However, Chapter 2 estimated a smaller relapse rate (21.8%) than these systematic reviews, despite investigating cohort studies that implemented similar, or indeed longer, follow-up periods. This may indicate that relapse of anxiety is less common than relapse of depression following psychological interventions. Indeed, a recent systematic review of relapse of anxiety-related disorders following CBT also estimated a smaller relapse rate of 14% (Levy et al., 2021). Further research is needed to investigate differences between these two mental health problems further in terms of relapse.

In addition, another clinical aspect that may moderate the prevalence of relapse is the intensity of the treatment that patients receive. For example, the study discussed in Chapter 4 estimated a relapse rate of 19.1% following high-intensity psychotherapy for the treatment of depression and anxiety within IAPT services. This rate is smaller than the LiCBT relapse rate of 37% that was reported after a similar length of follow-up time (4 months) in the original WYLOW study (Ali et al., 2017). Consequently, this may indicate that, although still relatively common, relapse is rarer following high-intensity treatment than following low-intensity treatment. As discussed in Chapter 4, this may be due to LiCBT being a briefer intervention that is highly manualised and less bespoke (Bennett-Levy et al., 2010). However, it is also possible that this identified difference in relapse rates is due to Chapter 4 only investigating one follow-up assessment at one point in time, while the WYLOW study investigated multiple assessments on a monthly basis. Therefore, the WYLOW study had greater chance of observing relapse occurrence, with Vittengl et al. (2007) previously demonstrating that having shorter gaps in follow-up assessments is associated with observing higher relapse rates. Consequently, further research is needed to investigate potential differences between low- and high-intensity treatments in terms of relapse.

The findings from Chapter 4 also suggest that relapse rates following CBT and other high-intensity psychotherapies may be similar. This would provide further evidence to the argument that CBT is non-superior to non-CBT-based psychotherapies in the treatment of common mental health problems, with recent studies arguing this case in relation to short-term outcomes (Barkham et al., 2021; Cuijpers et al., 2013a; King et al., 2014). However, further research is needed to investigate comparisons between CBT and specific non-CBT-based psychotherapies in terms of relapse.

In summary, taking into account both previous research and the research conducted within this thesis, relapse appears to occur at a rate of 14-53% following psychological interventions delivered in routine services. Prevalence seems to vary depending on methodological factors, the intensity of treatment provided, and the mental health problem that was initially treated. Further research employing more consistent methodology is needed to enable more robust comparisons between clinical aspects to be made in terms of relapse prevalence.

**Predicting Relapse**

As discussed previously throughout this thesis, the ability to predict relapse would allow for patients at greatest risk to be identified and thus targeted with relapse prevention interventions. This would ensure that those patients receive the support that they require to provide them the greatest opportunity to maintain their treatment gains, while also ensuring that services provide these interventions in a cost-effective manner (i.e., by not providing them to every recovered patient). The findings from Chapters 3 and 4 demonstrate that it is possible to predict relapse following psychological interventions. The application of data-driven, machine learning approaches resulted in the development of prognostic models that displayed promising predictive capacity. Furthermore, the developed models with the greatest predictive performance across the two chapters were the two models that were trained using information that was acquired during follow-up. This indicates that the prediction of relapse is more accurate when information related to the patient *after* the completion of treatment is available. This therefore illustrates the importance of providing follow-up reviews, as they enable improved predictions related to the occurrence of relapse events to be made.

However, the developed models had a number of limitations, primarily related to the models being developed using relatively small samples. For example, the overall size of these samples prevented any of the models from being evaluated externally on hold-out samples, with all evaluations being performed on subsamples that were involved in the training of the models. Considering that predictions tend to be less accurate when models are applied in new samples (Chekroud et al., 2016; Delgadillo & Gonzalez Salas Duhne, 2020; Delgadillo et al., 2017; Leighton et al., 2019), it is not possible to know how well the developed models generalise to external datasets. Indeed, it is especially likely that the models developed in Chapter 4 would perform worse in new samples, due it being likely that they overfit to the original data. Furthermore, an additional limitation of the developed models is that although the models were associated with reasonably high AUC statistics (range = 0.70-0.83), other conventional evaluation metrics of predictive performance were frequently poor. Indeed, only Model 4 in Chapter 3 performed relatively well (>60%) across all metrics. Considering that Model 4 is unable to predict relapses until three months have passed post-treatment and its lack of evaluation in an external sample, none of the developed models would be appropriate for use in clinical practice due to all of the discussed limitations.

Nevertheless, the developed models in Chapter 3 and 4 still demonstrate the potential of applying machine learning approaches for the purposes of predicting relapse. Therefore, considering that many of the discussed limitations are likely to be associated with underpowered samples, future research that applies these approaches with the use of adequately powered samples holds promise in developing clinically acceptable prognostic tools.

**Relapse Risk Factors**

The systematic review discussed in Chapter 2 highlighted the limited research that has been conducted in the area of predictors of relapse following CBT, with little currently being known regarding what risk factors are associated with relapse. Identifying specific predictors would improve our ability to develop clinically acceptable predictive models for relapse, while also potentially enabling more bespoke relapse prevention interventions to be developed and provided.

The systematic review did identify the presence of residual symptoms at the end of treatment as being a consistently well-supported predictor of relapse. Indeed, this variable has previously been regarded as a well-established risk factor of relapse (Bockting et al., 2015; Wojnarowski et al., 2019), and was also identified as important by the developed prognostic models discussed in Chapters 3 and 4. Different explanations for why residual symptoms may predict relapse have been provided, with Paykel (2008) suggesting that residual symptoms represent the persistence of an original disorder in a milder form. Meanwhile, Harkness et al. (2014) posited that the presence of residual symptoms predicts relapse as it is associated with a greater vulnerability to experiencing future life-events as intolerably stressful. The findings from Chapter 5 of this thesis may potentially suggest an alternative explanation, with residual symptoms potentially reflecting the unsuccessful acquisition of an internal locus of control, resulting in patients not feeling confident in applying internal coping strategies and thus seeking external support when in distress. Further research is required to explore these potential mechanisms that explain why residual symptoms is a consistently observed predictor of relapse. Other clinical variables were identified across the thesis chapters as being potentially important risk factors associated with relapse, including the presence of a personality disorder (Chapter 2), the degree of improvement over treatment (Chapter 2), trajectories of change over treatment (Chapter 3), and baseline severity of depression (Chapter 4). Previous research into these specific factors has been limited, and therefore they may be fruitful targets for future research.

Moreover, psychosocial factors were also consistently observed as being important for the prediction of relapse. For example, such factors include the experience of stressful events (Chapters 2 and 5), neighbourhood deprivation (Chapter 4), and social relationship issues (Chapter 5). The most promising potential psychosocial risk factor associated with relapse across the thesis was work-related stress and unemployment. Indeed, unemployment was identified as being a highly relevant predictor of relapse following LiCBT, with every participant who was unemployed at the beginning of treatment experiencing a relapse (Chapter 3). As discussed in Chapter 5, these findings regarding work and unemployment indicate that extreme aspects of work appear to be associated with a risk of relapse, supporting similar conclusions regarding associations between employment and mental health (Waddell & Burton, 2006).

This emphasises the importance of considering patients’ employment status and work context when delivering treatment and considering relapse prevention. Indeed, the provision of employment support was highlighted as being one of the initial key principles underlying the operation of IAPT services (Clark, 2011). A recent report has suggested that the provision of employment support in IAPT services can increase patients’ levels of confidence, improve their mental health and wellbeing, and help them remain in or return to work (Loveless, 2019). This report has also identified benefits to services, primarily through freeing up therapist time to focus more directly on providing therapy, which in turn reduces both burn-out and waiting times. This demonstrates that the provision of employment support in mental healthcare services provides benefits to both patients and services, and based on the findings from this thesis, may also reduce the risk of relapse. Meanwhile, looking beyond mental healthcare, a number of interventions have also been developed for preventing burnout and stress in the workplace (Bagnall et al., 2016). For example, a review of 23 systematic reviews identified that such interventions conducted at the individual-level can improve mental health outcomes, with CBT producing the largest effects compared to other interventions (Bhui et al., 2012). In addition, although similar interventions conducted at the organisational-level have been comparatively under-researched, they have been suggested to potentially produce longer-lasting effects than individual-level interventions (Awa et al., 2010; Bagnal et al., 2016). This demonstrates that interventions for managing work-related stress are available, and should be implemented by organisations to improve the mental health and wellbeing of employees, and consequently reduce workplace absence and disengagement.

Overall, the findings of this thesis suggest that psychosocial factors need to be considered when assessing if patients are at risk of relapse. However, compared to clinical variables, psychosocial variables have been comparatively neglected in relapse research (Chapter 2; Wojnarowski et al., 2019). Research is therefore often ignoring an aspect of patients’ lives that may potentially play a significant role in the occurrence of the outcome of interest. Improved understanding of the associations between these factors and relapse may allow for more accurate predictions, and more personalised relapse prevention plans to be provided. Therefore, it should be a priority for future research to examine psychosocial variables, alongside clinical variables, when investigating risk factors associated with relapse following psychological interventions.

**Measurement of Relapse**

As discussed in Chapter 5, this thesis has identified a potential issue with how relapse is frequently assessed in contemporary research. Specifically, it was identified that less than half of participants, who had been classified as relapsed following an experience of clinically significant symptom deterioration, agreed with their relapse classification. This raises concerns regarding whether currently used statistical methods of assessing the occurrence of relapse are accurate. Participants who did not explicitly agree with their relapse classification frequently discussed that their symptom increases were temporary. This may indicate that these participants had experienced a lapse, rather than a relapse, with their symptom deterioration not being maintained over an extended period of time.

If this is the case, then the application of reliable change indices at one point of time as an assessment of relapse occurrence may not be appropriate, as they may not to be specifically assessing relapse occurrence, but also the occurrence of a temporary lapse. Considering that the effective management and amelioration of lapses could constitute an outcome of effective therapy, non-specific relapse measurement criteria that incorrectly classify lapses as relapses could in fact be underestimating the long-term effectiveness of psychological interventions. Furthermore, application of an imprecise measure of relapse would also limit our ability to develop clinically effective predictive models. As discussed in Chapter 5, one approach to address this potential issue may be to ensure that when the occurrence of relapse is being measured, the temporal stability of symptom increases is also assessed, so that lapses and relapses can be adequately differentiated.

However, there may be alternative options for how the long-term effectiveness of psychological treatment should be assessed. For example, in a recent randomized controlled trial of an Internet-delivered relapse prevention intervention for depression, Kordy et al. (2016) assessed number of ‘well weeks’ (i.e., weeks with at most mild symptoms assessed by the Longitudinal Interval Follow-Up Evaluation [LIFE]; Keller et al., 1987) over a period of 24 months as their main outcome variable of interest. In addition, they also measured relapse occurrence, defined as having a Psychiatric Status Rating score greater than 5 for two consecutive weeks (also assessed using the LIFE). They identified that all participants had a large number of ‘unwell weeks’, but also that there were a large number of transitions between ‘well weeks’ and ‘unwell weeks’, further demonstrating the potential issue of applying a ‘one-point-in-time’ measurement of relapse. Furthermore, they found that although the intervention was associated with relapse rates similar to that of treatment-as-usual, the intervention was associated with patients experiencing significantly more ‘well weeks’, through faster transitions from unwell to well, and slower transitions from well to unwell. This study demonstrates that only measuring relapse rates when assessing long-term outcomes of psychological interventions may underestimate effectiveness and prevent positive outcomes from being identified, while also not providing information regarding temporal stability of treatment gains. Therefore, it is important for future research investigating long-term outcomes of psychological interventions to consider what it is that is actually wanted to be known at an individual level about patient experiences following treatment completion, and this may help guide decisions regarding how these outcomes should be measured.

**Consequences of Relapse**

Findings from this thesis also highlight the impact that relapse can have on both patients and services. For example, participants in Chapter 5 who were classified as having relapsed often discussed experiencing worsening or persistent symptoms, and that their deteriorations were having enough of an impact to make them consider seeking further support from healthcare services. Indeed, this was the most common coping strategy that participants reported when discussing how they would manage their symptom increases. This highlighted that the experience of relapse often leads patients to re-enter the healthcare system, consequently having an impact on service efficiency.

This impact on service efficiency was demonstrated in Chapter 6, with relapse occurrence contributing greatly to the 13.7% treatment return rate. Specifically, over one third of treatment-returning patients were shown to have experienced a demonstrable relapse upon the beginning of their second treatment episode. The detrimental effects of treatment return upon services was emphasised in Chapter 6, with over one quarter of all contacts being provided to less than one seventh of all patients. Furthermore, results indicated that treatment may become less effective with each additional episode of treatment received. Consequently, it is possible that returning patients receive less benefit from additional treatment than if it was provided to patients waiting for treatment who are new to treatment. This is an important point for discussion considering the high waiting lists for treatment that currently exist (Triggle, 2019), and that referrals and treatment uptake continue to increase (IAPT, 2019). Considering this, it is therefore important for approaches to reduce the rate of treatment return to be investigated. Unfortunately, the prediction of which cases are at greatest risk of returning to treatment proved highly challenging when using routine data. Therefore, future research should prioritise on increasing recovery rates, and reducing relapse rates, as these may be too factors that exacerbate the ‘revolving door’ issue within mental healthcare services.

**Clinical Implications**

**Clinical Practice**

The thesis findings and interpretations provide a number of recommendations for clinical practice of evidence-based psychological interventions for depression and anxiety. Most importantly, the findings emphasise that relapse prevention must be a valued and fundamental component of the treatment process, with the development of relapse prevention blueprints being a core component of treatment. This is particularly highlighted for LiCBT, considering that time is more limited for the delivery of these interventions, and that relapse rates may potentially be higher following LiCBT than high-intensity psychotherapy. It is also important that the development of relapse prevention plans is ideographically tailored to individual patients, with psychosocial aspects being considered during development. Indeed, what is important and what might work for one patient may be different to what is important and what might work for another patient. An additional important aspect of relapse prevention is that it should involve psychoeducation regarding the key differences between lapses and relapses, with it being emphasised that lapses are a normal part of life. This will ensure that patients do not catastrophise upon the experience of temporary symptom increases, and will help them learn to accept that ‘bad weeks’ are common and normal. Finally, it is important that the opportunity of receiving booster sessions if necessary post-treatment is provided to patients, as these have been demonstrated previously to help prevent relapse (Gearing et al., 2013; Whisman, 1990. Indeed, it is currently recommended within IAPT to offer such sessions to patients if necessary (National Collaborating Centre for Mental Health, 2020). Based on the findings and conclusion from this thesis, the delivery of these aspects of relapse prevention may provide patients with skills and resources that can help them manage the occurrence of lapses or unwell weeks, and thus maintain their treatment gains long-term.

**Clinical Services and Policy**

This thesis also provides several implications for the provision of mental health services. Primarily, the detrimental effects of relapse on service efficiency has been emphasised, with patients who experience relapse often returning to services for additional treatment. This process exacerbates a ‘revolving door’ cycle already occurring in mental health services. Therefore, it is important that relapse prevention is taken seriously by services, and that relapse prevention interventions are available to patients who complete treatment with remission of symptoms. It is especially important that these interventions are provided to those patients at greatest risk of relapse. However, this requires an understanding of what risk factors are associated with relapse occurrence, and as highlighted by this thesis, current understanding related to this is somewhat limited. For the time being, services may consider targeting relapse prevention intervention to patients who display residual symptoms at the end of treatment, as this is currently the only consistently supported risk factor associated with relapse.

For understanding related to relapse to be improved, and for specific and robust risk factors of relapse to be identified, it is important that structured follow-up is incorporated into clinical practice. Alongside providing patients with an opportunity to consolidate the skills that they learned during treatment and for early signs of relapse to be detected, this would allow for significantly more data to become available for research into relapse. However, as demonstrated in Chapter 4, it is important that follow-up is commissioned as part of the intervention package and offered in a systematic manner. Otherwise, it appears that only a minority of patients will be offered or attend follow-up reviews, with those patients at greatest risk of relapse being less likely to receive such reviews and thus systematically excluded from follow-up support. Furthermore, it is important that data related to follow-up reviews is consistently recorded and managed, as inconsistent recording limits the extent to which research investigating relapse can be conducted. Access to substantially larger samples would allow for more questions related to relapse to be explored (e.g., differences between services and treatments in terms of relapse rates), for different variables to be prospectively tested as potential predictors of relapse, and for clinically acceptable prognostic models for relapse to be developed and evaluated. Without more structured follow-up being provided in mental health services, these advances in relapse research cannot be made and our understanding will remain limited.

Another implication for services is that unemployment was identified as a potentially potent risk factor for relapse, especially for LiCBT. It is therefore important that the employment status of patients is assessed upon their entry to services, and that they have access to an employment adviser if employment (lack of, or at risk of losing) is an issue. As discussed previously, the provision of employment support was one of several key principles initially laid out in the original general framework for IAPT for the operation of services (Clark, 2011). It was initially intended for there to be a 1:8 ratio between employment advisors and therapist within IAPT, however this was not implemented. However, there is currently an ongoing programme to extend the employment advice component of IAPT provision, with the aim being to bring the current ratio closer to 1:8 (Loveless, 2019). The findings of this thesis highlight the importance of this component, as employment support may allow for patients to remain in or return to work, which appears to be a protective factor against relapse when compared to unemployment.

Finally, there also needs to be a recalibration in how policy evaluates the effectiveness of psychological treatments provided in mental health services. For instance, IAPT currently defines ‘recovery’ as occurring when patients score above the clinical cut-off on either PHQ-9 and/or GAD-7 at their pre-treatment assessment, and then score below both cut-offs post-treatment (National Collaborating Centre for Mental Health, 2020). This therefore means that a patient who scores 11 on the PHQ-9 in their assessment appointment and then scores 9 in their second session would be classed as having recovered. However, such a small reduction is unlikely to be clinically relevant to patients. This is especially true if patients have residual symptoms when classed as recovered (as in the previously provided example), considering that this thesis has highlighted residual symptoms as a risk factor for relapse. Considering this issue, it may be more appropriate for IAPT to adopt ‘reliable recovery’ (i.e., deemed to occur when patients score below the clinical cut-offs and their reduction from baseline represents statistically reliable and clinically significant improvement) as its primary short-term outcome measure.

Furthermore, IAPT’s definition of ‘recovery’ does not align with other definitions, which propose that full recovery does not occur until an extended period of remission has occurred (Frank et al., 1991; Rush et al., 2006). Consequently, IAPT currently does not take into account this temporal aspect of recovery, by only focussing on a remission of symptoms that occurs at one point in time. Therefore, although undoubtedly important, IAPT needs to look beyond simple short-term 50% recovery targets, and start considering long-term outcomes to be just as important. It has been demonstrated that a significant proportion of patients who finish treatment with remission of symptoms are not able to maintain their treatment gains, and this has detrimental consequences on service efficiency long-term. Therefore, to ensure that patients are provided with *evidence-based* treatments for common mental health problems, evidence related to long-term effectiveness of these interventions needs to be established and monitored.

**General Strengths and Limitations of the Thesis**

**Strengths**

This thesis had several strengths with regards to the approaches that were used to investigate relapse following psychological interventions. For instance, a range of methods and analytical approaches were adopted to examine the various thesis aims, including systematic review and meta-analytic techniques, retrospective analysis of longitudinal outcomes, machine learning approaches, and qualitative analysis. The range of analytical techniques implemented enabled a range of different research questions related to relapse to be explored. In addition, these questions were primarily explored utilising large datasets. As discussed previously, this is rare for investigations of relapse following psychological treatments, due to the limited provision of follow-up in services and the limited research into long-term outcomes of psychotherapy. The specific application of machine learning approaches enabled these large, complex, multidimensional datasets to be analysed to develop predictive models of relapse and identify potentially important predictors of relapse. In addition, the research in this thesis benefited from the advantages associated with using practice-based, routine outcome data. Specifically, routine outcome monitoring, as implemented through IAPT, provided a wealth of longitudinal session-by-session data on a range of variables and outcomes. Furthermore, use of this data also enhances the generalizability of the findings to IAPT services, while also allowing for many of the thesis recommendations to be implementable by services within current frameworks and systems.

**Limitations**

There were also a number of limitations across the thesis studies, with many being associated with the fact that the thesis primarily involved the retrospective, secondary analysis of practice-based, routine outcome data. For instance, the nature of this data meant that the studies less rigorously designed, with patients not being randomised and the study context being highly uncontrolled. This limits the internal validity the studies conducted, with there potentially being confounding variables that may have influenced the results. For example, in Chapter 4 follow-up durations were highly variable, and only a minority of patients actually received follow-up reviews. In addition, the uncontrolled study designs prevented checks of treatment fidelity or competency from being conducted. Therefore, it is not possible to know if relapse prevention was carried out during the initial treatment episode, or if treatment was delivered to patients to an appropriate standard. In addition, another limitation associated with analysing practice-based data is that missing data and inconsistent reporting were commonplace. This particularly caused issues in Chapter 4 where it was not clear what treatment sessions represented a follow-up review appointment, and consequently, assumptions needed to be put in place, limiting the robustness of the findings. Finally, an additional limitation associated with the use of practice-based, routine outcome data is that the studies conducted were limited to the available data. Consequently, only routinely collected information, which was not collected for the purposes of addressing any specific research question, could be investigated as potential risk factors of relapse or treatment return. This meant that certain theoretically derived variables, such as potential psychosocial risk factors (e.g., the experience of stressful life events, work-related stress etc.), could not be explored.

The limited access to robust and quality data, along with the limited use of theoretically derived variables, have previously been highlighted as significant barriers to the effective application of machine learning approaches (Aafjes-van Doorn et al., 2021; Rose, 2018). Indeed, although machine learning has been touted as having great potential to enhance and transform prediction in various medical disciplines (Delgadillo, 2021; Rose, 2018), multiple restrictions to its effective application have been identified, and these limitations have also been demonstrated within this thesis. For instance, one major limiting factor for the effective use of machine learning approaches is having a sufficient sample size, with very large samples being required for both training and validating developed models. For example, recent studies have demonstrated that, although required sample sizes vary depending on multiple factors (e.g., the event rate of the outcome, the number of investigated predictors etc.), it is generally expected that hundreds of cases are required when developing a multivariable prediction model, and moreover, additional hundreds of cases are required when externally validating a model (Riley et al., 2019; Riley et al., 2021). Therefore, as discussed previously, the studies in Chapters 3 and 4 were underpowered for the development of predictive models, and the limited samples also prevented the developed models from being externally validated. Therefore, although the use of machine learning approaches in Chapters 3 and 4 enabled an exploration into risk factors of relapse and the generation of hypotheses for future research, it is vital that future studies applying such methods to develop clinically effective prognostic tools do so with the use of adequately powered samples.

Another limitation of the use of machine learning approaches is that their underlying inherent complexity limits the ability for developed models to be interpreted. Indeed, prediction models developed by such methods have often labelled as ‘black boxes’, whereby it is not possible to understand how the models operate (Adadi & Berrada, 2018; Collins & Moons, 2019). The specific machine learning approach adopted in this thesis, XGBoost, is a highly complex algorithm involving the development of dozens of non-linear decision trees (Chen & Guestrin, 2016). Although this inherent complexity enables the analysis of large, multidimensional, non-linear datasets, it also makes it relatively difficult to understand how specific predictions are made and how specific predictors may relate to each other and the outcome. This limited interpretability prevents the face-value assessment of developed models, and hinders the ability for independent validation and evaluation from other researchers (Collins & Moons, 2019). In addition, the ‘black box’ nature of models developed by machine learning approaches has been argued to be potentially off-putting to some clinicians, patients, and researchers, who have a greater interest in theory and causal mechanisms and how prognostic models make their predictions (Aafjes-van Doorn et al., 2021; Shortliffe & Sepúlveda, 2018; Vayena et al., 2018). Furthermore, the inherent complexity of machine learning approaches also means that there are multiple decisions to be made by the researcher in the process of applying such an approach, including decisions related to how variables should be processed before implementation and what values should be assigned for various hyperparameters (Delgadillo, 2021). Consequently, considering the inherent complexity of machine learning approaches, with their comparatively limited interpretability and large number of researcher degrees of freedom, it is vital that the application of machine learning approaches in psychotherapy research is both rigorously conducted and transparently reported (Collins et al., 2015; Collins et al., 2021).

Looking beyond the restrictions of machine learning approaches, there were also limitations in this thesis associated with the applied methods of measurement. For example, inconsistent reliable change indices for the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001) and the Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006) were applied across the thesis for the assessment of relapse. Differences in reliable change indices were applied for the following reasons: 1) Chapter 3 applied indices of ≥5 for both measures to ensure consistency with the original WYLOW studies (Ali et al., 2017; Delgadillo et al., 2018b); 2) Chapter 4 applied indices of ≥6 and ≥5 for the PHQ-9 and GAD-7 respectively due to more recent recommendations made by Richards and Borglin (2011); and 3) Chapter 6 applied indices of ≥6 and ≥4 for the PHQ-9 and GAD-7 respectively to ensure consistency with outcome measures (other than relapse) that were utilized within IAPT and investigated in that study. The inconsistent application of reliable change indices resulted in different rates of relapse to be estimated, and consequently different patients to be classified as being relapsed. However, as can be seen in Table 7.1, application of the different reliable change indices did not change relapse rates within the three studies to a significant extent. Therefore, this variation may not have had a large effect on the results.

An additional limitation of this thesis is that the studies conducted were highly centred within the IAPT and English context. Therefore, although the findings may generalise well to services within the IAPT programme, it cannot generalise to other mental health settings or countries. Furthermore, due to the IAPT focus, the studies primarily focussed on CBT interventions. Although non-CBT-based psychotherapies were investigated in Chapters 4 and 6, due to limitations with the available data it was not possible to distinguish specific treatment modalities (e.g., interpersonal therapy, counselling for depression). Therefore, the findings of this thesis that are specifically related to CBT may not generalise to other modes of psychological treatment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 7.1**  *Relapse Rates for Study Chapters 3, 4, and 6 Recalculated using all Pairs of Reliable Change Indices (RCI) Applied across the Thesis (95% Confidence Intervals)* | | | |
|  | RCI –  ≥5 (PHQ-9); ≥5 (GAD-7) | RCI –  ≥6 (PHQ-9); ≥5 (GAD-7) | RCI –  ≥6 (PHQ-9); ≥4 (GAD-7) |
| Chapter 3ab | 69.2% (63.7% - 74.3%) | 65.9% (60.3% - 71.2%) | 68.2% (62.7% - 73.4%) |
| Chapter 4 | 20.2% (15.5% - 25.6%) | 19.1% (14.5% - 24.4%) | 20.6% (15.8% - 26.0%) |
| Chapter 6 | 36.9% (35.1% - 38.7%) | 36.1% (34.4% - 37.9%) | 37.3% (35.5% - 39.1%) |
| aThe calculation of relapse rates for this table for Chapter 3 excluded *n*=12 participants who were classified as having relapsed in the original study due to re-entering treatment services, and not due to changes in outcome scores.  bThis study involved participants being removed from the study once they had been classified as having relapsed in the original study. Therefore, if different RCI did not classify them as having relapse, there was no additional opportunity for these participants to experience a relapse according to the new criteria. | | | |

Finally, this thesis is also limited by its lack of investigation of potential effects related to therapists/practitioners or geographical regions in relation to relapse rates. For instance, it is possible that certain therapists or specific services produce lower relapse rates for their patients than others. Indeed, it has been demonstrated that therapist effects exist in relation to short-term outcomes of psychological interventions (Green et al., 2014; Johns et al., 2019), and that different IAPT services are associated with highly different rates of recovery (Baker, 2021). Therefore, it is plausible that similar effects exist for long-term outcomes, and this thesis is limited by not examining such potential effects.

**Recommendations for Future Research Directions**

This thesis has added to the evidence base for psychological treatments for depression and anxiety in routine care, and provided an initial exploration into the occurrence of relapse. However, further research is required to both develop understanding of relapse, and improve our ability to predict it. A priority of future research should be conducting more controlled, prospective, longitudinal cohort studies to investigate the occurrence of relapse. This would enable the following study design aspects to be implemented and thus allow for more robust conclusions to be drawn: 1) the implementation of follow-up in a consistent and structured manner; 2) the assessment of treatment fidelity and competency; 3) the investigation of variables other than routinely collected information (e.g. the occurrence of stressful life events): 4) the prospective testing of potential risk factors associated with relapse (e.g., unemployment for LiCBT, neighbourhood deprivation for high-intensity psychotherapy etc.); and 5) the measurement of relapse to be assessed in a manner that allows for temporal stability of symptom increases to be accounted for.

The current research was unable to provide any firm conclusions regarding the differences between CBT and other high-intensity psychotherapies, or between high-intensity interventions and LiCBT, in terms of relapse rates or associated risk factors. Therefore, future research should aim to explore any potential differences between these interventions in relapse prevalence. Furthermore, this research could also consider further exploring any potential moderating effects that pharmacotherapy may have on relapse following these different intervention formats. Moreover, further research should explore differences between therapists and services in terms of relapse rates.

Research should also continue to develop predictive models that are capable of identifying cases at risk of relapse following treatment. However, significantly larger samples are required for the development of clinically effective, acceptable models to be possible. Although this currently appears a challenging task, there are two possible avenues for such samples to be developed: 1) standardised incorporation of systematic follow-up in routine practice; or 2) the synthesis and harmonisation of individual participant data across separate, similarly conducted research studies. This latter approach is currently being adopted for the development and validation of a prognostic model capable of predicting relapse of depression following non-pharmacological interventions in primary care (Moriarty et al., 2021). If this developed model displays effective accuracy in both internal and external validation, this would further demonstrate that relapse can be predicted following psychological interventions, Furthermore, it would provide a useful blueprint for how a similar prognostic model could be developed for the purposes of predicting anxiety-related disorders in primary care.

In addition to the continued development of predictive models, future research should also focus on developing low-intensity relapse prevention interventions that can easily be embedded into routine services. Indeed, the development and evaluation of low cost, brief relapse prevention interventions have thus far been limited (Rodgers et al., 2012). However, two low-intensity interventions have recently been developed that display promise. First, the SMArT (Self-Management after Therapy) intervention has been developed, which involves the use of implementation intentions to develop specific relapse prevention plans, and consists of only one face-to-face session and three subsequent telephone follow-up reviews (Lucock et al., 2018). Second, a personalised post-treatment smart-messaging intervention has also been proposed, in which patients are regularly sent personalised relapse prevention text messages that they had written themselves at the end of treatment, and which are tailored to the patients’ current mood state (Malins et al., 2020). Both of these interventions have been demonstrated to have potential effectiveness (Lucock et al., 2021; Malins et al., 2020), however large-scale randomised controlled trials are required to fully assess their efficacy and cost-effectiveness. Nevertheless, they demonstrate the promise in development of low cost, brief interventions for the purpose of relapse prevention.

Related to this, future research is needed to investigate the provision and delivery of follow-up reviews and booster sessions following psychological interventions in routine services. This thesis has further highlighted that follow-up is rare in routine services, and that when it is provided it is delivered in a highly unstructured manner. It is also not clear what is actually involved in the delivery of follow-up and booster sessions following treatment. Therefore, research is required to understand how follow-up reviews and boosters are delivered across therapists and services, and to understand what barriers are associated with them being provided and attended. This may allow for changes to be introduced that facilitate the effective provision of follow-up after treatment.

Furthermore, future research should also prioritise further qualitative explorations of relapse. This thesis has demonstrated how qualitative research can enrich quantitative relapse research, by enabling different questions to be asked, patient’s perspectives to be taken into consideration, and new hypotheses related to relapse to be generated. However, the qualitative study outlined in Chapter 5 was limited by its questionnaire design, and therefore the depth of participants’ responses was limited. Consequently, future studies should consider utilising semi-structured interview designs to explore relapse following psychological interventions. These interviews could further explore the differences between lapses and relapses, and also explore the perspectives of patients who appear to maintain their treatment gains long-term and remain in remission.

Finally, further research is required to investigate the ‘revolving door’ phenomenon in mental health services. This thesis has identified the extent to which treatment return occurs in one IAPT service, but it is not currently known if there are varying rates of treatment return across different services. In addition, although it was identified that over one third of treatment returning patients experienced a demonstrable relapse upon their re-entry to the mental health service, this is the full extent of what is currently known regarding how relapse is associated with treatment return. Therefore, further research is required to explore this association in greater detail. For example, it may be helpful for future research to explore what prompts patients who relapse to consider, and subsequently decide, on returning for additional treatment. Furthermore, the health economic costs of treatment return are needed to be assessed, in order to fully understand the costs that the ‘revolving door’ process has on mental health services.

**Overall Conclusions**

This thesis has highlighted that relapse is relatively common following evidence-based psychological interventions for depression and anxiety. Despite this, relapse remains a comparatively under-researched area in psychotherapy research. Considering the detrimental effects that relapse can have on patients, through the re-experience of suffering and distress, and on services, through service efficiency being hindered due to frequent treatment return, the importance of further research to understand its occurrence cannot be overstated.

However, this thesis has demonstrated the potential of different research approaches that could be adopted again in the future to further improve our understanding and predictions of relapse. In addition, a number of potential risk factors of relapse have been identified, such as unemployment and work-related stress, which may be potentially fruitful targets for future research. Furthermore, recommendations for how practitioners and services can assist in preventing and understanding relapse have been suggested, including psychoeducation of the differences between lapses and relapses, and the incorporation of structured follow-up into intervention packages.

Although there is clearly a lot more to learn about relapse following psychological interventions, this thesis has demonstrated that our understanding is gradually improving, and that there are a number of promising directions for this understanding to be improved even further. Realisation of this potential could allow services to provide patients with the greatest opportunities possible to prevent relapses from occurring, and thus facilitate treatment gains to be maintained long-term.

**REFERENCES**

References marked with an asterisk indicate studies included in the systematic review in Chapter 2.

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Adadi, A. & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., & Delgadillo, J. (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*, 1–8. https://doi.org/10.1016/j.brat.2017.04.006

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders (3rd ed., revised)*. American Psychiatric Association.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Association.

Andrilla, C. H. A., Patterson, D. G., Garberson, L. A., Coulthard, C., & Larson, E. H. (2018). Geographic variation in the supply of selected behavioural health providers. *American Journal of Preventive Medicine, 54*(6 Suppl 3), S199-S207. https://doi.org/10.1016/j.amepre.2018.01.004

Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology, 80,* 750–765. https://doi.org/10.1037/a0028310

Awa, W. L., Plaumann, M., & Walter, U. (2010). Burnout prevention: a review of intervention programs. *Patient Education and Counseling*, *78*(2), 184–190. https://doi.org/10.1016/j.pec.2009.04.008

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice, 10*, 125-143. https://doi.org/10.1093/clipsy.bpg015

Bagnall, A., Jones, R., Akter, H., & Woodall, J. R. (2016). *Interventions to prevent burnout in high risk individuals: Evidence review.* Public Health England. Retrieved from https://eprints.leedsbeckett.ac.uk/id/eprint/2371/

Baker, C. (2018). *Mental health statistics for England: Prevalence, services and funding*. House of Commons Library. Retrieved from http://allcatsrgrey.org.uk/wp/wpfb-file/sn06988-1-pdf/

Baker, C. (2021). *Mental health statistics: Prevalence, services and funding in England*. House of Commons Library. Retrieved from https://commonslibrary.parliament.uk/research-briefings/sn06988/

Barth, J., Munder, T., Gerger, H., Nuesch, E., Trelle, S., Znoj, H., Juni, P., & Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine, 10*(5), e1001454. https://doi.org/10.1371/journal.pmed.1001454

Barkham, M., Saxon, D., Hardy, G. E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A. D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Ardern, K., Bishop-Edwards, L., Ashley, K., . . . Brazier, J. E. (2021). Person-centred experiential therapy versus cognitive behavioural therapy delivered in the English Improving Access to Psychological Therapies service for the treatment of moderate or severe depression (PRaCTICED): A pragmatic, randomised, non-inferiority trial. *The Lancet Psychiatry, 8*(6), 487–499. https://doi.org/10.1016/S2215-0366(21)00083-3

Beard, J. I. L., & Delgadillo, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety*. https://doi.org/10.1002/da.22931

Beck, A. T. (1964). Thinking and depression: II. Theory and therapy. *Archives of General Psychiatry, 10,* 561–571. https://doi.org/10.1001/archpsyc.1964.01720240015003

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond.* Guilford Press.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8-14. https://doi.org/10.1016/j.npls.2016.01.001

Bennett-Levy, J., Richards, D. A., & Farrand, P. (2010). Low intensity CBT interventions: A revolution in mental health care. In J. Bennett-Levy, D. A. Richards, P. Farrand, H. Christensen, K. M. Griffiths, D. J. Kavanaugh, . . . C. Williams (Eds.), *Oxford guides in* *cognitive behavioural therapy: Oxford guide to low intensity CBT interventions* (pp. 3-18). Oxford University Press

Berk, M., Ng, F., Dodd, S., Callaly, T., Campbell, S., Bernardo, M., & Trauer, T. (2008). The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. *Journal of Evaluation in Clinical Practice, 14(6),* 979-983. https://doi.org/10.1111/j.1365-2753.2007.00921.x

Bhaskar, R. (1975). *A Realist Theory of Science.* Verso.

Bhui, K. S., Dinos, S., Stansfeld, S. A., & White, P. D. (2012). A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. *Journal of Environmental and Public Health*, *2012*, 515874. https://doi.org/10.1155/2012/515874

Blackburn, I. M., James, I. A., Milne, D. L., & Reichelt, F. K. (2001). *Cognitive therapy scale—revised (CTS-R).* Newcastle Cognitive and Behavioural Therapies Centre.

Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W., & Dobson, K. (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review, 41*, 16-26. https://doi.org/10.1016/j.cpr.2015.02.003

Bockting C. L., Schene A. H., Spinhoven P., Koeter M. W., Wouters L. F., Huyser J., & Kamphuis J. H. (2005). Preventing relapse/recurrence in recurrent depression with cognitive therapy: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 73,* 647–657. https://doi.org/10.1037/0022-006x.73.4.647

Boerema, A. M., Cuijpers, P., Beekman, A. T. F., Hellenthal, A., Voorrips, L., & van Straten, A. (2016). Is duration of psychological treatment for depression related to return into treatment? *Social Psychiatry and Psychiatric Epidemiology, 51*(11), 1495-1507. https://doi.org/10.1007/s00127-016-1267-7

Bonde, J. P. E. (2008). Psychosocial factors at work and risk of depression: A systematic review of the epidemiological evidence. *Occupational and Environmental Medicine, 65(7),* 438-445. https://doi.org/10.1136/oem.2007.038430

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., Deisenhofer, A., Lutz, W., & Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes: Development and validation of a prediction model using routinely collected symptom data. *Lancet Digital Health, 3*(4), 231-240. https://doi.org/10.1016/S2589-7500(21)00018-2

Bouton M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. https://doi.org/10.1016/s0006-3223(02)01546-9

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. *British Journal of Psychiatry, 186,* 11-17. http://dx.doi.org/10.1192/bjp.186.1.11

Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30(7),* 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

\*Braga, D. T., Cordioli, A. V., Niederaur, K., & Manfro, G. G. (2005). Cognitive-behavioral group therapy for obsessive–compulsive disorder: a 1-year follow-up. *Acta Psychiatrica Scandinavica, 112*(3), 180-186. https://doi.org/10.1111/j.1600-0447.2005.00559.x

\*Braga, D. T., Manfro, G. G., Niederauer, K., & Cordioli, A. V. (2010). Full remission and relapse of obsessive-compulsive symptoms after cognitive-behavioral group therapy: a two-year follow-up. *Revista Brasileira De Psiquiatria, 32*(2), 164-168. https://doi.org/10.1590/s1516-44462010000200012

Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., Shea, T., & Keller, M. B. (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-year prospective study. *American Journal of Psychiatry, 162,* 1179-1187. https://doi.org/10.1176/appi.ajp.162.6.1179

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Burcusa, S. L., & Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical Psychology Review, 27*, 959-985. https://doi.org/10.1016/j.cpr.2007.02.005

Centre for Mental Health. (2020). *The Economic and Social Costs of Mental Health Problems in 2009/2010*. Retrieved from https://www.centreformentalhealth.org.uk/publications/economic-and-social-costs-mental-health-problems-200910

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16,* 321-357. https://doi.org/10.1613/jair.953

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *Lancet Psychiatry, 3*, 243-250. https://doi.org/10.1016/S2215-0366(15)00471-X

Chellingsworth, M., Farrand, P., & Small, F. (2013). *Relapse prevention: Toolkit recovery guide.* Clinical Education Development and Research, University of Exeter.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2019). Xgboost: Extreme Gradient Boosting. R package version 0.82.1. https://CRAN.R-project.org/package=xgboost

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2021). Xgboost: Extreme Gradient Boosting. R package version 1.4.1.1. https://CRAN.R-project.org/package=xgboost

Cheng, C., Cheung, S., Chio, J. H., & Chan, M. S. (2013). Cultural meaning of perceived control: A meta-analysis of locus of control and psychological symptoms across 18 cultural regions. *Psychological Bulletin, 139(1),* 152-188. https://doi.org/10.1037/a0028596

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry, 13*(2), 153-160. https://doi.org/10.1002/wps.20128

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23,* 318–327. https://doi.org/10.3109/09540261.2011.606803

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, *14*, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Clark, D. M. (2019). *IAPT at 10: Achievements and challenges.* NHS. Retrieved from https://www.england.nhs.uk/blog/iapt-at-10-achievements-and-challenges

Clark, D. M, Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): An analysis of public data. *The Lancet, 391(10121)*, 679-686. https://doi.org/10.1016/S0140-6736(17)32133-5

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, *47*(11), 910–920. https://doi.org/10.1016/j.brat.2009.07.010

Cochrane Collaboration. (2014). *Data Collection Form for Intervention Reviews: RCTS and Non-RCTS, Version 3.* Available at: https://dplp.cochrane.org/data-extraction-forms

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J., & Driessen, E. (2020). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive behavioural versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research, 30(2)*, 137-150. https://doi.org/10.1080/10503307.2018.1563312

Collins, G. S., Dhiman, P., Navarro, C. L. A., Ma, J., Hooft, L., Reitsma, J. B., Logullo, P., Beam, A. L., Peng, L., Van Calster, B., van Smeden, M., Riley, R. D., & Moons, K. G. M. (2021). Protocol for development of a reporting guidelines (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open, 11*, e048008. https://doi.org/10.1136/bmjopen-2020-048008

Collins, G. S. & Moon, K. G. M. (2019). Reporting of artificial intelligence prediction models. *The Lancet, 393*(10181), 1577-1579. https://doi.org/10.1016/S0140-6736(19)30037-6

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*, *102*(3), 148-158. <https://doi.org/10.1002/bjs.9736>

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58,* 10-23. https://doi.org/10.1016%2Fj.brat.2014.04.006

Critical Appraisal Skills Programme. (2018). *CASP Cohort Study Checklist.* Available at: https://casp-uk.net/casp-tools-checklists/

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52(4),* 281–302. https://doi.org/10.1037/h0040957

Cuijpers, P., Berking. M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013a). A meta-analysis of cognitive behavioural therapy for adult depression, alone and in comparison with other treatments. *Canadian Journal of Psychiatry, 58*(7), 376-385. https://doi.org/10.1177/070674371305800702

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry, 15*(3), 245–258. https://doi.org/10.1002/wps.20346

Cuijpers, P., Hollon, S. D., van Straten, A., Berking, M., Bockting, C. L. H., & Andersson, G. (2013b). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation medication? A meta-analysis. *BMJ Open, 26*(3). http://dx.doi.org/10.1136/bmjopen-2012-00254.

Cuijpers, P., Huibers, M., Ebert, D. D., Koole, S. L., & Andersson, G. (2013c). How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders, 149*, 1-13. https://doi.org/10.1016/j.jad.2013.02.030

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S.D., & van Straten, A. (2014). The effects of psychtherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders, 159,* 118-126. https://doi.org/10.1016/j.jad.2014.02.026

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021). Psychotherapies for depression: A network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry, 20*(2), 283-293. https://doi.org/10.1002/wps.20860

Dattilio, F. M., Edwards, D. J. A., & Fishman, D. B. (2010). Case studies within a mixed methods paradigm: Toward a resolution of the alienation between researcher and practitioner in psychotherapy research. *Psychotherapy: Theory, Research, Practice, Training, 47(4),* 427–441. https://doi.org/10.1037/a0021181

De Jong, K., Conijn, J.M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review, 85*, 102002. https://doi.org/10.1016/j.cpr.2021.102002

De Jonge, M., Bockting, C. L. H., Kikkert, M. J., van Djik, M. K., van Schaik, D. J. F., Peen, J., Hollon, S. D., & Dekker, J. J. M. (2019). Preventive cognitive therapy versus care as usual in cognitive behavioural therapy responders: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 87*(6), 521-529. https://doi.org/10.1037/ccp0000395

Delgadillo, J. (2021). Machine learning: A primer for psychotherapy researchers. *Psychotherapy Research, 31*(1), 1-4. https://doi.org/10.1080/10503307.2020.1859638

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (2016). On poverty, politics and psychology: The socioeconomic gradient of mental healthcare utilisation and outcomes. *The British Journal of Psychiatry, 209*(5), 429–430. https://doi.org/10.1192/bjp.bp.115.171017

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M., Nevin, J., O’Hayon, H., Patel, U., Sainty, A., Spencer, P., & McMillan, D. (2018a). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5*(7), 564-572. https://doi.org/10.1016/S2215-0366(18)30162-7

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive-behavioral therapy versus person-centred counselling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88,* 14-24. https://doi.org/10.1037/ccp0000476

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology, 85*(9), 835-853. https://doi.org/10.1037/ccp0000231

Delgadillo, J., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., . . . Ali, S. (2018b). Relapse and recurrence of common mental health problems after low intensity cognitive behavioural therapy: The WYLOW longitudinal cohort study. *Psychotherapy and Psychosomatics, 87*, 116-117. https://doi.org/10.1159/000485386

Department for Levelling Up, Housing and Communities. (2019). *English indices of deprivation.* Retrieved from https://www.gov.uk/government/collections/english-indices-of-deprivation.

DiMauro, F., Domingues, J., Fernandez, G., & Tolin, D. F. (2013a). Corrigendum to “Long-term effectiveness of CBT for anxiety disorders in an adult outpatient clinic sample: A follow-up study.” *Behaviour Research and Therapy, 51(6)*, 332. https://doi.org/10.1016/j.brat.2013.02.003

\*DiMauro, F., Domingues, J., Fernandez, G., & Tolin, D. F. (2013b). Long-term effectiveness of CBT for anxiety disorders in an adult outpatient clinic sample: A follow-up study. *Behaviour Research and Therapy, 51(2)*, 82-86. https://doi.org/10.1016/j.brat.2012.10.003

Elliott, R. (2010). Psychotherapy change process research: Realizing the promise. *Psychotherapy Research, 20(2),* 123–135. https://doi.org/10.1080/10503300903470743

Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence, 20*(1), 18-36. https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x

Evans, M. D., Hollon, S. D., DeRubeis, R. J., Piasecki, J. M., Grove, W. M., Garvey, M. J., & Tuason, V. B. (1992). Differential relapse following cognitive therapy and pharmacotherapy for depression. *Archives of General Psychiatry, 49*(10)*,* 802–808. https://doi.org/10.1001/archpsyc.1992.01820100046009

\*Fava, G. A., Grandi, S., Rafanelli, C., Ruini, C., Conti, S., & Belluardo, P. (2001a). Long-term outcome of social phobia treated by exposure. *Psychological Medicine, 31,* 899-905. https://doi.org/10.1017/s0033291701004020

\*Fava, G. A., Rafanelli, C., Grandi, S., Conti, S., Ruini, C., Mangelli, L., & Belluardo, P. (2001b). Long-term outcome of panic disorder with agoraphobia treated by exposure. *Psychological Medicine, 31,* 891-898. https://doi.org/10.1017/s0033291701003592

\*Fava, G. A., Zielezny, M., Savron, G., & Grandi, S. (1995). Long-term effects of behavioural treatment for panic disorder with agoraphobia. *British Journal of Psychiatry, 166,* 87-92. https://doi.org/10.1192/bjp.166.1.87

Finegan, M., Firth, N., & Delgadillo, J. (2020). Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: The role of area-level income and crime. *Psychotherapy Research, 30*(4), 546-554. https://doi.org/10.1080/10503307.2019.1649500

Fletcher, A. J. (2017). Applying critical realism in qualitative research: Methodology meets method. *International Journal of Social Research Methodology, 20*(2), 181-194. https://doi.org/10.1080/13645579.2016.1144401

Frank, E., Kupfer, D. J., Buysse, D. J., Swartz, H. A., Pilkonis, P. A., Houck, P. R., Rucci, P., Novick, D. M., Grochocinski, V. J., & Stapf, D. M. (2007). Randomized trial of weekly, twice-monthly, and monthly interpersonal psychotherapy as maintenance treatment for women with recurrent depression. *The American Journal of Psychiatry, 164*(5), 761-767. http://dx.doi.org/10.1176/appi.ajp.164.5.761.

Frank, E., Kupfer, D. J., Perel, J. M., & Cornes, C. (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry, 47*(12), 1093-1099. http://dx.doi.org/10.1001/archpsyc.1990.01810240013002

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W, Rush, J., & Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry, 48*(9), 851-855. https://doi.org/10.1001/archpsyc.1991.01810330075011

French, L. R. M., Turner, K. M., Dawson, S., & Moran, P. (2017). Psychological treatment of depression and anxiety in patients with co-morbid personality disorder: A scoping study of trial evidence. *Personality and Mental Health, 11(2),* 101-117. https://doi.org/10.1002/pmh.1372

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29(5)*, 1189-1232. https://doi.org/10.1214/aos/1013203451

Fryers, T., Melzer, D., & Jenkins, R. (2003). Social inequalities and the common mental disorders. *Social Psychiatry and Psychiatric Epidemiology, 38*(5), 229–237. https://doi.org/10.1007/s00127-003-0627-2

Freud, S. (1917). Mourning and melancholia. *Standard Edition*, *14*, 237–259.

Furukawa, T. A., Weitz, E. S., Tanaka, S., Hollon, S. D., Hofmann, S. G., Andersson, G., Twisk, J., DeRubeis, R. J., Dimidjian, S., Hegerl, U., Mergl, R., Jarrett, R. B., Vittengl, J. R., Watanabe, N., & Cuijpers, P. (2017). Initial severity of depression and efficacy of cognitive-behavioural therapy: Individual-participant data meta-analysis of pill-placebo-controlled trials. *British Journal of Psychiatry, 210*(3), 190-196. https://doi.org/10.1192/bjp.bp.116.187773

Gearing, R E., Schwalbe, C. S. J., Lee, R. H., & Hoagwood, K. E. (2013). The effectiveness of booster sessions in CBT treatment for child an adolescent mood and anxiety disorders. *Depression and Anxiety, 30*(9), 800-808. https://doi.org/10.1002/da.22118

Goldner, E. M., Jones, W., & Fang., M. L. (2011). Access to and waiting time for psychiatrist services in Canadian urban area: A study in real time. *Canadian Journal of Psychiatry, 56*(8), 474-480. https://doi.org/10.1177%2F070674371105600805

Goldstein, D. J., Potter, W. Z., Ciraulo, D. A., & Shader, R. I. (2011). Biological theories of depression and implications for current and new treatments. *Pharmacotherapy of Depression: Second Edition*, 1–32. https://doi.org/10.1007/978-1-60327-435- 7\_1

Goodman, W. K., Price, L. H., Rasmussen, A. S., Mazure C., Feleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: Development, use, and reliability. *Archives of General Psychiatry, 46,* 1006-1016. https://doi.org/10.1001/archpsyc.1989.01810110048007

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT psychological wellbeing practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy, 63,* 43-54. https://doi.org/10.1016/j.brat.2014.08.009

Greenberg, L. S. (1991). Research on the process of change. *Psychotherapy Research, 1(1),* 3–16. https://doi.org/10.1080/10503309112331334011

Guy, W. (1976). *Assessment manual for psychopharmacology.* United States Government Printing Office.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications, 73,* 220-239. https://doi.org/10.1016/j.eswa.2016.12.035

Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatra Scandinavica, 122,* 184-191. https://doi.org/10.1111/j.1600-0447.2009.01519.x

Harkness, K. L., Bagby, R. M. and Kennedy, S. H. (2012). Childhood maltreatment and differential treatment response and recurrence in adult major depressive disorder. *Journal of Consulting and Clinical Psychology, 80,* 342–353. https://doi.org/10.1037/a0027665

Harkness, K. L., Theriault, J. E., Stewart, J. G., & Bagby, R. M. (2014). Acute and chronic stress exposure predicts 1-year recurrence in adult outpatients with residual depression symptoms following response to treatment. *Depression and Anxiety, 31,* 1-8. https://doi.org/10.1002/da.22177

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* Springer.

Hatton, C. M., Paton, L. W., McMillan, D., Cussens, J., Gilbody, S., & Tiffin, P. (2019). Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare. *Journal of Affective Disorders, 246*, 857-860. https://doi.org/10.1016/j.jad.2018.12.095

Health Education England. (2020). *Adult IAPT Workforce Census 2020.* NHS. Retrieved from https://www.hee.nhs.uk/sites/default/files/documents/IAPT%20Census%202020%20-%20National%20report%20Final%20-%20August%202021.pdf

Health Resources and Services Administration. (2016). *National Projections of Supply and Demand for Selected Behavioral Health Practitioners: 2013-2025.* Health Resources and Services Administation.

\*Heldt, E., Kipper, L., Blaya, C., Salum, G. A., Hirakata, V. N., Otto, M. W., & Manfro, G. G. (2011). Predictors of relapse in the second follow-up year post cognitive-behavior therapy for panic disorder. *Revista Brasileira de Psiquiatria, 33(1),* 23-29. https://doi.org/10.1590/S1516-44462010005000005

Heldt E., Manfro, G. G., Kipper, L., Blaya, C., Isolan, L., & Otto, M. W. (2006). One-year follow-up of pharmacotherapy resistant patients with panic disorder treated with cognitive-behaviour therapy: Outcome and predictors of remission. *Behaviour Research and Therapy, 44(5),* 657-665. https://doi.org/10.1016/j.brat.2005.05.003

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioural therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69*(4), 621-632. https://doi.org/10.4088/jcp.v69n0415

Hollon, S. D., Stewart, M. O., & Strunk, D. (2006). Enduring effects for cognitive behaviour therapy in the treatment of depression and anxiety. *Annual Review of Psychology, 57,* 285-315. https://doi.org/10.1146/annurev.psych.57.102904.190044

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*, 678-685. https://doi.org/10.1037/0022-006X.61.4.678

Improving Access to Psychological Therapies. (2019). *Talking therapies: New statistics show an increase in referrals, numbers starting treatment and recovery rates during 2018-19.* NHS Digital. Retrieved from https://digital.nhs.uk/news-and-events/latest-news/iapt-2018-19

Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine, 46*(12), 2455-2465. https://doi.org/10.1017/S0033291716001367

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59(1)*, 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Jarrett, R. B., Kraft, D., Doyle, J., Foster, B. M., Eaves, G., & Silver, C. (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase. *Archives of General Psychiatry, 58*(4), 381-388. https://doi.org/10.1001/archpsyc.58.4.381

Jarrett, R. B., Vittengl, J. R., & Clark, L. A. (2008). How much cognitive therapy, for which patients, will prevent depressive relapse? *Journal of Affective Disorders, 111*(2-3), 185-1192. http://dx.doi.org/10.1016/j.jad.2008.02.011.

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review, 67,* 78–93. https://doi.org/10.1016/j.cpr.2018.08.004

Johnston, D. A., Harvey, S. B., Glozier, N., Calvo, R. A., Christensen, H., & Deady, M. (2019). The relationship between depression symptoms, absenteeism and presenteeism. *Journal of Affective Disorders, 256*, 536-540. https://doi.org/10.1016/j.jad.2019.06.041

Kazdin, A. E. (2018). Treatment gap: Barriers to providing and receiving services. In A. E. Kazdin (Ed.), *Innovations in psychosocial interventions and their delivery: Leveraging cutting-edge science to improve the world’s mental health* (pp. 51-76). Oxford University Press.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44*, 540-548. https://doi.org/10.1001/archpsyc.1987.01800180050009

Kellett, S., Simmonds-Buckley, M., Limon, E., Stride, C., Hughes, L., Hague, J., & Millings, A. (2021). Defining the assessment and treatment competencies to deliver low intensity cognitive behavioural therapy: A multi-centre validation study. *Behavior Therapy, 52*(1), 15-27.https://doi.org/10.1016/j.beth.2020.01.006

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry, 156*(6), 837-841. https://doi.org/10.1176/ajp.156.6.837

Kendler, K. S., Thornton, L. M., & Gardner, C. O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the “kindling” hypothesis. *American Journal of Psychiatry, 157*(8), 1243-1251. https://doi.org/10.1176/appi.ajp.157.8.1243

Kessler, R. C. (1997). The effects of stressful life events on depression. *Annual Reviews of Psychology, 48*, 191-214. https://doi.org/10.1146/annurev.psych.48.1.191

Kilcullen, J. R., Castonguay, L. G., Janis, R. A., Hallquist, M. N., Hayes, J. A., & Locke, B. D. (2021). Predicting future courses of psychotherapy within a grouped LASSO framework*. Psychotherapy Research, 31*(1), 63-77. https://doi.org/10.1080/10503307.2020.1762948

King, M., Marston, L., & Bower, P. (2014). Comparison of non-directive counselling and cognitive behaviour therapy for patients presenting in general practice with an ICD-10 depressive episode: a randomized control trial. *Psychological Medicine, 44*(9), 1835–1844. https://doi.org/10.1017/S0033291713002377

Kordy, H., Wolf, M., Aulich, K., Bürgy, M., Hegerl, U., Hüsing, J., Puschner, B., Rummel-Kluge C., Vedder, H., & Backenstrass, M. (2016). Internet-delivered disease management for recurrent Depression: Multicenter randomized controlled trial. *Psychotherapy and Psychosomatics, 85,* 91-98. https://doi.org/10.1159/000441951

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PhQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9)*,* 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317-325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159. https://doi.org/10.2307/2529310

Law, P. C. F., Too, L. S., Butterworth, P., Witt, K., Reavley, N., & Milner, A. J. (2020). A systematic review on the effect of work-related stressors on mental health of young workers. *International Archives of Occupational and Environmental Health, 93,* 611-622. https://doi.org/10.1007/s00420-020-01516-7

Layard, R., & Clark, D. M. (2014). *Thrive: The power of psychological therapy*. Allen Lane, Penguin Group.

Leaker, D., & Kumar, K., (2021). *Sickness absense in the UK labour market: 2020.* Office for National Statistics. Retrieved from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/sicknessabsenceinthelabourmarket/2020

Leighton, S. P., Upthegrove, R., Krishnadas, R., Benros, M. E., Broome, M. R., Gkoutos, G. V., Liddle, P. F., Singh, S. P., Everard, L., Jones, P. B., Fowler, D., Sharma, V., Freemantle, N., Christensen, R. H. B., Albert, N., Nordentoft, M., Schwannauer, M., Cavanagh, J., Gumley, A. I., . . . Mallikarjun, P. K. (2019). Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: A machine learning approach. *The Lancet Digital Health, 1*(6), 261-270. https://doi.org/10.1016/S2589-7500(19)30121-9

Lemmens, L. H. J. M., van Bronswijk, S. C., Peeters, F., Arntz, A., Hollon, S. D., & Huijbers, M. J. H. (2019). Long-term outcome acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: Follow-up of a randomized controlled trial. *Psychological Medicine, 49*(3), 465-473. https://doi.org/10.1017/S0033291718001083

Leon, A. C., Shear, M. K., Klerman, G. L., Portera, L., Rosenbaum, J. F., & Goldenberg, I. (1993). A comparison of symptom determinants of patient and clinician global ratings in patients with panic disorder and depression. *Journal of Clinical Psychopharmacology, 13(5),* 327–331. https://doi.org/10.1097/00004714-199310000-00005

Levy, H. C., O’Bryan, E. M., & Tolin, D. F. (2021). A meta-analysis of relapse rates in cognitive-behavioral therapy for anxiety disorder. *Journal of Anxiety Disorder, 81,* 102407. https://doi.org/10.1016/j.janxdis.2021.102407

Lewinsohn, P. M. (1974). A behavioural approach to depression. In R. J. Friedman & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research.* John Wiley & Sons Inc.

Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports, 8,* 2861. https://doi.org/10.1038/s41598-018-21243-x

\*Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., von Witzleben I., Schroeder, B., & Fiegenbaum, W. (2005). Who comes, who stays, who profits? Predicting refusal, dropout, success, and relapse in a short intervention for social phobia. *Psychotherapy Research, 15(3),* 210-225. https://psycnet.apa.org/doi/10.1080/10503300512331387834

Lopez, V., Fernandez, A., Garcia, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250,* 113-141. https://doi.org/10.1016/j.ins.2013.07.007

Loveless, L. (2019). *Employment Advisers in Improving Access to Psychological Therapies: Process Evaluation Report.* Department for Work and Pensions. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/817628/employment-advisers-in-improving-access-to-psychological-therapies-process-evaluation-report.pdf

Lovell, K., & Richards, D. (2012). *A recovery programme for depression*. Rethink Mental Illness.

Lucock, M., Bartys, S., Cupac, J., Delgadillo, J., Denton, C., Gaines, S., McMillan, D., Prestwich, A., & Stebbings, R. (2018). Using implementation intentions to prevent relapse after psychological treatment for depression – The SMArT intervention. *Behavioural and Cognitive Psychotherapy*, *46*(5), 626–632. https://doi.org/10.1017/S1352465818000255

Lucock, M., Borthwick, R., Cupac, J., Elliott, R., Howell, R., Kendal, S., Khan, W., Sandford, D., & Tolley, B. (2021). Using implementation intentions to prevent relapse after remission from psychological treatment for depression: The SMArT intervention. *Psychotherapy Research.* Advance online publication. https://doi.org/10.1080/10503307.2021.1959079

Malins, S., Biswas, S., Patel, S., Levene, J., Moghaddam, N., & Moriss, R. (2020). Preventing relapse with personalized smart-messaging after cognitive behavioural therapy: A proof-of-concept evaluation. *British Journal of Clinical Psychology, 59*(2), 241-259. https://doi.org/10.1111/bjc.12244

McLeod, J. (2011). The role of qualitative methods in outcome research. In J. McLeod (Ed.), *Qualitative research in counselling and psychotherapy* (2nd ed., pp. 161–180). Sage.

McMillan, D., Gilbody, S., & Richards, D. A. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders, 127,* 122-129. https://doi.org/10.1016/j.jad.2010.04.030

Midgley, N., Ansaldo, F., & Target, M. (2014). The meaningful assessment of therapy outcomes: Incorporating a qualitative study into a randomized controlled trial evaluating the treatment of adolescent depression*. Psychotherapy, 51(1),* 128–137. https://doi.org/10.1037/a0034179

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6,* e1000097. https://doi.org/10.1136/bmj.b2535

Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine, 162(2),* 1-73. https://doi.org/10.7326/M14-0698

Moriarty, A. S., Paton, L. W., Snell, K. I. E., Riley, R. D., Buckman, J. E. J., Gilbody, S., Chew-Graham, C. A., Ali, S., Pilling, S., Meader, N., Philips, B., Coventry, P. A., Delgadillo, J., Richards, D. A., Salisbury, C., & McMillan, D. (2021). 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), 12. https://doi.org/10.1186/s41512-021-00101-x

Morgan, D.L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research, 3*(1), 112-121. https://doi.org/10.1177/104973239300300107

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*(5), 461-464. https://doi.org/10.1192/bjp.180.5.461

National Collaborating Centre for Mental Health (2020). *The improving access to psychological therapies manual*. NHS. Retrieved from https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

National IAPT Team. (2015). *National curriculum for the education of Psychological Wellbeing Practitioners, Third edition.* University College London.

National Institute for Health and Care Excellence. (2011). *Common mental health disorders: Identification and pathways to care.* National Institute for Health and Care Excellence. Retrieved from https://www.nice.org.uk/guidance/cg123/resources/common-mental-health-problems-identification-and-pathways-to-care-pdf-35109448223173

Netterstrøm, B., Conrad, N., Bech, P., Fink, P., Olsen, O., Rugulies, R., & Stansfield, S. A. (2008). The relation between work-related psychosocial factors and the development of depression. *Epidemiologic Reviews, 30(1),* 118-132. https://doi.org/10.1093/epirev/mxn004

Nieuwenhuijsen, K., Bruinvels, D., & Frings-Dresen, M. (2010). Psychosocial work environment and stress-related disorders, a systematic review. *Occupational Medicine, 60(4),* 277-286. https://doi.org/10.1093/occmed/kqq081

Nutt, D. J., Ballenger, J. C., Sheehan, D., & Wittchen, H. U. (2002). Generalized anxiety disorder: comorbidity, comparative biology and treatment. *The International Journal of Neuropsychopharmacology, 5*(4), 315–325. https://doi.org/10.1017/S1461145702003048

Ogawa, S., Furukawa, T. A., Nakano, Y., Funayama, T., Watanabem N., Noguchi, Y., & Sasaki, M. (2010). Interoceptive hypersensitivity as prognostic factor among patients with panic disorder who have received cognitive behavioral therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 41,* 325-329. https://doi.org/10.1016/j.jbtep.2010.03.002

Olatunji, B. O., Cisler, J. M., & Tolin, D. F. (2007). Quality of life in the anxiety disorders: a meta-analytic review. *Clinical Psychology Review, 27*(5), 572–581. https://doi.org/10.1016/j.cpr.2007.01.015

\*Otto, M. W., Pollack, M. H., & Sabatino, S. A. (1996). Maintenance of remission following cognitive behavior therapy for panic disorder: Possible deleterious effects of concurrent medication treatment. *Behavior Therapy, 27*, 473-482. https://doi.org/10.1016/S0005-7894(96)80028-1

Otto, M. W., Smits, J. A., & Reese, H. E. (2005). Combined psychotherapy and pharmacotherapy for mood and anxiety disorders in adults: review and analysis. *Clinical Psychology: Science and Practice, 12*, 72-86. https://doi.org/10.1093/clipsy.bpi009

Papakostas, G. I., Petersen, T., Mahal, Y., Mischoulon, D., Nierenberg, A. A., & Fava, M. (2004). Quality of life assessments in major depressive disorder: A review of the literature. *General Hospital Psychiatry, 26,* 13–17. https://doi.org/10.1016/j.genhosppsych.2003.07.004

Papini, S., Pisner, D., Shumake, J., Powers, M. B., Beevers, C. G., Rainey, E. E., Smits, J. A. J., & Warren, A. M. (2018). Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *Journal of Anxiety Disorders, 60,* 35-42. https://doi.org/10.1016/j.janxdis.2018.10.004

Paykel, E. S. (2008). Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience, 10*(4), 431-437. https://doi.org/10.31887/DCNS.2008.10.4/espaykel

Piet, J., & Hougaard, E. (2011). The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 31,* 1032-1040. https://doi.org/10.1016/j.cpr.2011.05.002

Pratt, L. A., Druss, B. G., Manderscheid, R. W., & Walker, E.R. (2016). Excess mortality due to depression and anxiety in the United States: Results from a nationally representative survey. *General Hospital Psychiatry, 39,* 39-45. https://doi-org.sheffield.idm.oclc.org/10.1016/j.genhosppsych.2015.12.003

Reeder, K., Park., A.L., & Chorpita B.F. (2020). Turning back to treatment: The effect of attendance and symptom outcomes on subsequent service use. *Administration and Policy in Mental Health and Mental Health Services Research, 47*, 641-647. https://doi.org/10.1007/s10488-020-01032-3

Reimherr, F. W., Amsterdam, J. D., Quitkin, F. M., Rosenbaum, J. F., Fava, M., Zajecka, J., Beasley, C. M., Michelson, D., Roback, P., & Sundell, K. (1998). Optimal length of continuation therapy in depression: a prospective assessment during long-term fluoxetine treatment. *American Journal of Psychiatry, 155*(9), 1247–1253. https://doi.org/10.1176/ajp.155.9.1247

Reynolds III, C.F., Dew, M. A., Pollock, B. G., Mulsant, B. H., Frank, E., Miller, M. D., Houck, P. R., Mazumdar, S., Butters, M. A., Stack, J. A., Schlernitzauer, M. A., Whyte, E. M., Gildengers, A., Karp, J., Lenze, E., Szanto, K., Bensasi, S., & Kupfer, D. J. (2006). Maintenance treatment of major depression in old age. *The New England Journal of Medicine, 354*(11), 1130–1138. http://dx.doi.org/10.1056/NEJMoa052619

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort evaluation. *Journal of Affective Disorders, 133,* 51–60. https://doi.org/10.1016/j.jad.2011.03.024

Riley, R. D., Debray, T. P. A., Collins, G. S., Archer, L., Ensor, J., van Smeden, M., & Snell, K. I. E. (2021). Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Statistics in Medicine, 40*(19), 4230-4251. https://doi.org/10.1002/sim.9025

Riley, R. D., Snell, K. I. E., Ensore, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2018). Minimum sample size for developing a multivariable prediction model: Part II – binary and time-to-event outcomes. *Statistics in Medicine, 38*(7), 1276-1296. https://doi.org/10.1002/sim.7992

Ritchie, J., & Spencer, L. (2002). Qualitative data analysis for applied policy research. In A.M. Huberman & M.B. Miles (Eds.), *The qualitative researcher’s companion* (pp. 305-329). SAGE Publications Inc.

Robinson, L., Delgadillo, J., & Kellett, S. (2019). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2019.1566676

Rodgers, M., Asaria, M., Walker, S., McMillan, D., Lucock, M., Harden, M., Palmer, S., & Eastwood, A. (2012). The clinical effectiveness and cost-effectiveness of low-intensity psychological interventions for the secondary prevention of relapse after depression: a systematic review. *Health Technology Assessment, 16*(28)*,* 1–130. https://doi.org/10.3310/hta16280

Roscoe, J. (2019). Has IAPT become a bit like Frankenstein’s monster? *CBT Today,* *47*(1), 16.

Rose, S. (2018). Machine learning for prediction in electionic health data. *JAMA Network Open, 1*(4), e181404. https://doi.org/10.1001/jamanetworkopen.2018.1404

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy, 36*(2), 129-147. https://doi.org/10.1017/S1352465808004141

Rush, A. J., Kraemer, H. C., Sackeim, H. A., Fava, M., Trivedi, M. H., Frank, E., Ninan, P. T., Thase, M. E., Gelenberg, A. J., Kupfer, D. J., Regier, D. A., Rosenbaum, J. F., Ray, O., & Schatzberg, A. F. (2006). Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology, 31*(9), 1841–1853. https://doi.org/10.1038/sj.npp.1301131

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55(1)*, 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Santini, Z. I., Koyanagi, A., Tyrovolas, S., Catherine, M., & Haro, J. M. (2015). The association between social relationships and depression: A systematic review. *Journal of Affective Disorders, 175*, 53-65. https://doi.org/10.1016/j.jad.2014.12.049

Scott, J., Palmer, S., Paykel, E., Teasdale, J., & Hayhurst, H. (2003). Use of cognitive therapy for relapse prevention in chronic depression: Cost effectiveness study. *British Journal of Psychiatry, 182,* 221–227. https://doi.org/10.1192/bjp.182.3.221

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse.* Guilford Press.

Segal, Z. V., Williams, J. M. G, Teasdale, J. D., & Gemar, M. (1996). A cognitive science perspective on kindling and episode sensitization in recurrent affective disorder. *Psychological Medicine, 26,* 371-380. https://doi.org/10.1017/S0033291700034760

Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. A. (1996). The measurement of disability. *International Clinical Psychopharmacology, 11*(3), 89–95. https://doi.org/10.1097/00004850-199606003-00015

Shortliffe, E. H. & Sepúlveda, M. J. (2018). Clinical decision support in the era of Artificial Intelligence. *JAMA, 320*, 2199-2200. https://doi.org/10.1001/jama.2018.17163

Siddall, L. B., Haffey, N. A., & Feinman, J. A. (1988). Intermittent brief psychotherapy in an HMO setting. *American Journal of Psychotherapy, 42*(1), 96–106. https://doi.org/10.1176/appi.psychotherapy.1988.42.1.96

Silva, M., Loureiro, A., & Cardoso, G. (2016). Social determinants of mental health: A review of the evidence. *The European Journal of Psychiatry, 30*(4), 259–292.

Somers, J. M., Goldner, E. M., Waraich, P., & Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: A systematic review of the literature. *The Canadian Journal of Psychiatry, 51*(2), 100-113. https://doi.org/10.1177/070674370605100206

Spitzer, R. L., Kroenke, R., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Stansfield, S., & Candy, B. (2006). Psychosocial work environment and mental health – A meta-analytic review. *Scandinavian Journal of Work, Environment and Health, 32(6),* 443-462. https://doi.org/10.5271/sjweh.1050

Stekhoven, D. J. (2013). *MissForest: Nonparametric missing value imputation using random forest.* R package version 1.4. Retrieved from https://cran.r-project.org/web/packages/missForest/

Stekhoven, D. J., & Bühlmann, R. (2012). MissForest: Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112-118. https://doi.org/10.1093/bioinformatics/btr597

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R., & Morganstein, D. (2003). Cost of lost productive work time among US workers with depression. *JAMA*, *289*(23), 3135–3144. https://doi.org/10.1001/jama.289.23.3135

Steyerberg, E. W., Bleeker, S. E., Moll, H. A., Grobbee, D. E., & Moons, K. G. M. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology, 56*(5), 441-447. https://doi.org/10.1016/s0895-4356(03)00047-7

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods, 8*(4), 537-553. https://doi.org/10.1002/jrsm.1260

Tahir, M. A., Kittler, J., Mikolajczyk, K., & Zhou, Y. (2009). A multiple expert approach to the class imbalance problem using inverse random under sampling. In J.A. Benediktsson, J. Kittler & F. Roli. (Eds.), *Multiple Classifier Systems: MCS 2009: Lecture Notes in Computer Science: Vol 5519* (pp. 82-91). Berlin, Heidelberg: Springer.

Thase, M. E., Simons, A. D., McGeary, J., Cahalane, J. F., Hughes, C., Harden, T., & Friedman, E. (1992). Relapse after cognitive behavior therapy of depression: potential implications for longer courses of treatment. *American Journal of Psychiatry, 149,* 1046–1052. https://doi.org/10.1176/ajp.149.8.1046

Too, L. S., Spittal, M. J., Bugeja, L., Reifels, L., Butterworth, P., & Pirkis, J. (2019). The association between mental disorders and suicide: A systematic review and meta-analysis of record linkage studies. *Journal of Affective Disorders, 259*, 302-313. https://doi.org/10.1016/j.jad.2019.08.054

Trautmann, S., Rehm, J., & Wittchen, H. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Reports, 17*(9), 1245-1249. https://doi.org/10.15252/embr.201642951

Triggle, N. (2019, December 5). *Hidden waits ‘leave mental health patients in limbo’.* British Broadcasting Corporation. Retrieved from https://www.bbc.co.uk/news/health-50658007

van Dis, E. A. M., van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L. H., van den Heuvel, R. M., Cuijpers, P., & Engelhard, I. M. (2019). Long-term outcomes of cognitive behavioural therapy for anxiety-related disorders: A systematic review and meta-analysis. *JAMA Psychiatry, 77(3),* 265-273. https://doi.org/10.1001/jamapsychiatry.2019.3986

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine, 15,* e1002689. https://doi.org/10.1371/journal.pmed.1002689

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9*, 215-248. https://doi.org/10.1146/annurev-clinpsy-050212-185542

Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive behavioural therapy's effects. *Journal of Consulting and Clinical Psychology, 75,* 475-488. https://doi.org/10.1037/0022-006X.75.3.475

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2009). Continuation-phase cognitive therapy's effects on remission and recovery from depression. *Journal of Consulting and Clinical Psychology, 77*(2), 367–371. https://doi.org/10.1037/a0015238

Wacharasak, S. (2019). *A collection of oversampling techniques for class imbalance problem based on SMOTE.* R package version 1.3.1. Retrieved from https://cran.r-project.org/web/packages/smotefamily.

Waddell, G., & Burton, A. K. (2006). *Is work good for your health and well-being?* London, UK: The Stationery Office.

Wang, J., Mann, F., Lloyd-Evans, B., Ma, R., & Johnson, S. (2018). Associations between loneliness and perceived social support and outcomes of mental health problems: A systematic review. *BMC Psychiatry, 18(1),* 156. https://doi.org/10.1186/s12888-018-1736-5

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2020). Improving access to Psycholoical Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology, 60*(1), 1-37. https://doi.org/10.1111/bjc.12259

Walker, E. R., McGee, R. E., & Druss, B. G. (2015). Mortality in mental disorders and global disease burden implications: A systematic review and meta-analysis. *JAMA Psychiatry, 72*(4), 334–341. https://doi.org/10.1001/jamapsychiatry.2014.2502

Warshaw, M. G., Dyck, I., Allsworth, J., Stout, R. L., & Keller, M. B. (2001). Maintaining reliability in a long-term psychiatric study: An ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. *Journal of Psychiatric Research, 35(5),* 297-305. https://doi.org/10.1016/S0022-3956(01)00030-9

Warshaw, M. G., Keller, M. B., & Stout, R. L. (1994). Reliability and validity of the longitudinal interval follow-up evaluation for assessing outcome of anxiety disorders. *Journal of Psychiatric Research, 28(6),* 531-545. https://doi.org/10.1016/0022-3956(94)90043-4

Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2000). *Comprehensive guide to interpersonal psychotherapy.* Basic Books.

Whisman, M. A. (1990). The efficacy of booster maintenance sessions in behavior therapy: Review and methodological critique. *Clinical Psychology Review, 10*(2), 155-170. https://doi.org/10.1016/0272-7358(90)90055-F

White, K. S., Payne, L. A., Gorman, J. M., Shear, M. K., Woods, S. W., Saksa, J. R., & Barlow, D. H. (2013). Does maintenance CBT contribute to long-term treatment response of panic disorder with or without agoraphobia? A randomised controlled clinical trial. *Journal of Consulting and Clinical Psychology, 81(1),* 47-57. https://doi.org/10.1037/a0030666

Wojnarowski, C., Firth, N., Finegan, M., & Delgadillo, J. (2019). Predictors of depression relapse and recurrence after cognitive behavioural therapy: A systematic review and meta-analysis. *Behavioural and Cognitive Psychotherapy, 47*(5), 514-529. https://doi.org/10.1017/S1352465819000080

World Health Organization. (2017). *Depression and Other Common Mental Disorders: Global Health Estimates.* Geneva, Switzerland: World Health Organization.

Wynants, L., van Smeden, M., McLernon, D. J., Timmerman, D., Steyerberg, E. W., & Van Calster, B. (2019). Three myths about risk thresholds for prediction models. *BMC Medicine, 17*, 192. https://doi.org/10.1186/s12916-019-1425-3

Yarkoni, T. & Westfall, J. (2017). Choosing prediction over explanation in Psychology: Lessons from Machine Learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

|  |  |
| --- | --- |
| **Table A1**  *Example Systematic Review Search Strategy (PsycINFO Database)* | |
| Search Term | Returned Papers |
| Cognitive Behavior Therapy.sh | 17973 |
| Cognitive Behavior\* Therap\*.ti,ab | 15989 |
| Cognitive Behaviour\* Therap\*.ti,ab | 4651 |
| CBT.ti,ab | 12487 |
| 1 OR 2 OR 3 OR 4 | 27798 |
| Relapse (Disorders).sh | 6552 |
| Relaps\*.ti,ab | 26858 |
| Recurr\*.ti,ab | 31063 |
| Recrudesc\*.ti,ab | 155 |
| 6 OR 7 OR 8 OR 9 | 56983 |
| Prediction.sh | 19349 |
| Predict\*.ti,ab | 422702 |
| 11 OR 12 | 424164 |
| Anxiety.sh | 57892 |
| exp Anxiety Disorders | 79555 |
| Separation Anxiety.sh | 1505 |
| Separation Anxiety Disorder.sh | 296 |
| Mutism.sh | 413 |
| exp Phobias | 12649 |
| Fear.sh | 17424 |
| Social Phobia.sh | 4403 |
| exp Panic | 1989 |
| Panic Disorder.sh | 7505 |
| Agoraphobia.sh | 2845 |
| Generalized Anxiety Disorder.sh | 2637 |
| Posttraumatic Stress Disorder.sh | 30613 |
| Complex PTSD.sh | 209 |
| Obsessive Compulsive Disorder.sh | 13329 |
| Anxiet\*.ti,ab | 183285 |
| Anxious.ti,ab | 20063 |
| Selective\* Mutism.ti,ab | 414 |
| Phobi\*.ti,ab | 15194 |
| Panic\*.ti,ab | 16186 |
| Fear.ti,ab | 60139 |
| Agoraphobi\*.ti,ab | 4897 |
| Posttraumatic Stress.ti,ab | 26909 |
| Post Traumatic Stress.ti,ab | 10633 |
| Post-Traumatic Stress.ti,ab | 10633 |
| PTSD.ti,ab | 29500 |
| Obsessive-Compulsive\*.ti,ab | 18992 |
| Obsessive Compulsive\*.ti,ab | 18992 |
| OCD.ti,ab | 10008 |
| 14…OR…42 | 301314 |
| 5 AND 10 AND 13 AND 43 | 65 |

**APPENDIX A**

**APPENDIX B**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table B1**  *Risk of Bias Ratings for Included Studies (Adapted from CASP Checklist for Cohort Studies)* | | | | | | | | | | | | |
| Study | Clear  issue | Sample | Predictor measures accuracy | Relapse measure accuracy | Confounding factors  identified | Confounding factors considered | Complete follow-up | Long enough follow-up | Complete  results | Precise  results | Results  believability | Overall  rating |
| Braga et al. (2005) | Good | Good | Good | Good | Good | Poor | Good | Good | Good | Good | Good | Good |
| Braga et al. (2010) | Good | Good | Good | Unclear | Good | Poor | Good | Good | Good | Good | Good | Good |
| DiMauro et al. (2013b) | Good | Good | Poor | Unclear | Good | Poor | Poor | Good | Good | Poor | Poor | Poor |
| Fava et al. (1995) | Good | Poor | Good | Good | Poor | Poor | Poor | Good | Good | Poor | Good | Fair |
| Fava et al. (2001a) | Good | Poor | Good | Good | Good | Good | Poor | Good | Good | Poor | Good | Good |
| Fava et al. (2001b) | Good | Poor | Good | Good | Good | Good | Poor | Good | Good | Poor | Good | Good |
| Heldt et al. (2011) | Good | Good | Good | Good | Poor | Poor | Good | Good | Good | Good | Good | Good |
| Lincoln et al. (2005) | Good | Unclear | Good | Good | Good | Good | Poor | Good | Good | Good | Good | Good |
| Otto et al. (1996) | Good | Unclear | Good | Good | Good | Good | Good | Good | Good | Poor | Good | Good |

**APPENDIX C**

|  |  |  |  |
| --- | --- | --- | --- |
| **Table C1**  *Significant and Non-Significant Predictors of Relapse of Anxiety Disorders after CBT* | | | |
| Study | Primary Disorder | Non-significant predictors (p>0.05) | Significant Predictors (p<0.05) |
| Braga et al. (2005) | OCD | Baseline intensity of overvalued ideas; baseline symptom severity; age at onset; co-morbid disorder | Degree of improvement; partial remission |
| Braga et al. (2010) | OCD | - | Partial remission |
| DiMauro et al. (2013b) | Mixed | Pre-treatment severity; post-treatment severity; post-treatment improvement; pre-treatment work functioning; pre-treatment family functioning; pre-treatment social functioning: post-treatment work functioning; post-treatment family functioning; post-treatment social functioning | - |
| Fava et al. (1995) | PD (with AG) | Age; gender; social class; marital status; education; employment; duration of disorder; co-morbid psychiatric disorder; baseline level of panic; baseline level of agoraphobia; baseline level of depressed mood; benzodiazepine use at end of treatment; residual depression | Personality disorder; residual agoraphobia |
| Fava et al. (2001a) | SA | Age; gender; social class; education; duration of illness; co-morbid psychiatric disorder; baseline level of social phobia; baseline level of generalized anxiety; baseline level of somatic anxiety; baseline level of depressed mood; post-treatment level generalized anxiety; post-treatment level of somatic anxiety; post-treatment level of depressed mood; antidepressant use at end of treatment | Personality disorder; residual social phobia; benzodiazepine use at end of treatment |
| Fava et al. (2001b) | PD (with AG) | Gender; social class; marital status; education; employment; duration of disorder; co-morbid psychiatric disorder; baseline level of panic; baseline level of agoraphobia; residual depression | Age; personality disorder; baseline level of depressed mood; antidepressant use at end of treatment; benzodiazepine use at end of treatment; residual agoraphobia |
| Heldt et al. (2011) | PD | Age; gender; marital status; education; age at onset; duration of disorder; baseline severity – agoraphobia; baseline severity – anticipatory anxiety; baseline severity – Hamilton anxiety; co-morbid mood disorder; co-morbid anxiety disorder; antidepressant use in follow-up; benzodiazepine use in follow-up; stressful life event (“loss”); stressful life event (“medical”); stressful life event (“others”) | Severity at one-year post-treatment – agoraphobia; severity at one-year post-treatment – anticipatory anxiety; severity at one-year post-treatment – Hamilton anxiety; stressful life event (“conflict”) |
| Lincoln et al. (2005) | SA | Age; gender; marital status; education; age at onset; prior outpatient treatment; prior inpatient treatment; prior medication use; severity of disorder; perceived impairment of disorder; baseline intensity of perceived distress; post-treatment intensity of perceived distress; number of physical symptoms; baseline physical arousal symptoms; post-treatment physical arousal symptoms; baseline bodily symptoms; post-treatment bodily symptoms; anxiety level across a range of feared situations (generalized subtype); comorbidity; baseline depression; post-treatment depression; baseline obsessive-compulsive ruminations; post-treatment obsessive-compulsive ruminations; baseline obsessive-compulsive behaviours; post-treatment obsessive-compulsive behaviours; baseline agoraphobic cognitions; post-treatment agoraphobic cognitions; baseline agoraphobic avoidance; post-treatment agoraphobic avoidance; baseline hypochondriasis; post-treatment hypochondriasis; chronic disease; baseline benzodiazepine use; baseline alcohol use; post-treatment alcohol use | Number of feared situations (generalized subtype) |
| Otto et al. (1996) | PD | Entering remission with medication; maintained on medication through follow-up; age; baseline panic attack frequency; baseline fear of anxiety sensations; baseline agoraphobic avoidance; co-morbid major depression; co-morbid anxiety disorder; history of childhood anxiety disorders; duration of adult disorder; previous remissions of greater than two months; previous remissions of greater than six months | Baseline disorder severity; agoraphobic subtype at baseline |

**APPENDIX D**

|  |  |
| --- | --- |
| **Table D1**  *Glossary for Predictors* | |
| Predictor | Definition |
| *Introduced in Model 1…* | |
| Age | The age of the patient. |
| Baseline GAD-7 | GAD-7 score at assessment. |
| Baseline PHQ-9 | PHQ-9 score at assessment. |
| Baseline WSAS | WSAS score at assessment. |
| Chronicity of Problem | The duration (months) of the mental health problem. |
| Diagnosis | Primary diagnosis recorded by clinician. |
| Disability | Patients having a disability or not. |
| Ethnicity | The ethnicity of the patient. |
| Expectancy of Treatment | Patient’s level of expectancy of treatment. |
| Family History | Patients having a family history of mental health problems or not. |
| Gender | The gender of the patient. |
| Long-Term Condition | Patients having a long-term physical condition or not. |
| Neighbourhood Deprivation | Level of deprivation of patient’s neighbourhood. |
| Previous Treatment Episodes | Patient’s number of previous treatment episodes. |
| Responded to Treatment Expectancy Question | Patients responded to expectancy question or not. |
| Taking Medication at Start | Patients taking medication at assessment or not. |
| Unemployment at Start | Patients unemployed at assessment or not. |
| *Introduced in Model 2…* | |
| Early Treatment Response GAD-7 | GAD-7 score at third treatment session minus score at first session. |
| Early Treatment Response PHQ-9 | PHQ-9 score at third treatment session minus score at first session. |
| Early Treatment Response WSAS | WSAS score at third treatment session minus score at first session. |
| GAD-7 at End | GAD-7 score at the end of treatment. |
| Linear Treatment Response (GoF) – GAD-7 | GoF of linear regression model applied to patient’s GAD-7 scores over treatment. |
| Linear Treatment Response (GoF) – PHQ-9 | GoF of linear regression model applied to patient’s PHQ-9 scores over treatment. |
| Linear Treatment Response (GoF) – WSAS | GoF of linear regression model applied to patient’s WSAS scores over treatment. |
| Linear Treatment Response (Slope) – GAD-7 | Slope of linear regression model applied to patient’s GAD-7 scores over treatment. |
| Linear Treatment Response (Slope) – PHQ-9 | Slope of linear regression model applied to patient’s PHQ-9 scores over treatment. |
| Linear Treatment Response (Slope) – WSAS | Slope of linear regression model applied to patient’s WSAS scores over treatment. |
| Number of Treatment Sessions | The number of sessions patients received. |
| PHQ-9 at End | PHQ-9 score at end of treatment. |
| Taking Medication at End | Patients taking medication at end of treatment or not. |
| Unemployment at End | Patients unemployment at end of treatment of not. |
| WSAS at End | WSAS score at end of treatment. |
| *Introduced in Model 3….* | |
| GAD-7 at FU1 | GAD-7 score at FU1. |
| GAD-7 Change from End to FU1 | The change in GAD-7 score from final treatment session to FU1. |
| PHQ-9 at FU1 | PHQ-9 score at FU1. |
| PHQ-9 Change from End to FU1 | The change in PHQ-9 score from final treatment session to FU1. |
| WSAS at FU1 | WSAS score at FU1. |
| WSAS Change from End to FU1 | The change in WSAS score from final treatment session to FU1. |
| Responded at FU1 | Patients responded to questionnaires at FU1. |
| *Introduced in Model 4…* | |
| GAD-7 at FU2 | GAD-7 score at FU2. |
| GAD-7 at FU3 | GAD-7 score at FU3. |
| GAD-7 Change from FU1 to FU2 | The change in GAD-7 score from FU1 to FU2. |
| GAD-7 Change from FU2 to FU3 | The change in GAD-7 score from FU2 to FU3. |
| PHQ-9 at FU2 | PHQ-9 score at FU2. |
| PHQ-9 at FU3 | PHQ-9 score at FU3. |
| PHQ-9 Change from FU1 to FU2 | The change in PHQ-9 score from FU1 to FU2. |
| PHQ-9 Change from FU2 to FU3 | The change in PHQ-9 score from FU2 to FU3. |
| WSAS at FU2 | WSAS score at FU2. |
| WSAS at FU3 | WSAS score at FU3. |
| WSAS Change from FU1 to FU2 | The change in WSAS score from FU1 to FU2. |
| WSAS Change from FU2 to FU3 | The change in WSAS score from FU2 to FU3. |
| Responded at FU2 | Patients responded to questionnaires at FU2. |
| Responded at FU3 | Patients responded to questionnaires at FU3. |
| Abbreviations: FU1, first month of follow-up; FU2, second month of follow-up; FU3, third month of follow-up; GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit (R2) for a regression model; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale. | |

**APPENDIX E**

The partial dependence plots below are organized with regards to the four developed models, and displayed in order of relative importance. The x-axis of each plot represents the values of a specific predictor, while the y-axis represents probabilities of relapse (log odds ratio). Values above zero on the y-axis indicate an increased risk for relapse according to the respective developed model, while values below zero indicate a higher probability for remaining in-remission.

**Model 1**

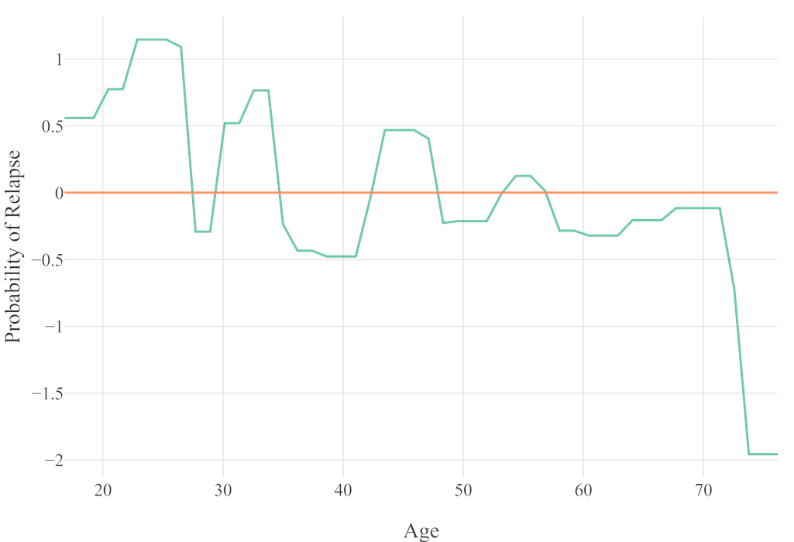


Figure E1. Partial dependence plot for the relationship between age and relapse in Model 1.



Figure E2. Partial dependence plot for the relationship between baseline WSAS and relapse in Model 1.

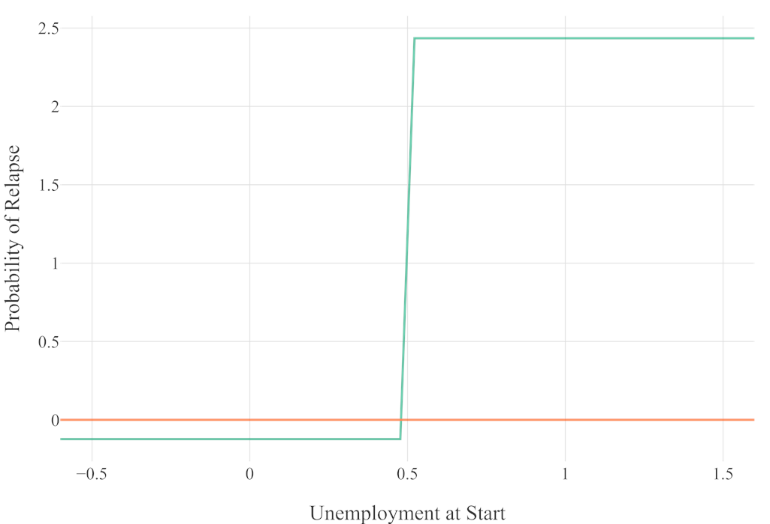


Figure E3. Partial dependence plot for the relationship between unemployment at start and relapse in Model 1. X-axis values: ‘0’=*employment*, ‘1’=*unemployment*.

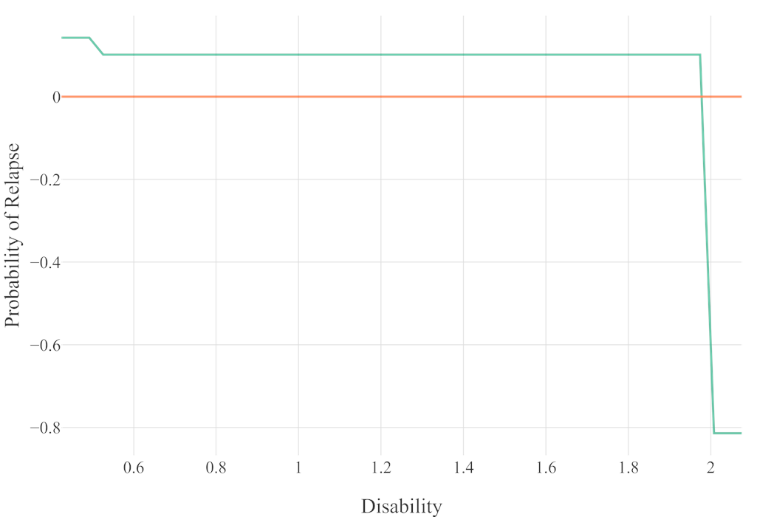


Figure E4. Partial dependence plot for the relationship between disability and relapse in Model 1. X-axis values: ‘0’=*no recorded disability*, ‘1’=*recorded disability,* ‘2’*=missing response*.

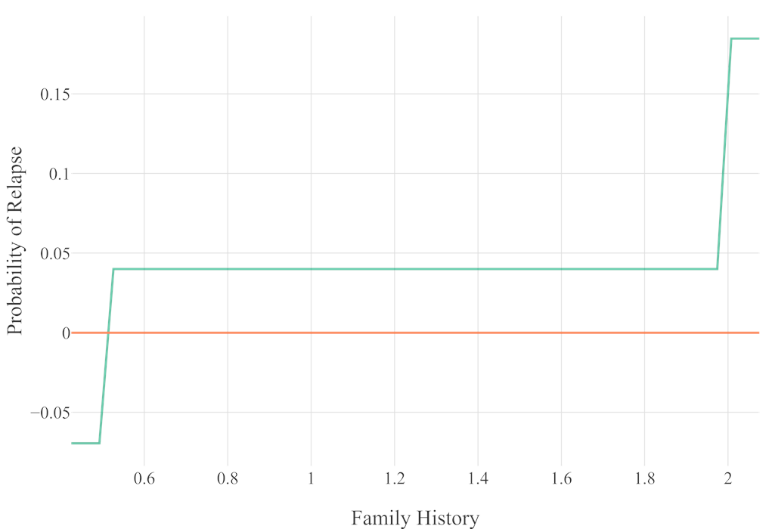


Figure E5. Partial dependence plot for the relationship between family history and relapse in Model 1. X-axis values: ‘0’=*no recorded family history*, ‘1’=*recorded family history,* ‘2’*=missing response*.

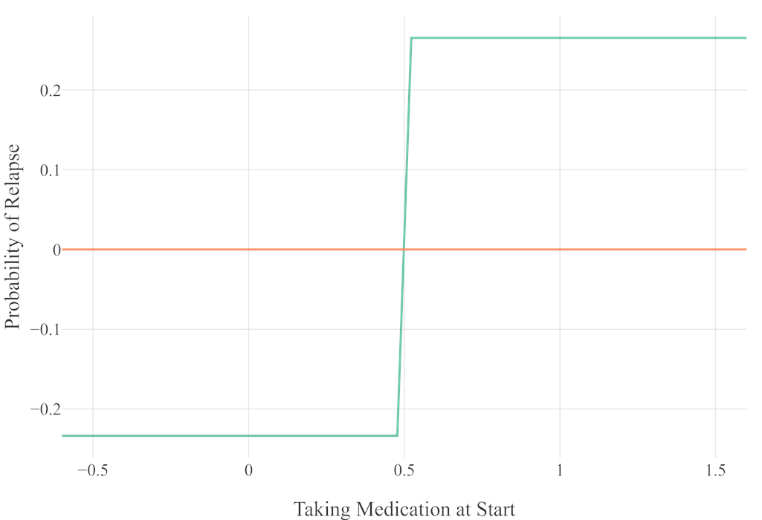


Figure E6. Partial dependence plot for the relationship between taking medication at start and relapse in Model 1. X-axis values: ‘0’=*not taking medication*, ‘1’=*taking medication.*

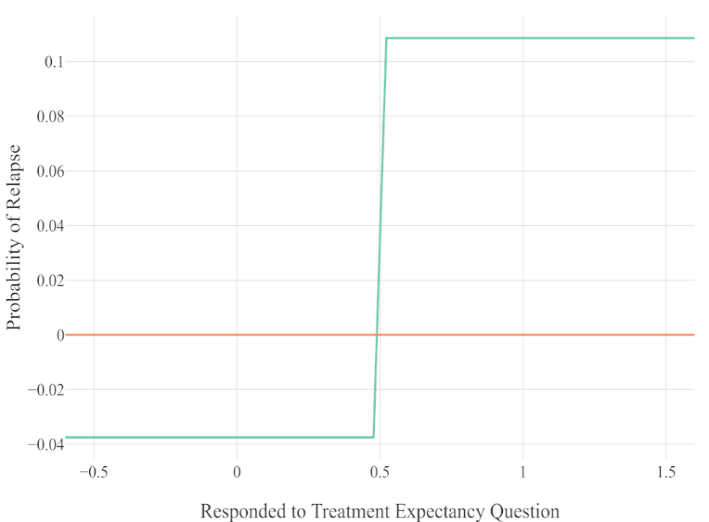


Figure E7. Partial dependence plot for the relationship between responding to treatment expectancy question and relapse in Model 1. X-axis values: ‘0’=*no response*, ‘1’=*response*

**Model 2**

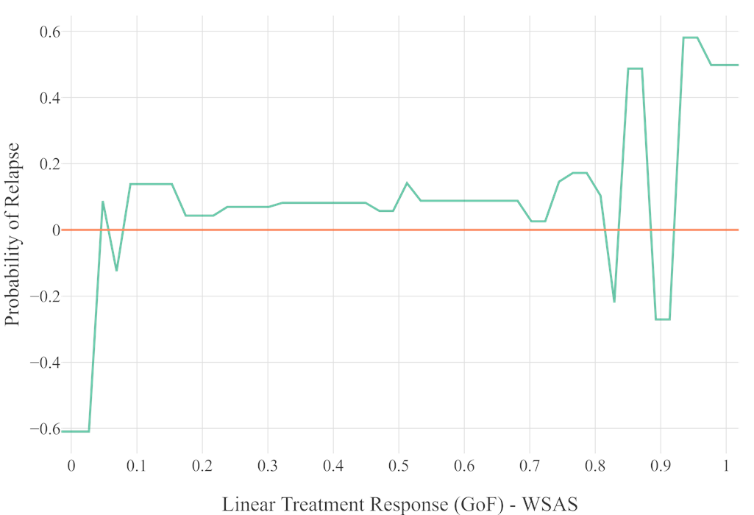


Figure E8. Partial dependence plot for the relationship between linear treatment response (GoF) in the WSAS and relapse in Model 2.

Figure E9. Partial dependence plot for the relationship between age and relapse in Model 2.

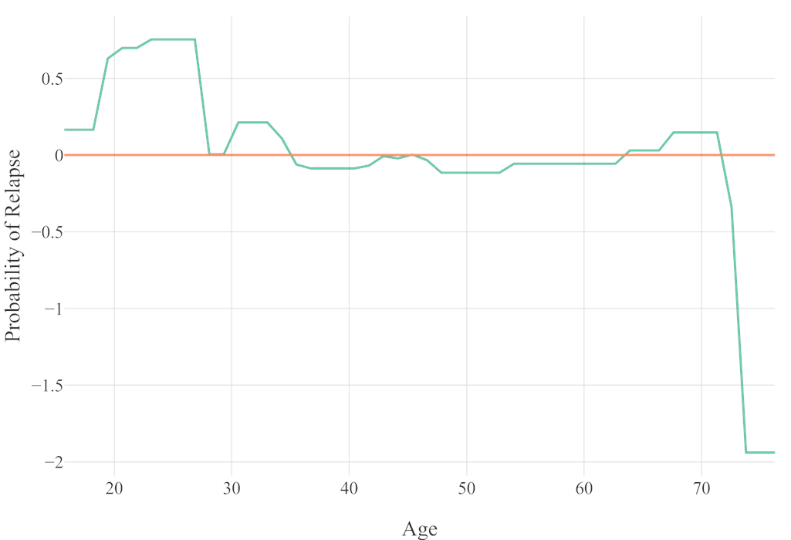
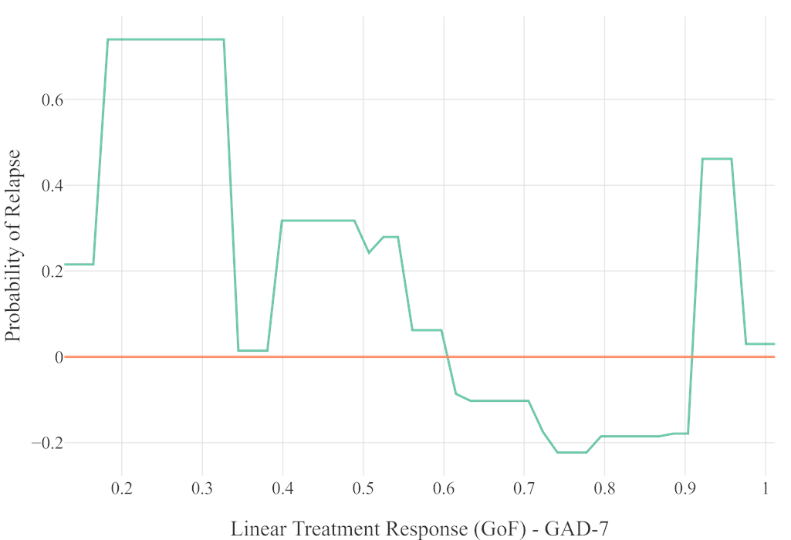


Figure E10. Partial dependence plot for the relationship between linear treatment response (GoF) in GAD-7 and relapse in Model 2.



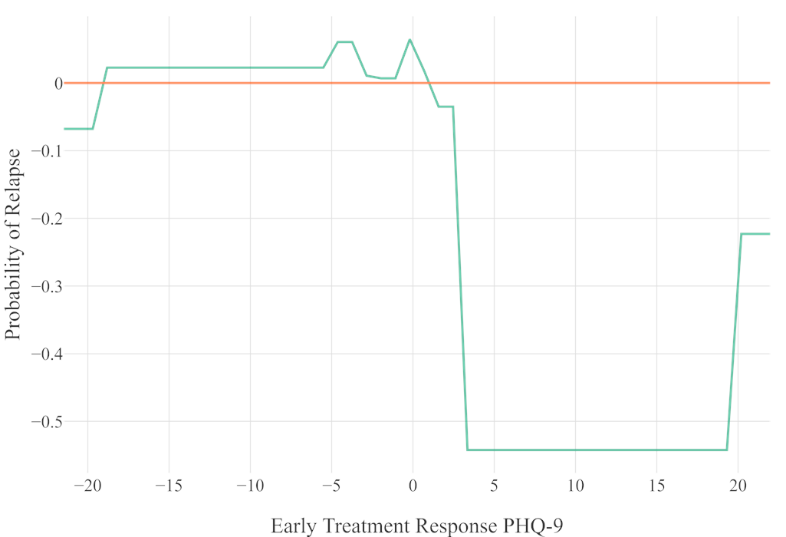


Figure E11. Partial dependence plot for the relationship between early treatment response in PHQ-9 and relapse in Model 2.

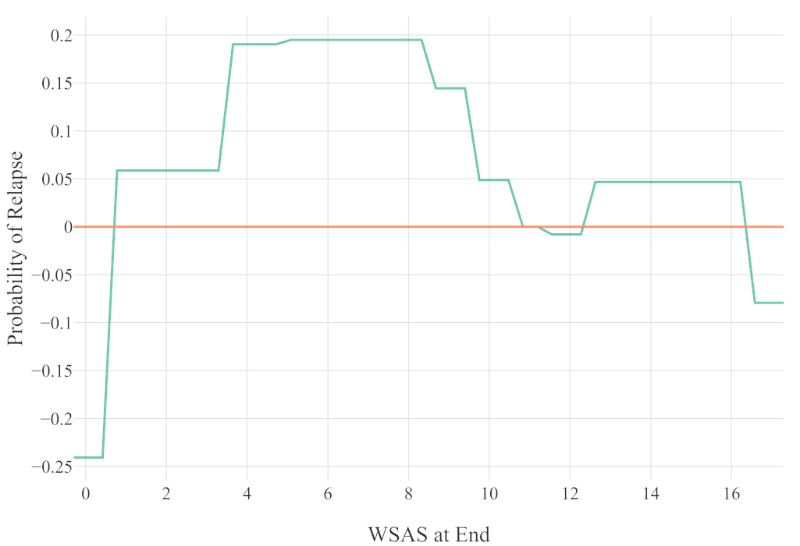


Figure E12. Partial dependence plot for the relationship between WSAS at the end of treatment and relapse in Model 2.

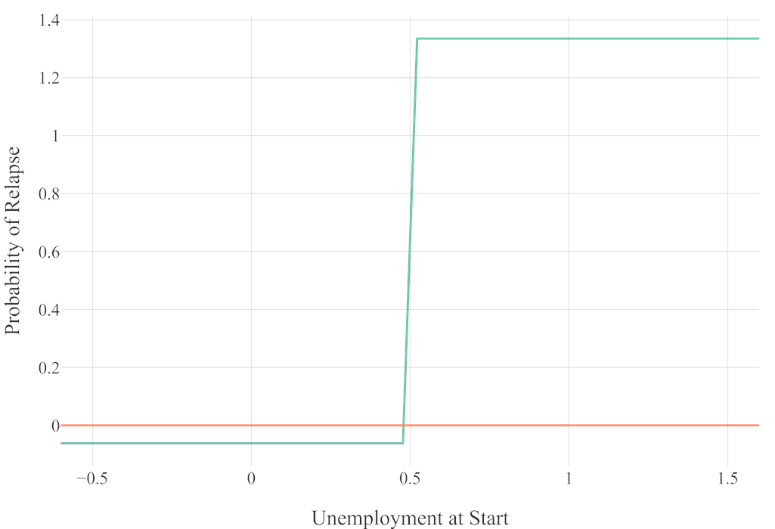


Figure E13. Partial dependence plot for the relationship between unemployment at start and relapse in Model 2. X-axis values: ‘0’=*employment*, ‘1’=*unemployment*.

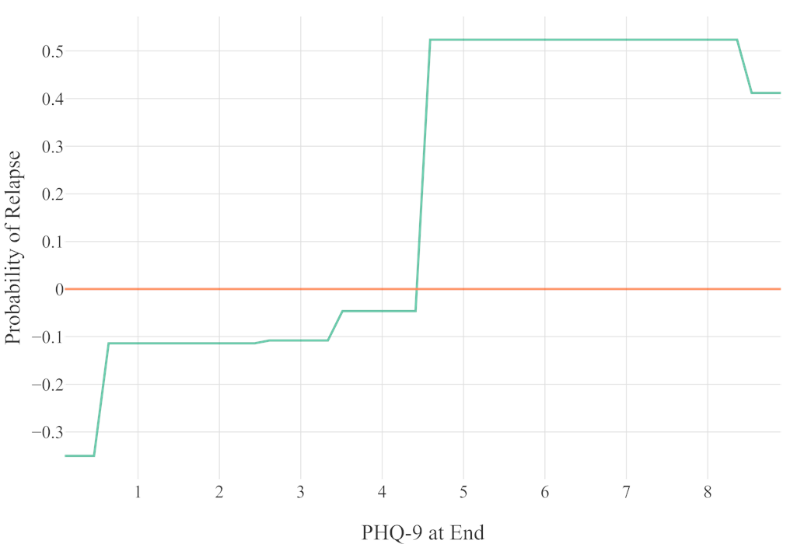


Figure E14. Partial dependence plot for the relationship between PHQ-9 at the end of treatment and relapse in Model 2.

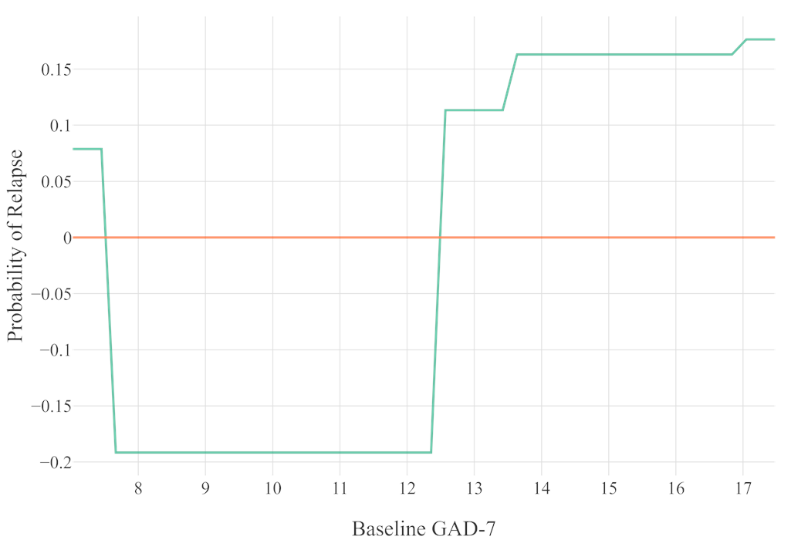


Figure E15. Partial dependence plot for the relationship between baseline GAD-7 and relapse in Model 2.

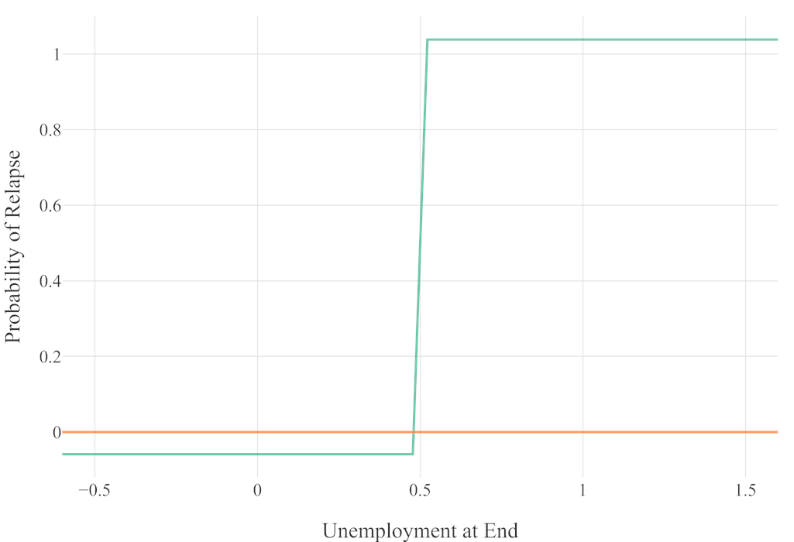


Figure E16. Partial dependence plot for the relationship between unemployment at end and relapse in Model 2. X-axis values: ‘0’=*employment*, ‘1’=*unemployment*.

Figure E17. Partial dependence plot for the relationship between previous treatment episodes and relapse in Model 2.

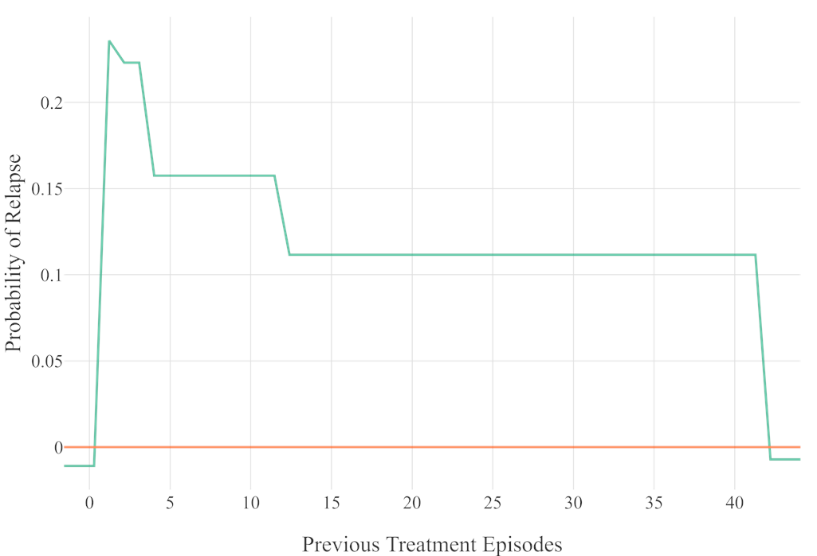


Figure E18. Partial dependence plot for the relationship between taking medication at the start and relapse in Model 2. X-axis values: ‘0’=*not taking medication*, ‘1’=*taking medication*.

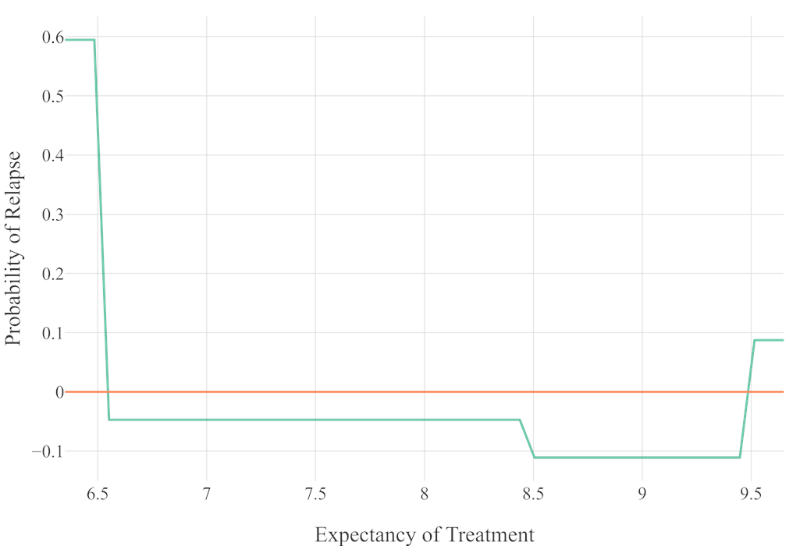
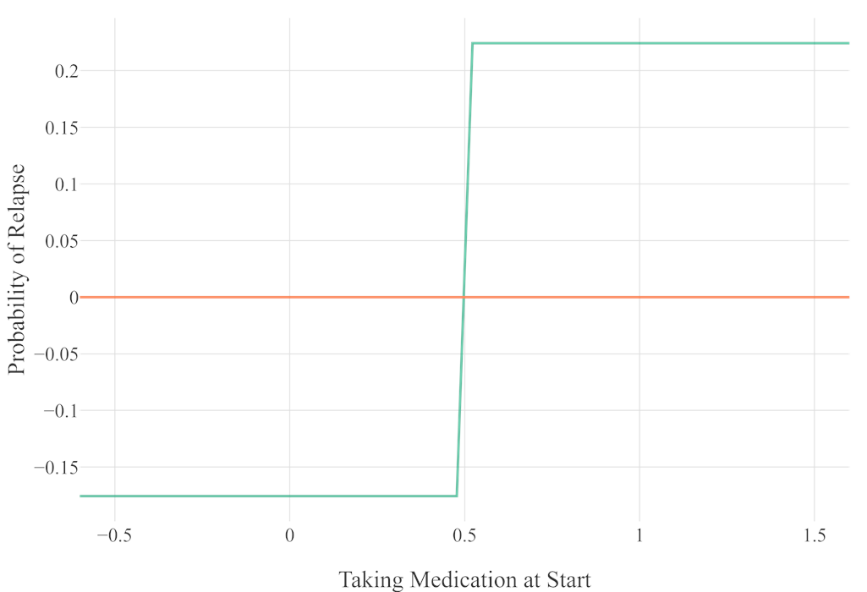


Figure E19. Partial dependence plot for the relationship between expectancy of treatment and relapse in Model 2.

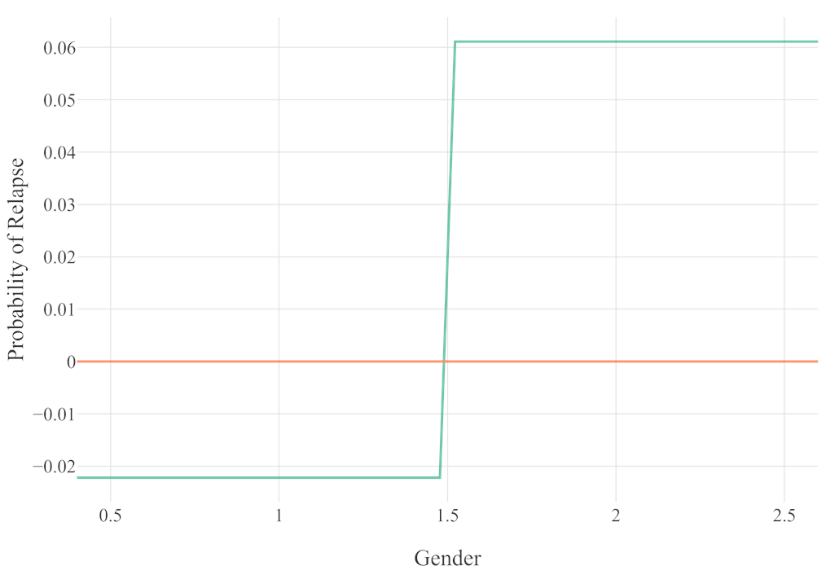


Figure E20. Partial dependence plot for the relationship between gender and relapse in Model 2. X-axis values: ‘1’=*male*, ‘2’=*female*.

**Model 3**

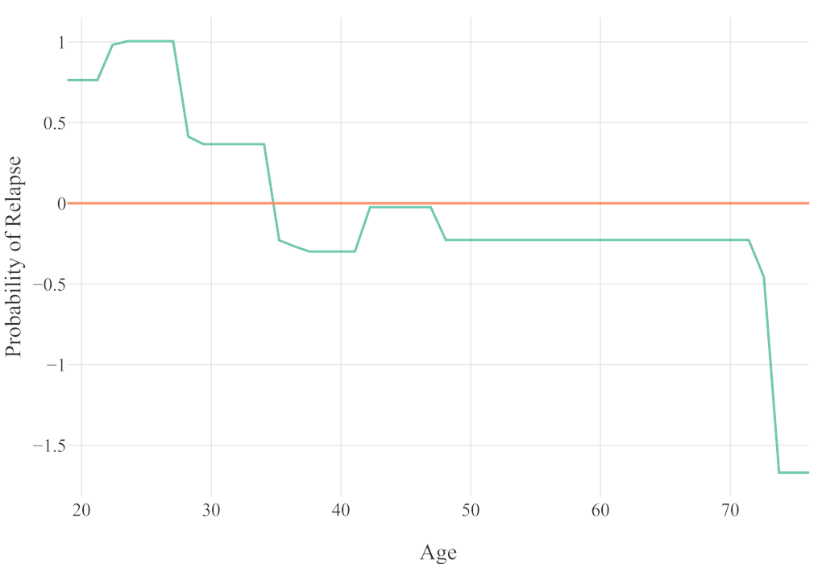


Figure E21. Partial dependence plot for the relationship between age and relapse in Model 3.

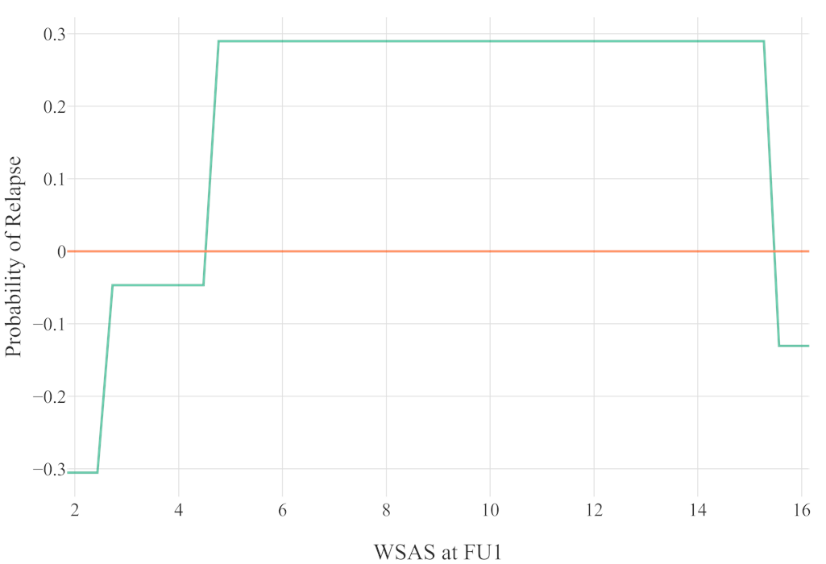


Figure E22. Partial dependence plot for the relationship between WSAS at FU1 and relapse in Model 3.

Figure E23. Partial dependence plot for the relationship between GAD-7 at FU1 and relapse in Model 3.

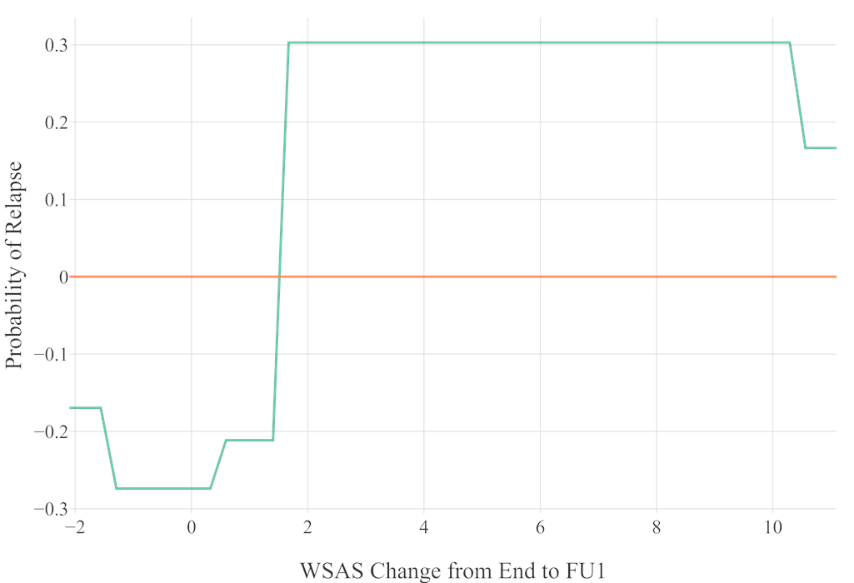
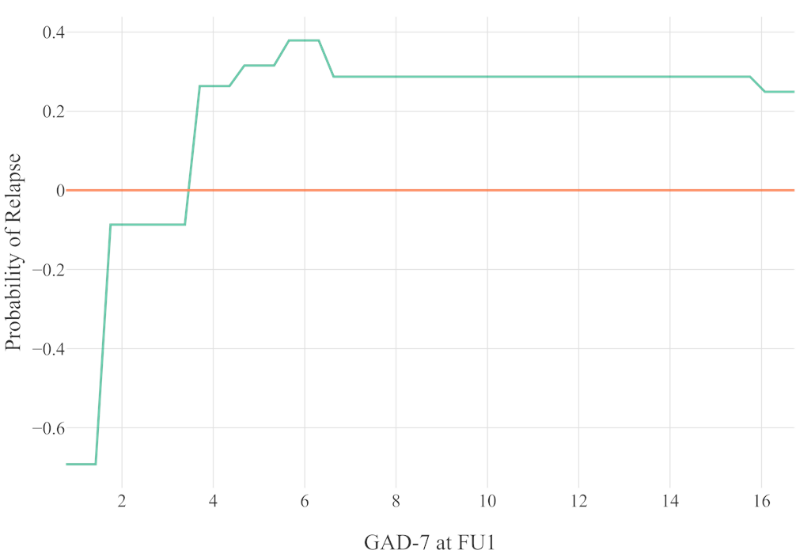


Figure E24. Partial dependence plot for the relationship between change in WSAS from end of treatment to FU1 and relapse in Model 3.

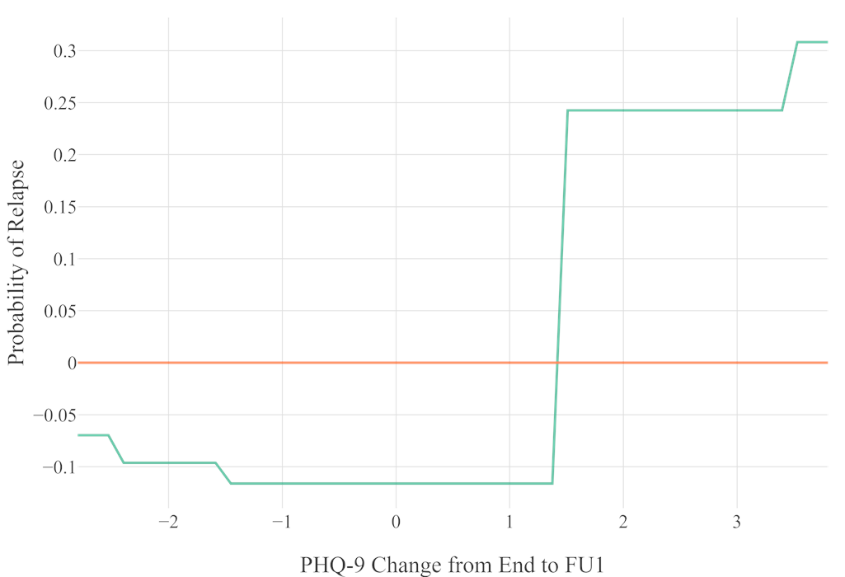


Figure E25. Partial dependence plot for the relationship between change in PHQ-9 from end of treatment to FU1 and relapse in Model 3.

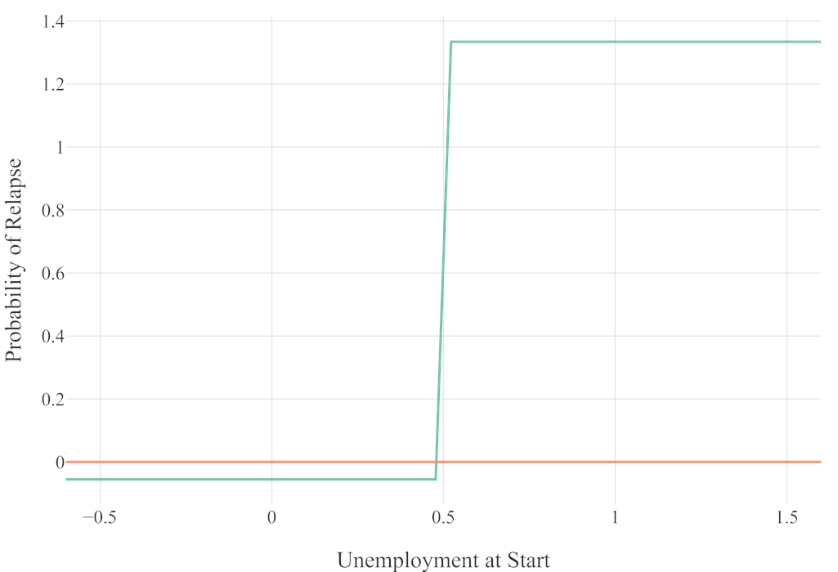


Figure E26. Partial dependence plot for the relationship between unemployed at start and relapse in Model 3. X-axis values: ‘0’=*employment*, ‘1’=*unemployment*.

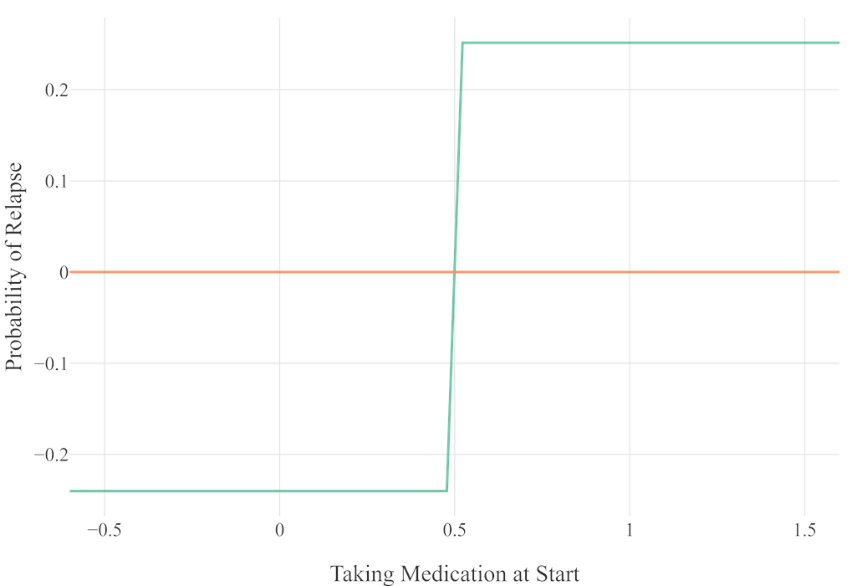


Figure E27. Partial dependence plot for the relationship between taking medication at start and relapse in Model 3. X-axis values: ‘0’=*not taking medication*; ‘1’=*taking medication*.

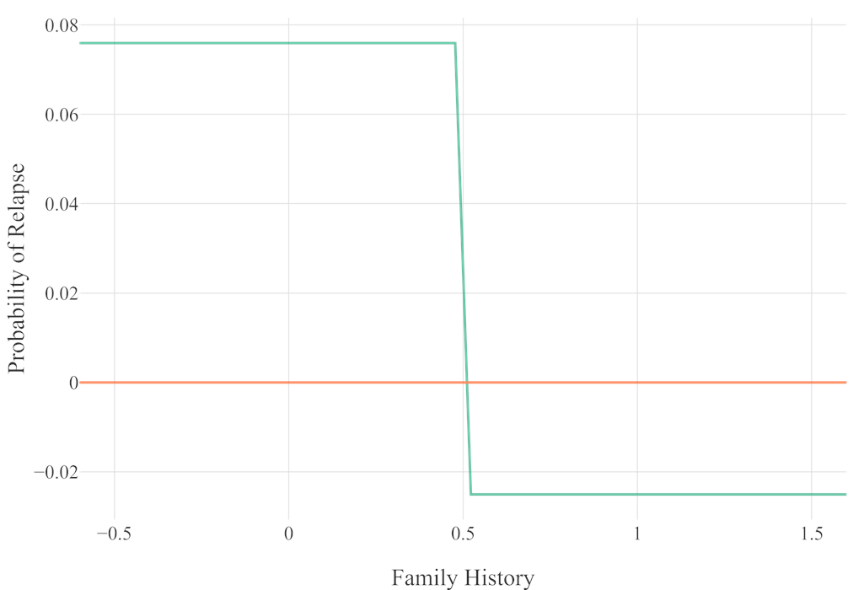


Figure E28. Partial dependence plot for the relationship between family history and relapse in Model 3. X-axis values: ‘0’=*no recorded family history*; ‘1’=*recorded family history*.

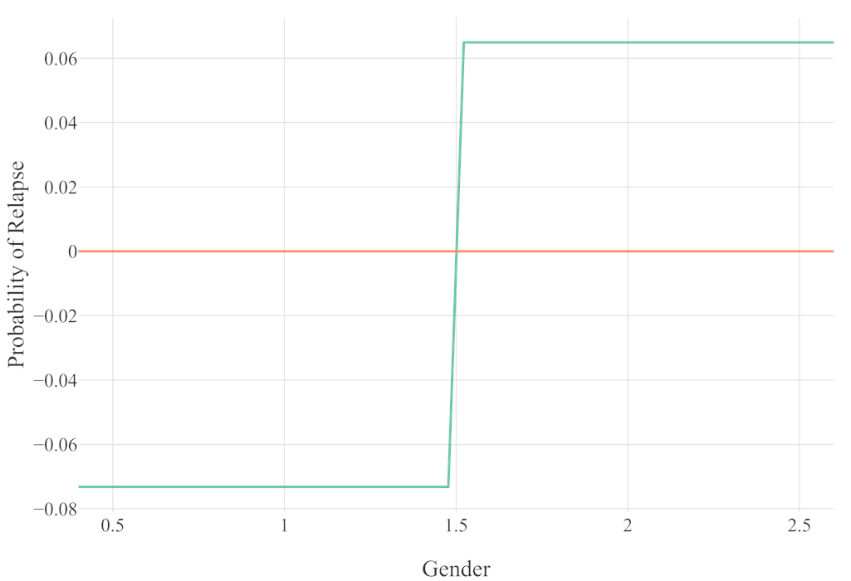


Figure E29. Partial dependence plot for the relationship between gender and relapse in Model 3. X-axis values: ‘1’=*male*, ‘2’=*female*.

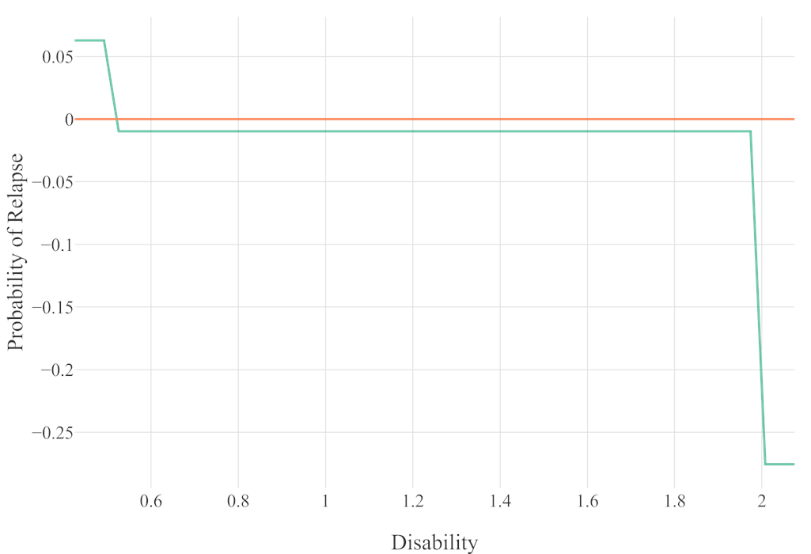
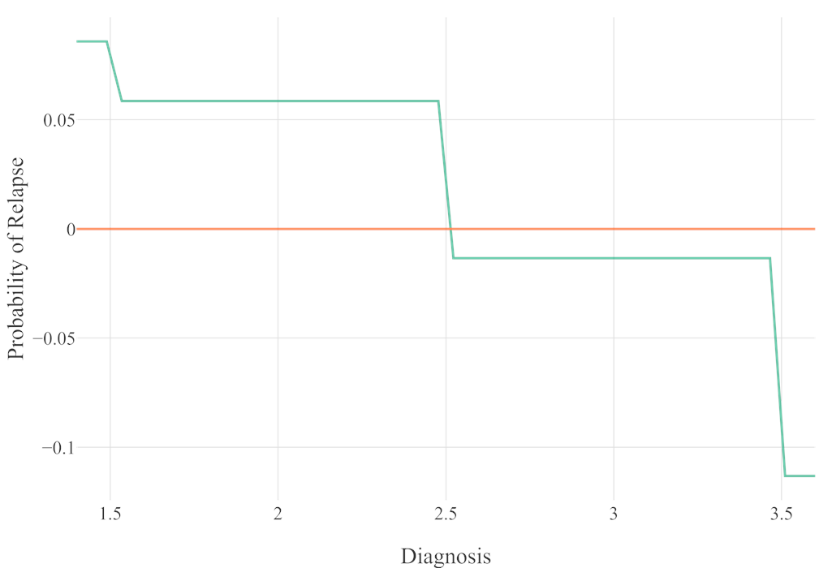


Figure E30. Partial dependence plot for the relationship between disability and relapse in Model 3. X-axis values: ‘0’=*no recorded disability*, ‘1’=*recorded disability,* ‘2’*=missing response*.

Figure E31. Partial dependence plot for the relationship between diagnosis and relapse in Model 3. X-axis values: ‘1’=*affective disorder*, ‘2’=*anxiety disorder,* ‘3’*=mixed anxiety and depression, ’4’=other (including PTSD and OCD)*



**Model 4**

Figure E32. Partial dependence plot for the relationship between GAD-7 at FU3 and relapse in Model 4.

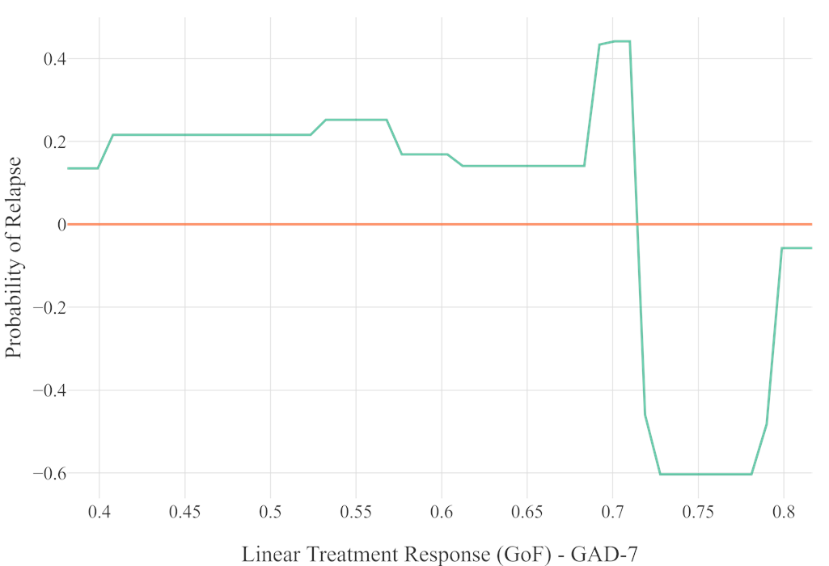
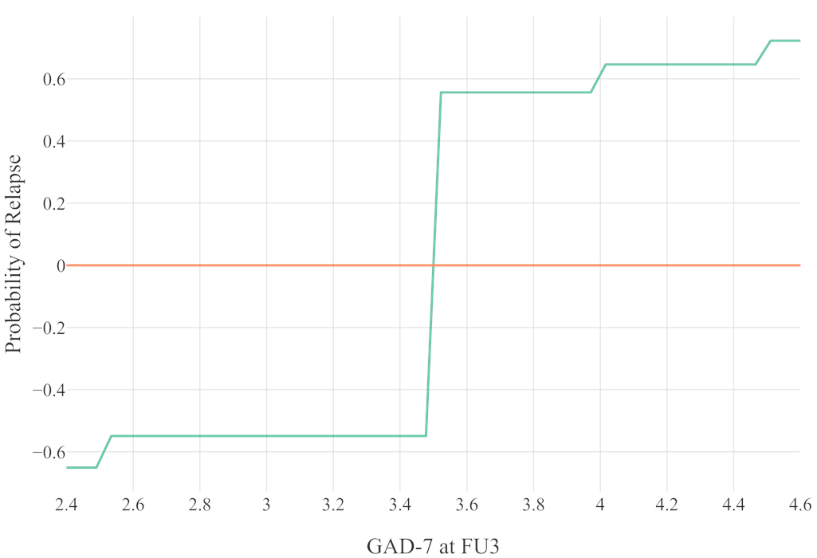


Figure E33. Partial dependence plot for the relationship between linear treatment response (GoF) in GAD-7 and relapse in Model 4.

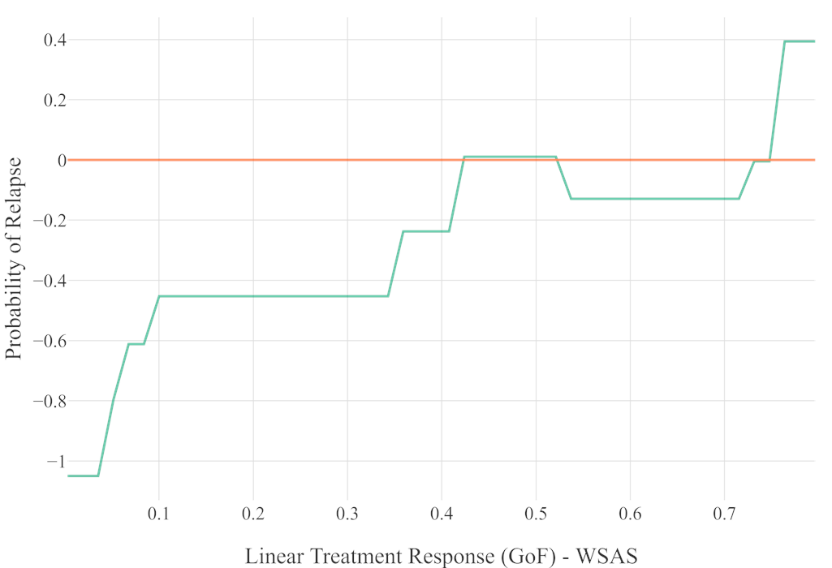
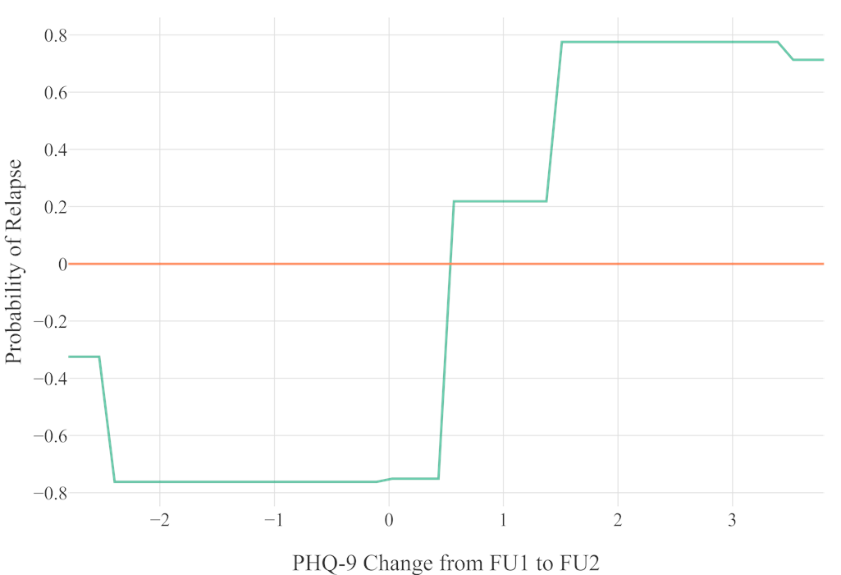


Figure E34. Partial dependence plot for the relationship between linear treatment response (GoF) in WSAS and relapse in Model 4.

Figure E35. Partial dependence plot for the relationship between change in PHQ-9 from FU1 to FU2 and relapse in Model 4.



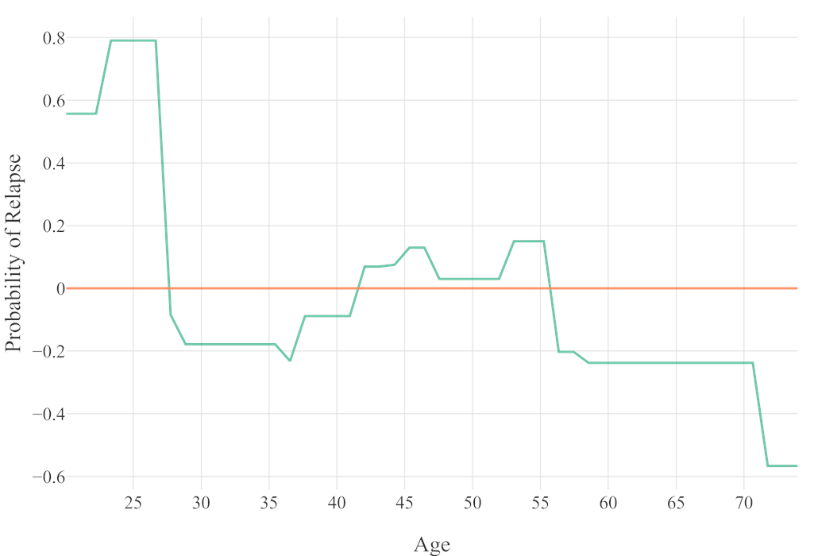


Figure E36. Partial dependence plot for the relationship between age and relapse in Model 4.

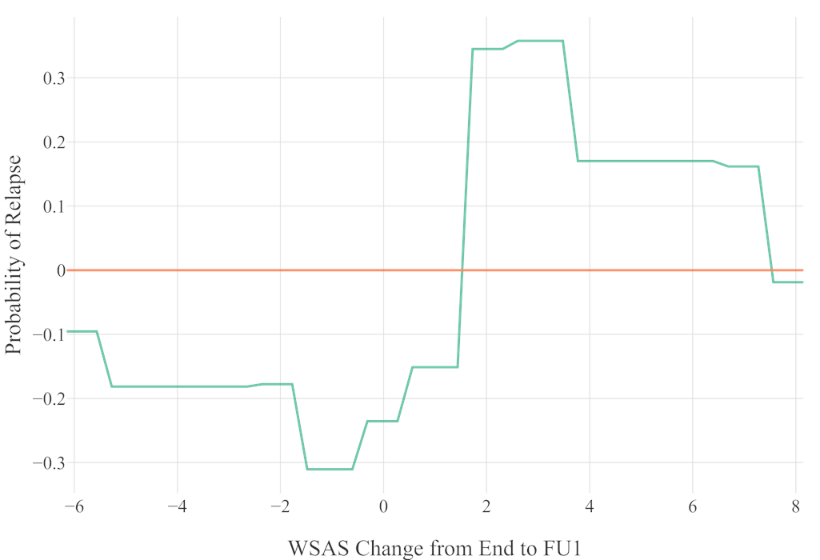


Figure E37. Partial dependence plot for the relationship between change in WSAS from end of treatment to FU1 and relapse in Model 4.

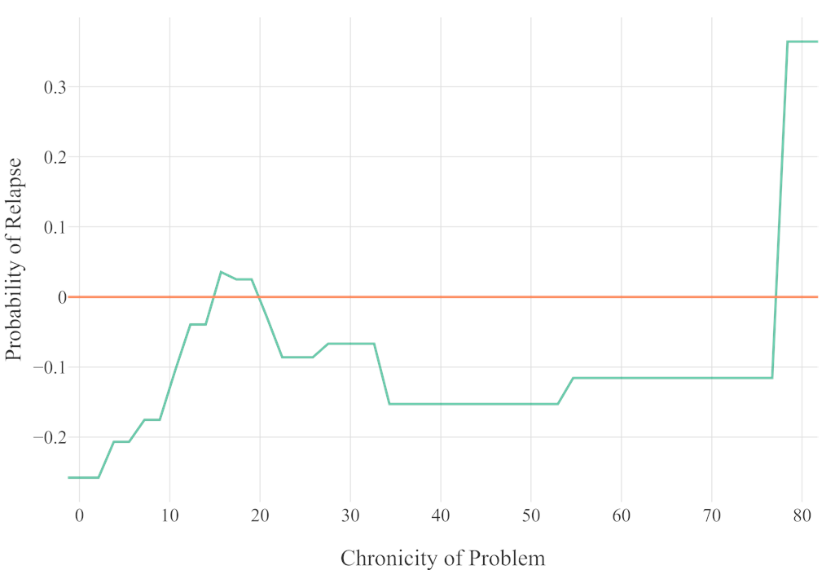


Figure E38. Partial dependence plot for the relationship between chronicity of problem and relapse in Model 4.

Figure E39. Partial dependence plot for the relationship between change in PHQ-9 from end of treatment to FU1 and relapse in Model 4.

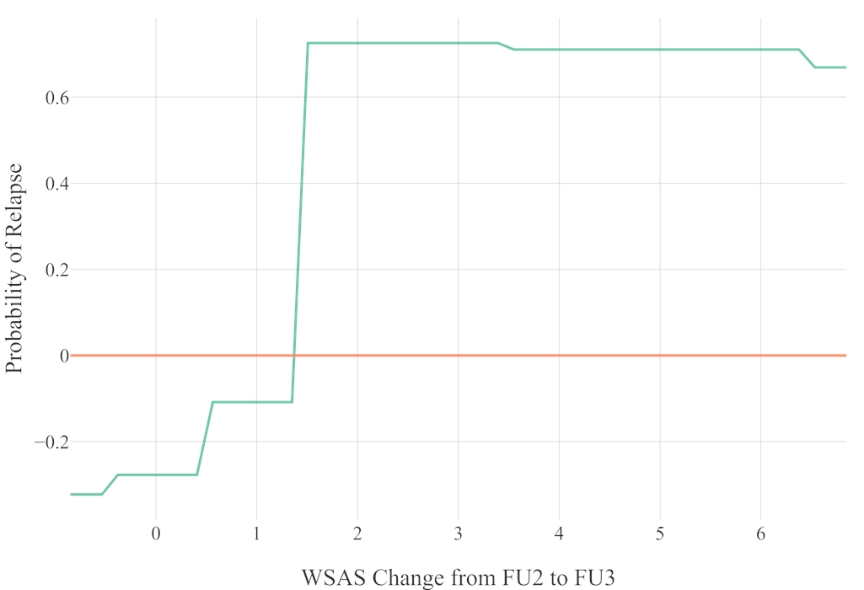


Figure E40. Partial dependence plot for the relationship between change in WSAS from FU2 to FU3 and relapse in Model 4.

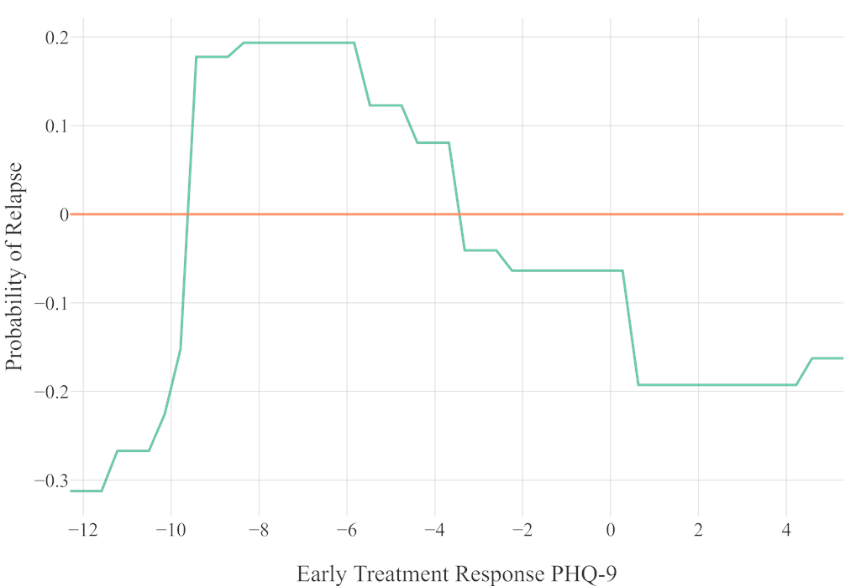
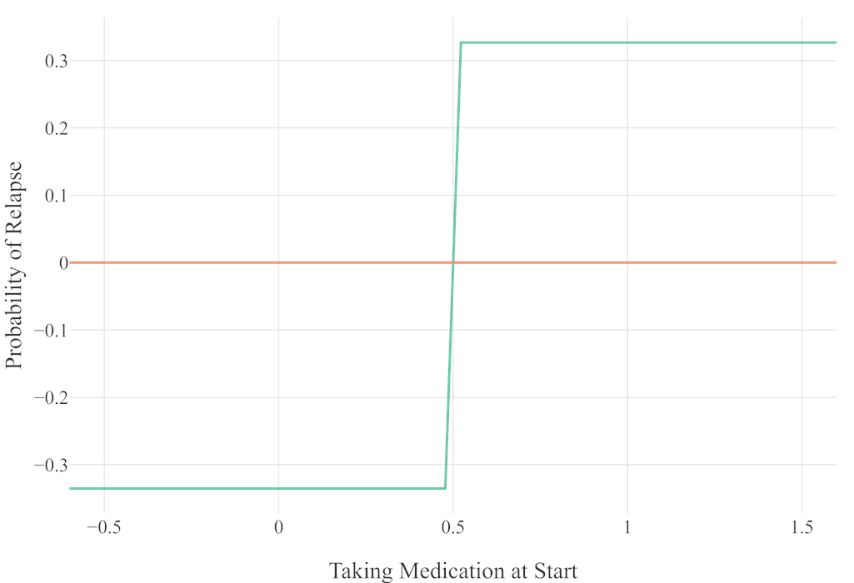


Figure E41. Partial dependence plot for the relationship between early treatment response in PHQ-9 and relapse in Model 4.

Figure E42. Partial dependence plot for the relationship between taking medication at start and relapse in Model 4. X-axis values: ‘0’=*not taking medication*, ‘1’=*taking medication*.



**APPENDIX F**

|  |  |  |
| --- | --- | --- |
| **Table F1**  *Glossary for Predictors* | | |
| Predictor | Definition | |
| Age | The age of the patient. | |
| Any Previous Treatment Within Episode | Patients receiving previous low- or high-intensity treatment in episode of care or not. | |
| Baseline GAD-7 | GAD-7 score at assessment. | |
| Baseline PHQ-9 | PHQ-9 score at assessment. | |
| Baseline WSAS | WSAS score at assessment. | |
| Computed Noise Variable (Categorical) | Noise variable computed based on base event rate of unemployment at start. | |
| Computed Noise Variable (Continuous) | Noise variable computed based on mean and standard deviation of PHQ-9 at end. | |
| Diagnosis (Affective Disorder) | Primary diagnosis of an affective disorder, recorded by clinician. | |
| Diagnosis (Mixed Anxiety/Depression) | Primary diagnosis of mixed anxiety and depression, recorded by clinician. | |
| Diagnosis (Anxiety Disorder) | Primary diagnosis of an anxiety disorder, recorded by clinician. | |
| Diagnosis (Other) | Primary diagnosis of a disorder other than those above, recorded by clinician. | |
| Early Treatment Response GAD-7 | GAD-7 score at third treatment session minus score at first session. | |
| Early Treatment Response PHQ-9 | PHQ-9 score at third treatment session minus score at first session. | |
| Early Treatment Response WSAS | WSAS score at third treatment session minus score at first session. | |
| Ethnicity | The ethnicity of the patient (White British vs other). | |
| GAD-7 at End | GAD-7 score at the end of treatment. | |
| Gender | The gender of the patient (male vs female). | |
| Linear Treatment Response (GoF) – GAD-7 | GoF of linear regression model applied to patient’s GAD-7 scores over treatment. | |
| Linear Treatment Response (GoF) – PHQ-9 | GoF of linear regression model applied to patient’s PHQ-9 scores over treatment. | |
| Linear Treatment Response (GoF) – WSAS | GoF of linear regression model applied to patient’s WSAS scores over treatment. | |
| Neighbourhood Deprivation | Level of deprivation of patient’s neighbourhood. |
| Number of Treatment Sessions Attended | The number of sessions patients received. |
| PHQ-9 at End | PHQ-9 score at end of treatment. |
| Taking Medication at Start | Patients taking medication at assessment or not. |
| Taking Medication at End | Patients taking medication at end of treatment or not. |
| Unemployment at Start | Patients unemployed at assessment or not. |
| Unemployment at End | Patients unemployment at end of treatment of not. |
| WSAS at End | WSAS score at end of treatment. |
| Abbreviations: GAD-7, Generalized Anxiety Disorder scale; GoF, goodness of fit (R2) for a regression model; PHQ-9, Patient Health Questionnaire; WSAS, Work and Social Adjustment Scale. | |

**APPENDIX G**

The partial dependence plots below are organized with regards to the four developed models, and displayed in order of relative importance. The x-axis of each plot represents the values of a specific predictor, while the y-axis represents probabilities of relapse (log odds ratio). Values above zero on the y-axis indicate an increased risk for relapse according to the respective developed model, while values below zero indicate a higher probability for remaining in-remission.

**Primary Model**

Figure G2. Partial dependence plot for the relationship between continuous noise variable and relapse in Primary Model.

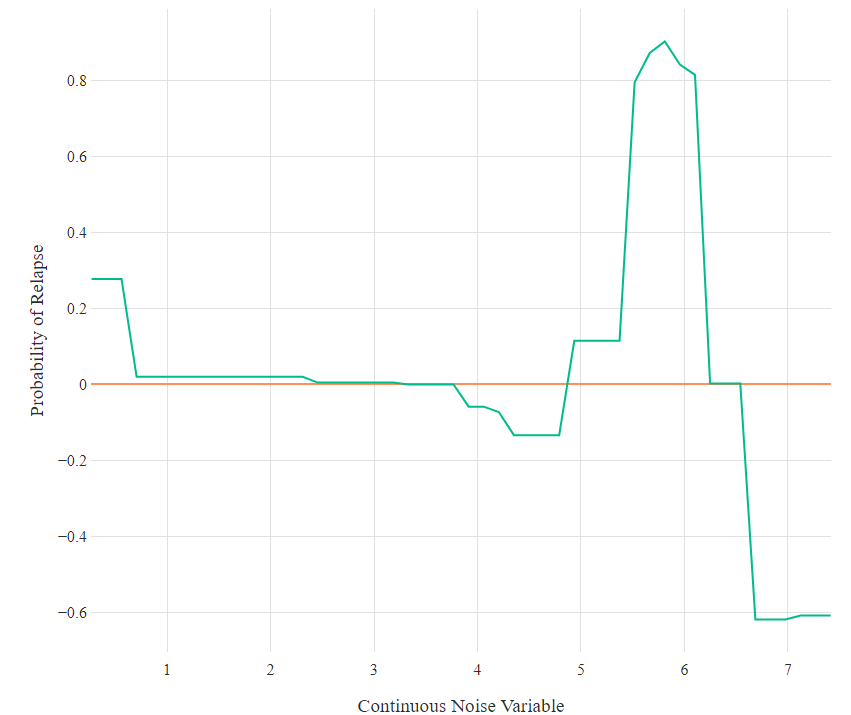


Figure G1. Partial dependence plot for the relationship between WSAS at the end of treatment and relapse in Primary Model.

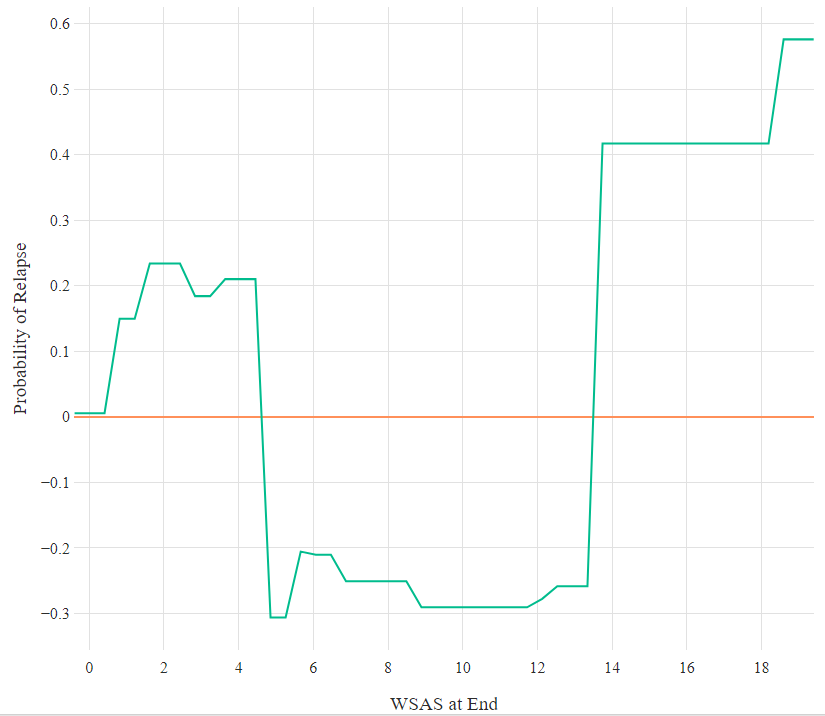


Figure G4. Partial dependence plot for the relationship between WSAS at the end of treatment and relapse in Primary Model.

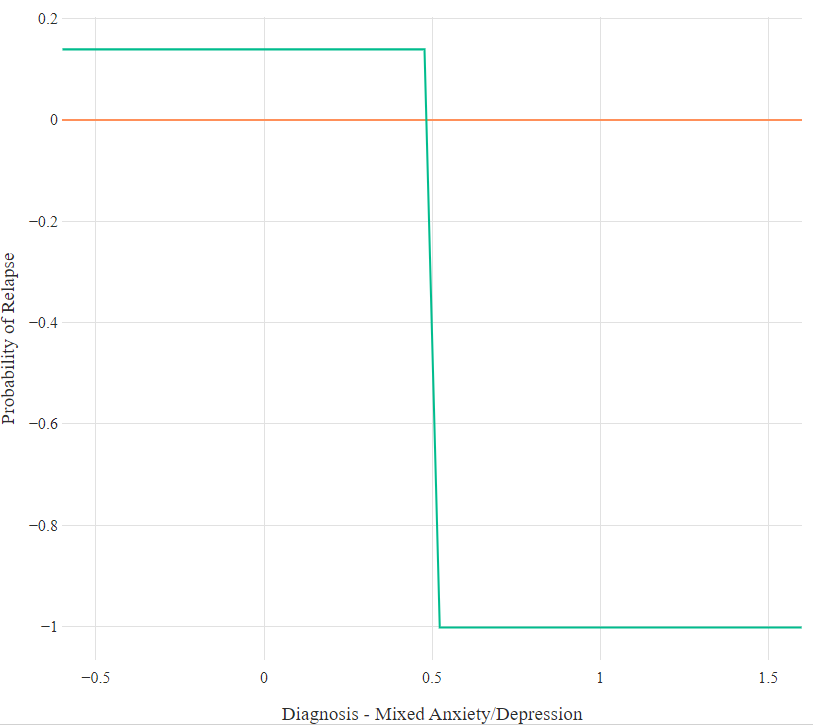


Figure G3. Partial dependence plot for the relationship between PHQ-9 at the end of treatment and relapse in Primary Model.

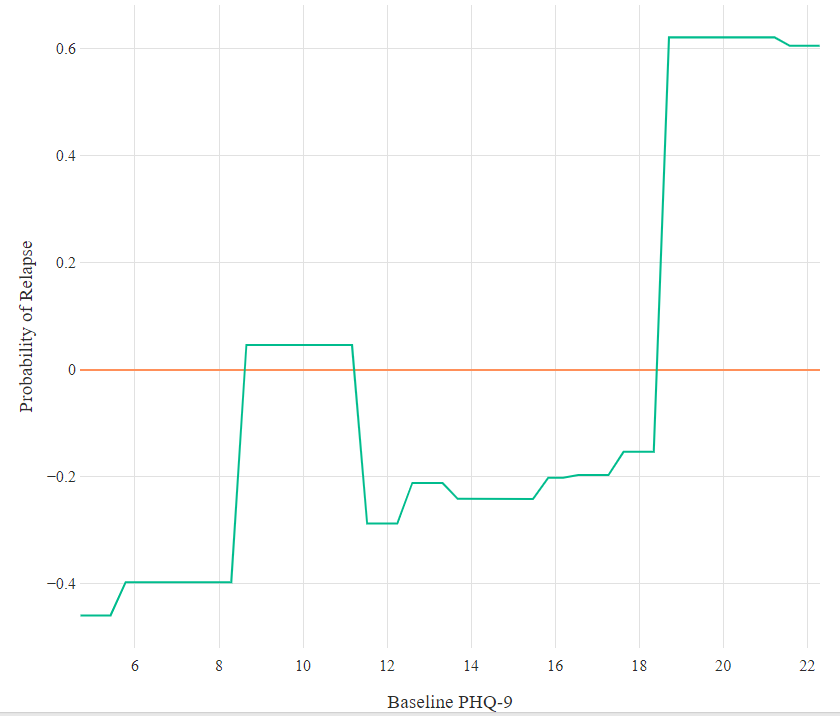


Figure G5. Partial dependence plot for the relationship between GAD-7 at the end of treatment and relapse in Primary Model.

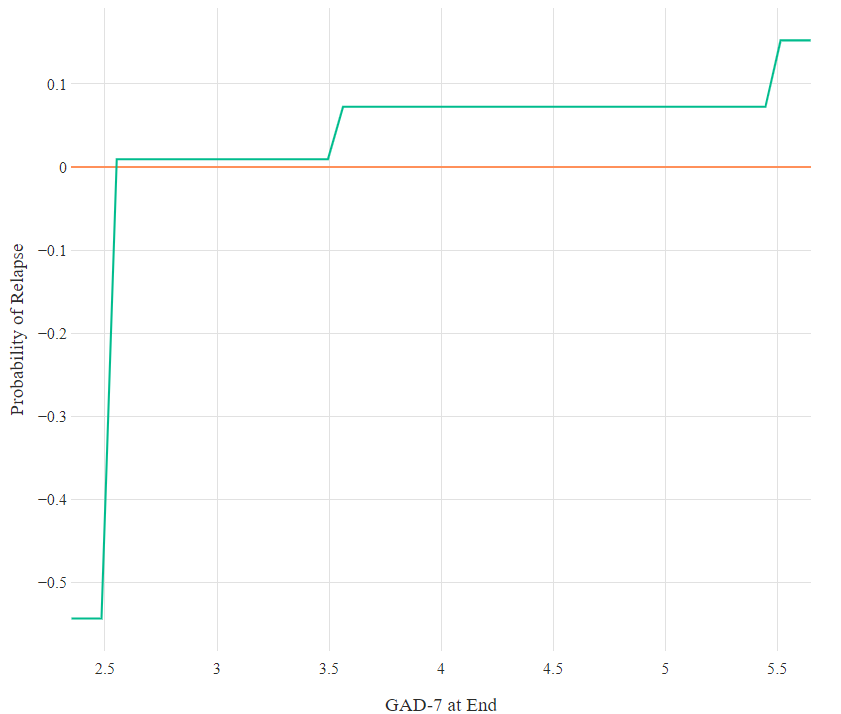


Figure G6. Partial dependence plot for the relationship between treatment modality and relapse in Primary Model. X-axis values: *‘0’=non-CBT, ‘1’=CBT.*

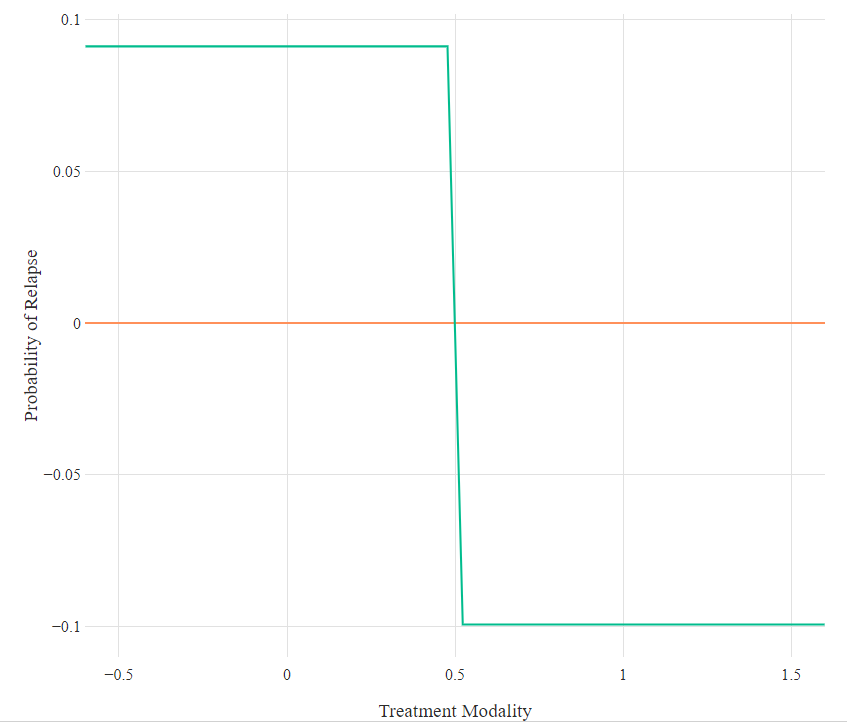


Figure G7. Partial dependence plot for the relationship between categorical noise variable and relapse in Primary Model.

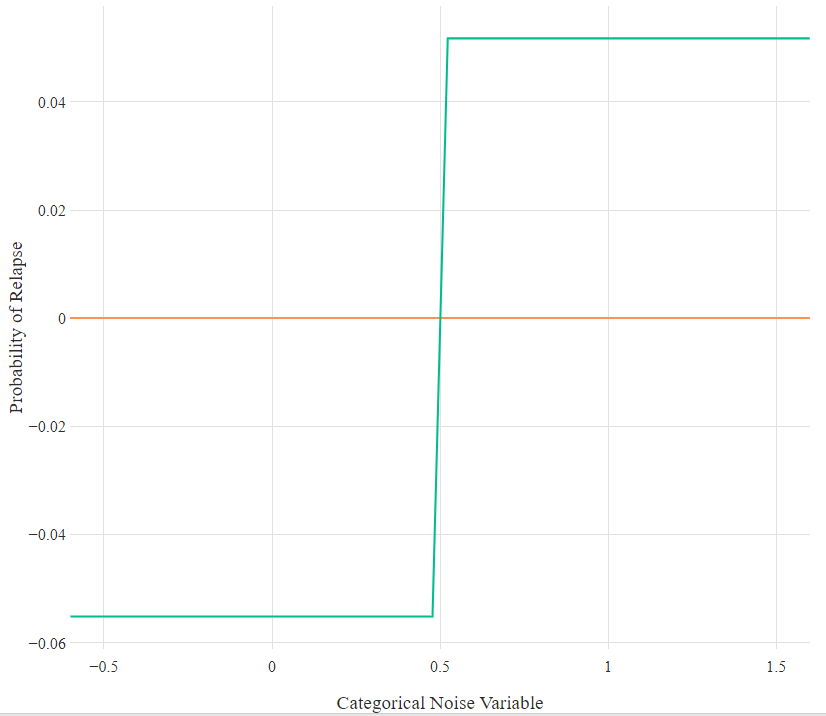


Figure G8. Partial dependence plot for the relationship between neighbourhood deprivation and relapse in Primary Model.

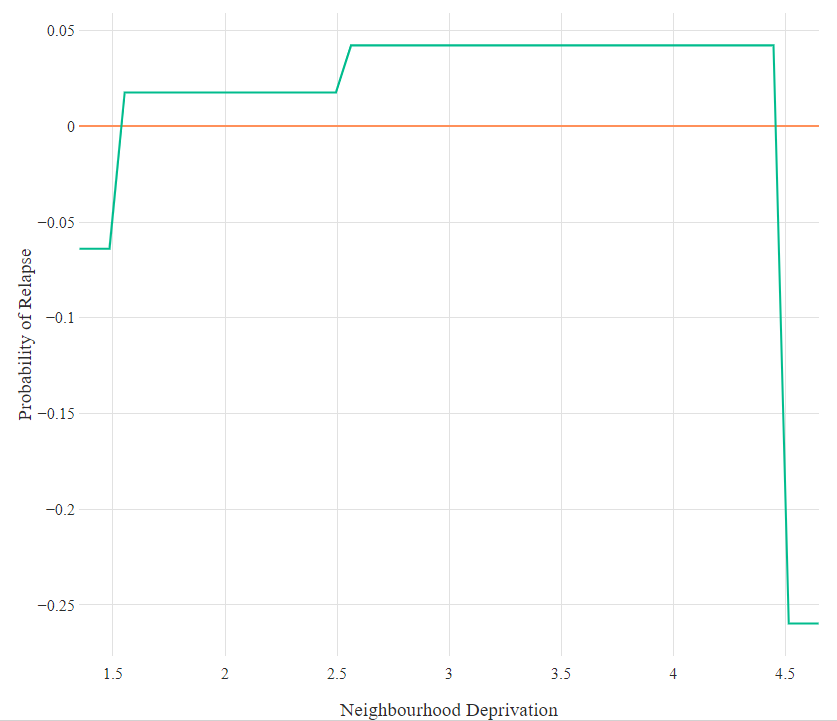


Figure G9. Partial dependence plot for the relationship between unemployment at start and relapse in Primary Model. X axis values: *‘0’=employment, ‘1’=unemployment.*

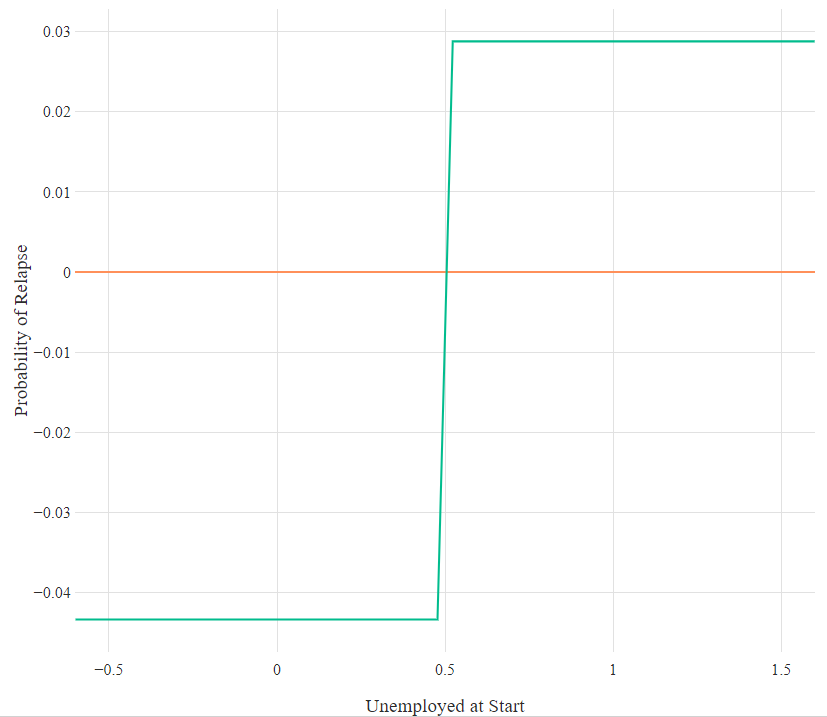
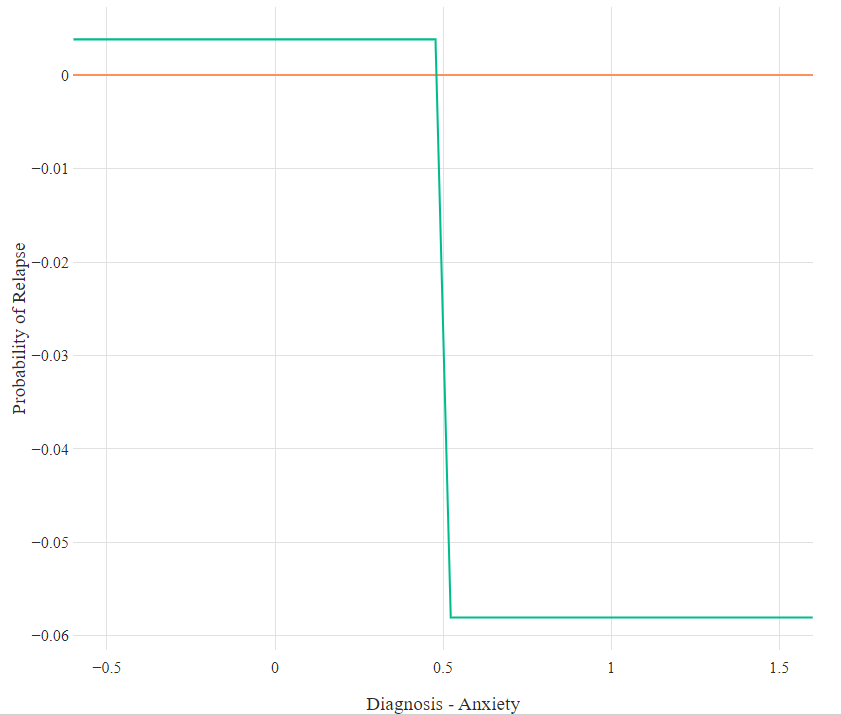


Figure G10. Partial dependence plot for the relationship between primary problem of anxiety and relapse in Primary Model. X-axis values: *“0”=not anxiety”, “1”=anxiety.*



**Secondary Model**

Figure G11. Partial dependence plot for the relationship between neighbourhood deprivaiton and relapse in Secondary Model.

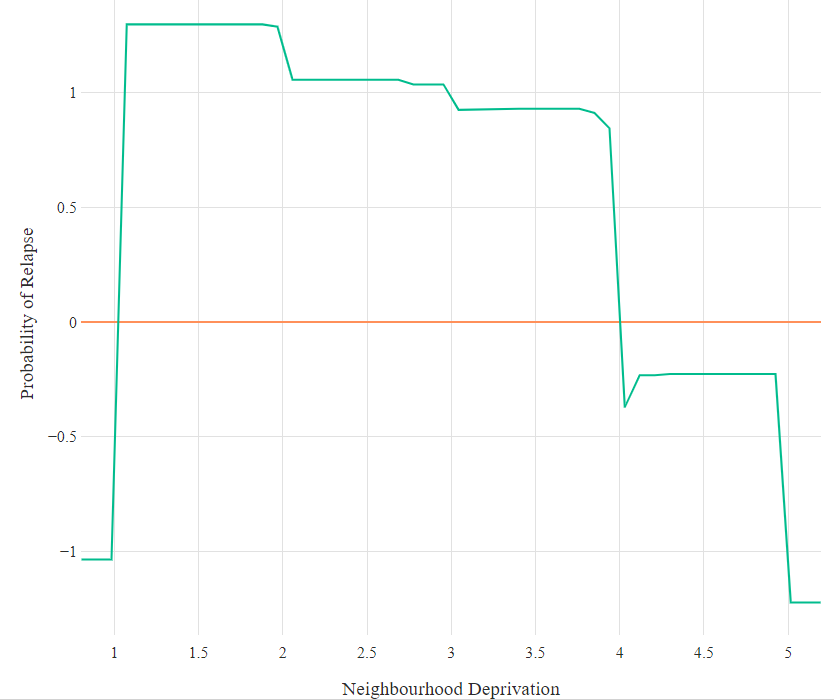


Figure G12. Partial dependence plot for the relationship between baseline WSAS and relapse in Secondary Model.

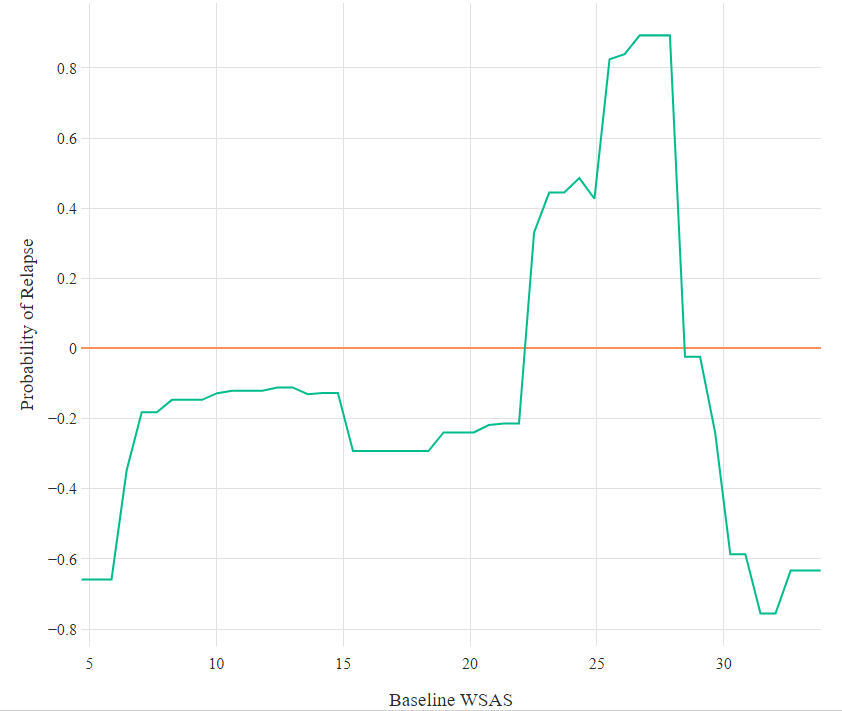


Figure G13. Partial dependence plot for the relationship between baseline GAD-7 and relapse in Secondary Model.

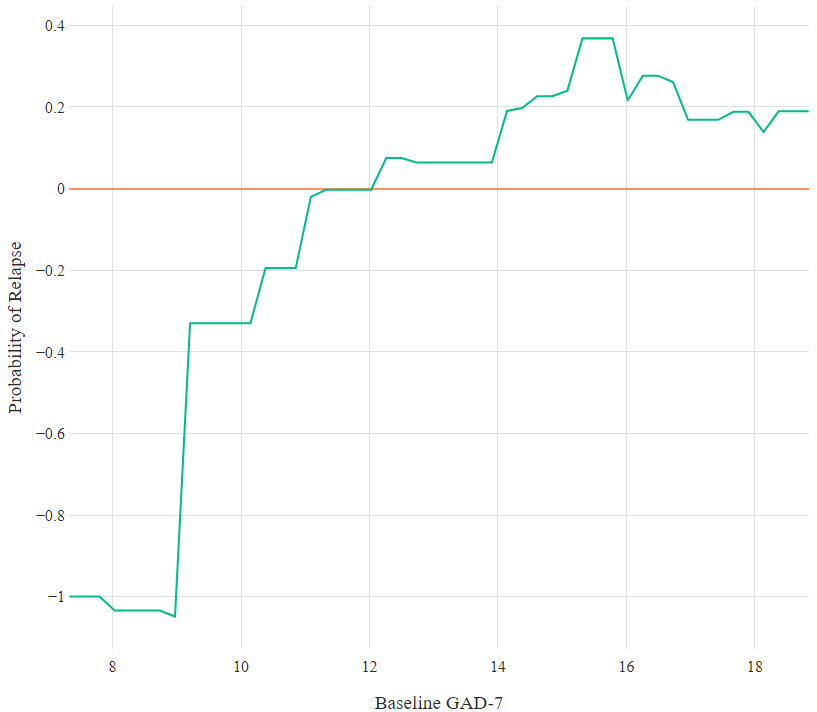


Figure G14. Partial dependence plot for the relationship between PHQ-9 at end of treatment and relapse in Secondary Model.

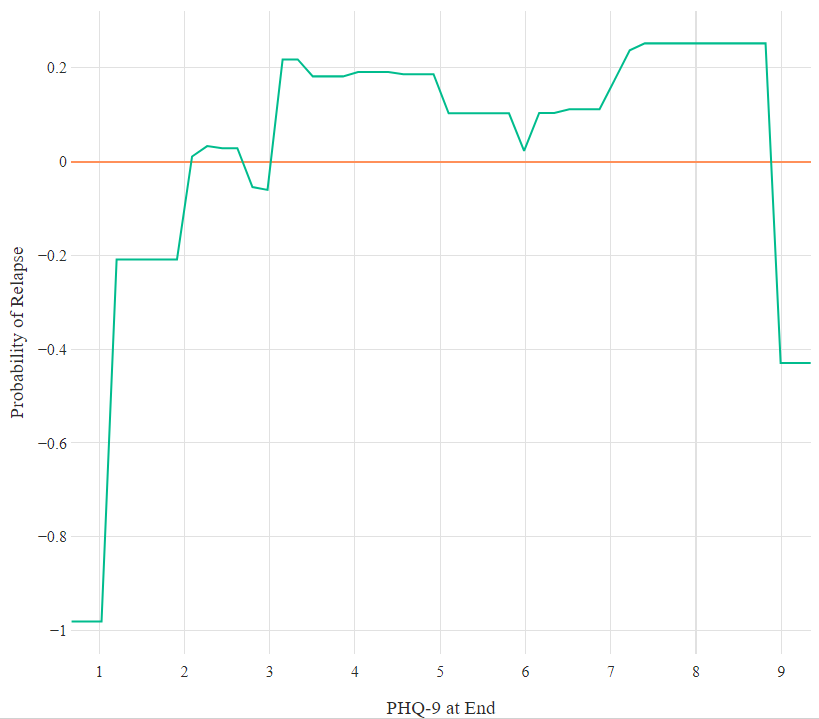


Figure G15. Partial dependence plot for the relationship between WSAS at end of treatment and relapse in Secondary Model.

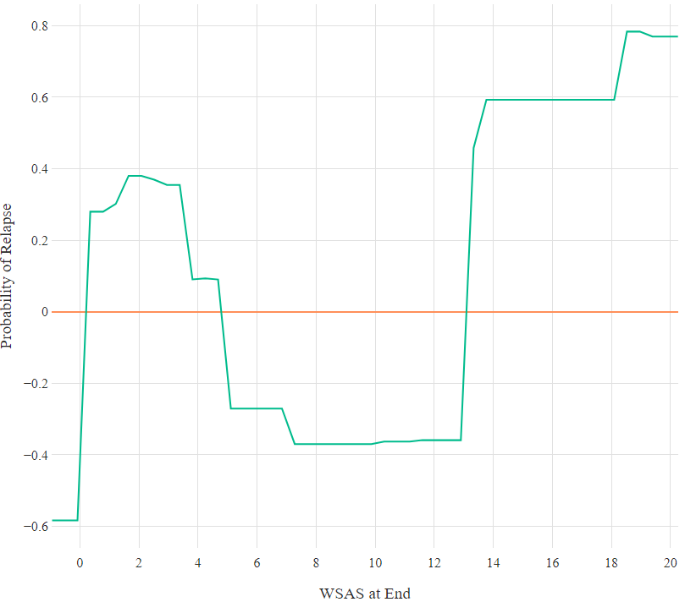


Figure G16. Partial dependence plot for the relationship between age and relapse in Secondary Model.

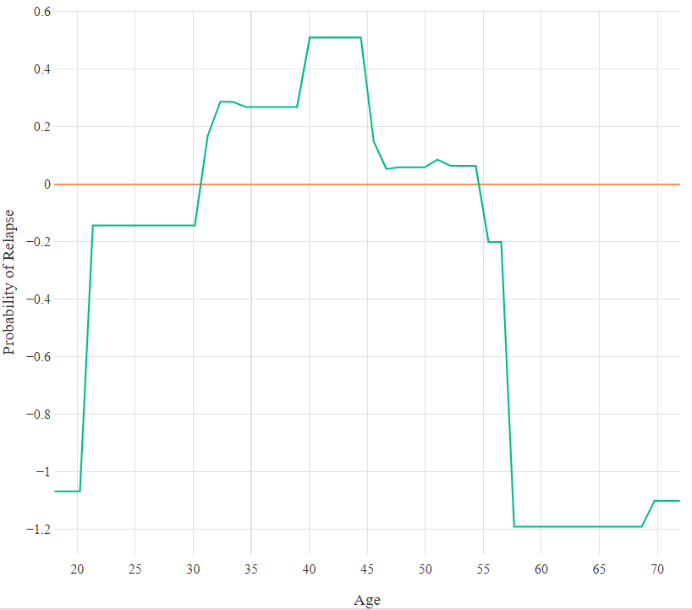


Figure G17. Partial dependence plot for the relationship between baseline PHQ-9 and relapse in Secondary Model.

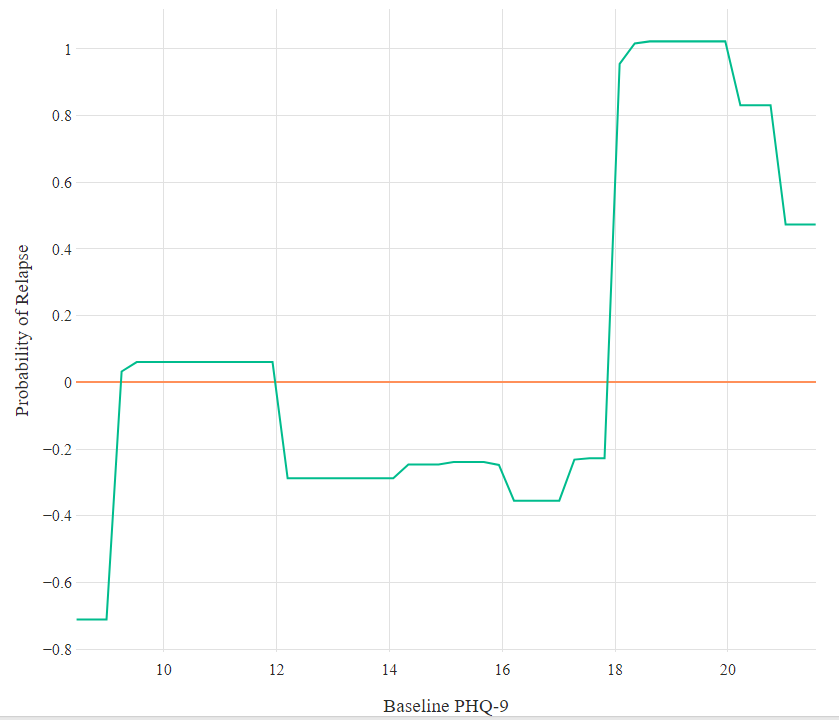


Figure G18. Partial dependence plot for the relationship between number of attended treatment sessions and relapse in Secondary Model.

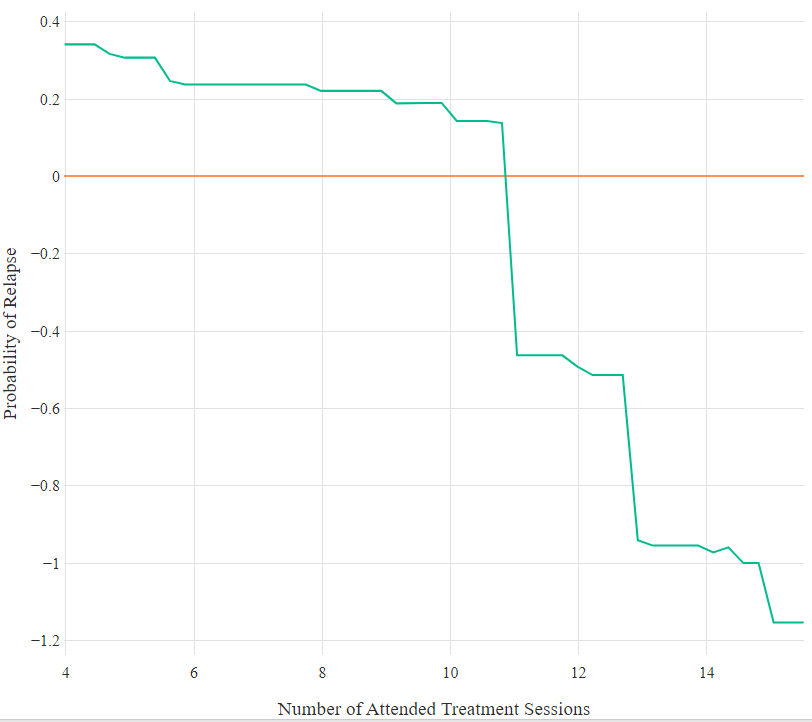


Figure G19. Partial dependence plot for the relationship between primary problem of mixed anxiety/depression and relapse in Secondary Model. X-axis values: *‘0’=not mixed anxiety/depression, ‘1’=mixed anxiety/depression.*

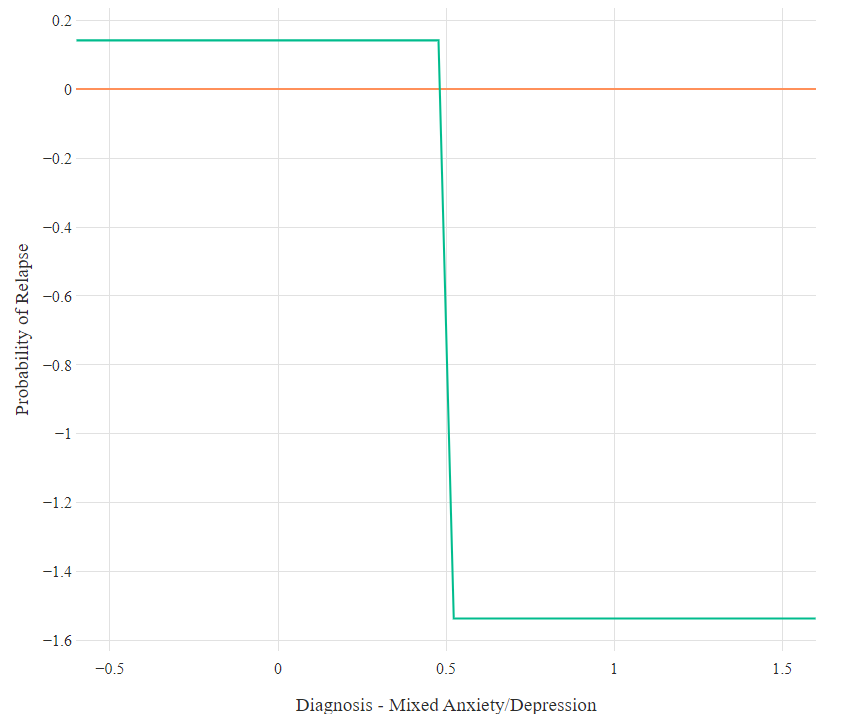


Figure G20. Partial dependence plot for the relationship between linear treatment response (GoF) in WSAS and relapse in Secondary Model.

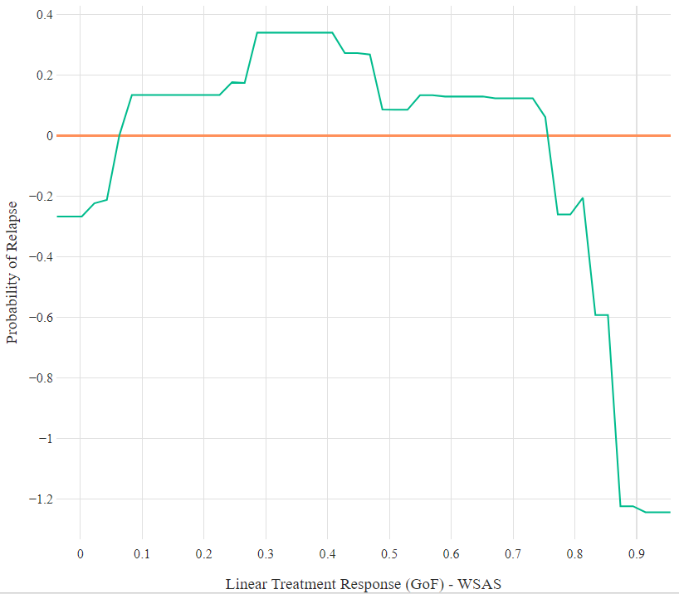


Figure G21. Partial dependence plot for the relationship between early treatment response in PHQ-9 and relapse in Secondary Model.

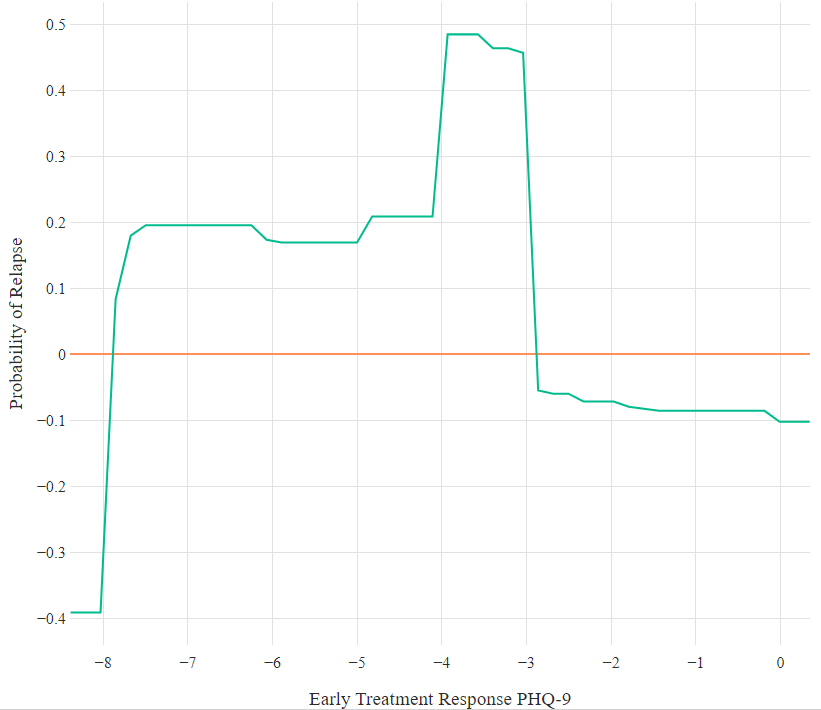


Figure G22. Partial dependence plot for the relationship between primary problem of ‘Other’ and relapse in Secondary Model. X-axis values: ‘*0’=not ‘other’, ‘1’=’other’*.

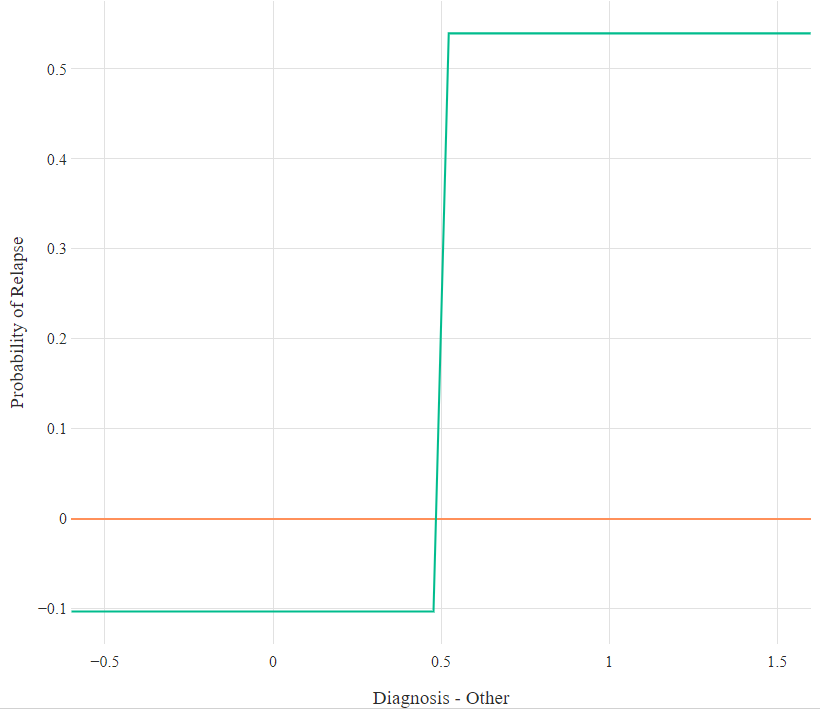


Figure G23. Partial dependence plot for the relationship between unemployment at start and relapse in Secondary Model. X-axis values: *‘0’=employment, ‘1’=unemployment’.*

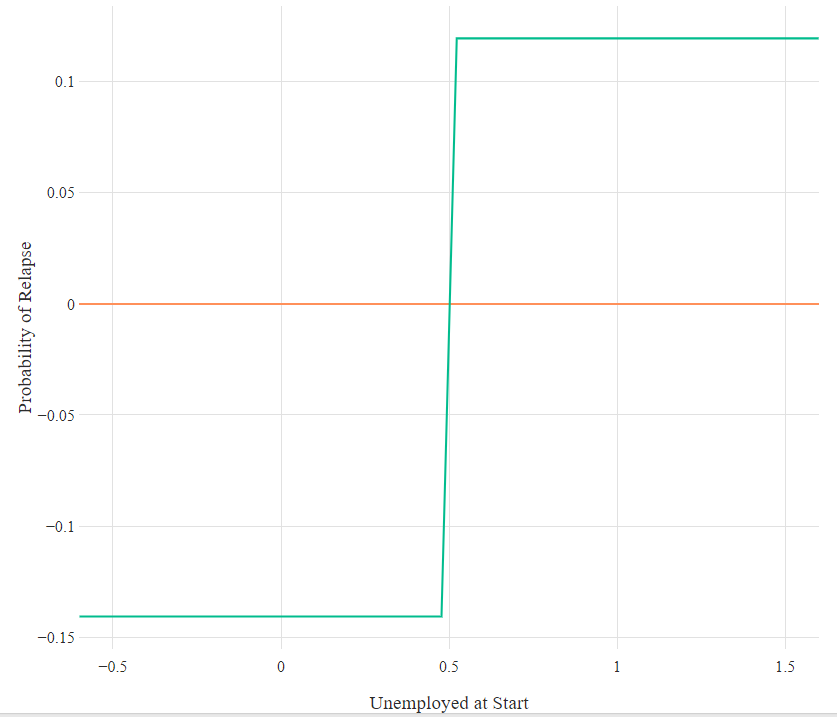


Figure G24. Partial dependence plot for the relationship between taking medication at start of treatment and relapse in Secondary Model. X-axis values: *‘0’=not taking, ‘1’=taking.*

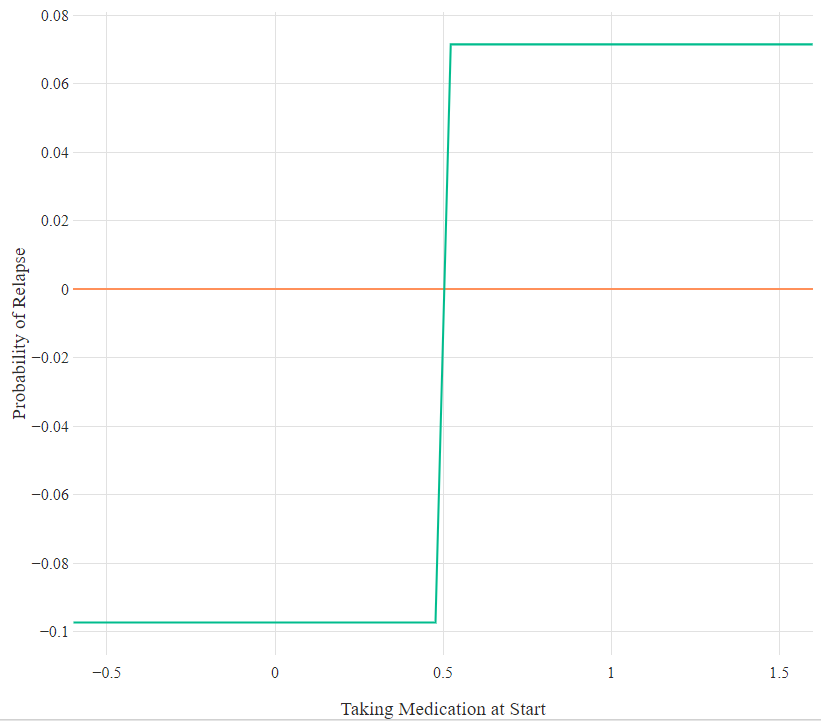


Figure G25. Partial dependence plot for the relationship between gender and relapse in Secondary Model. X-axis values: *‘1’=male, ‘2’=female.*

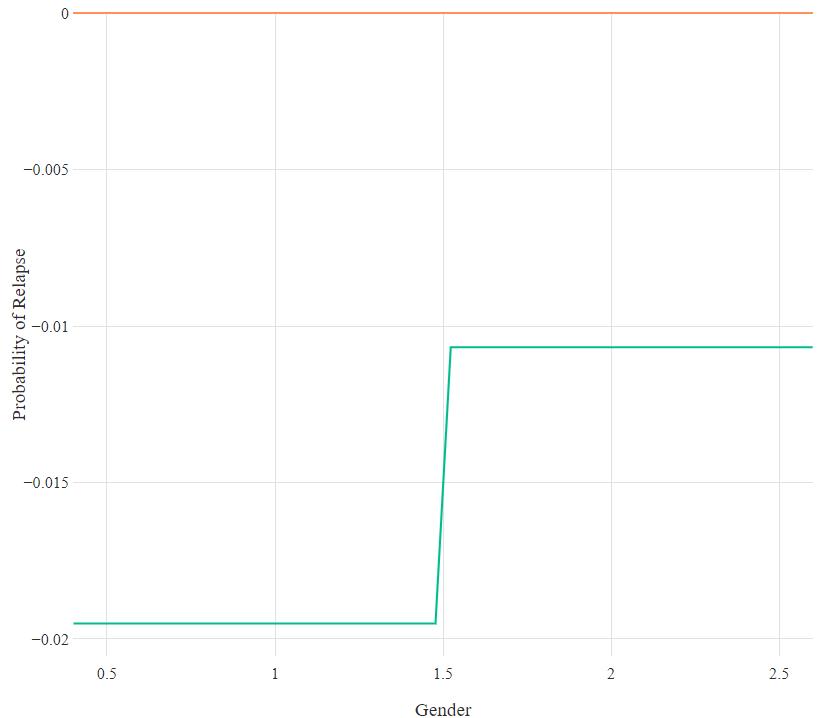


Figure G26. Partial dependence plot for the relationship between having previous treatment in episode and relapse in Secondary Model. X-axis values: *‘0’=no previous, ‘1’=previous.*

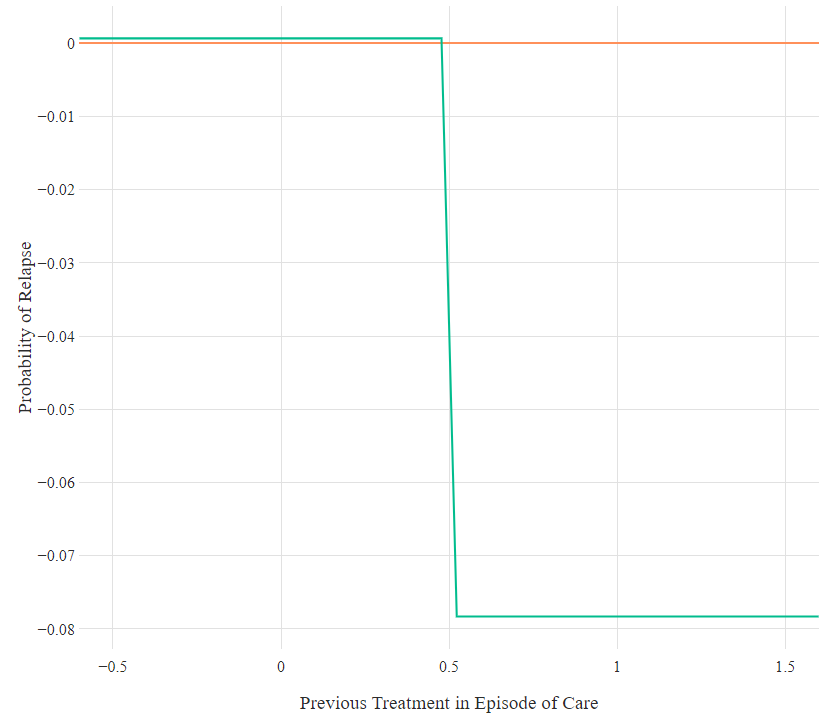


Figure G27. Partial dependence plot for the relationship between treatment modality and relapse in Secondary Model. X-axis values: *‘0’=non-CBT, ‘1’=CBT.*

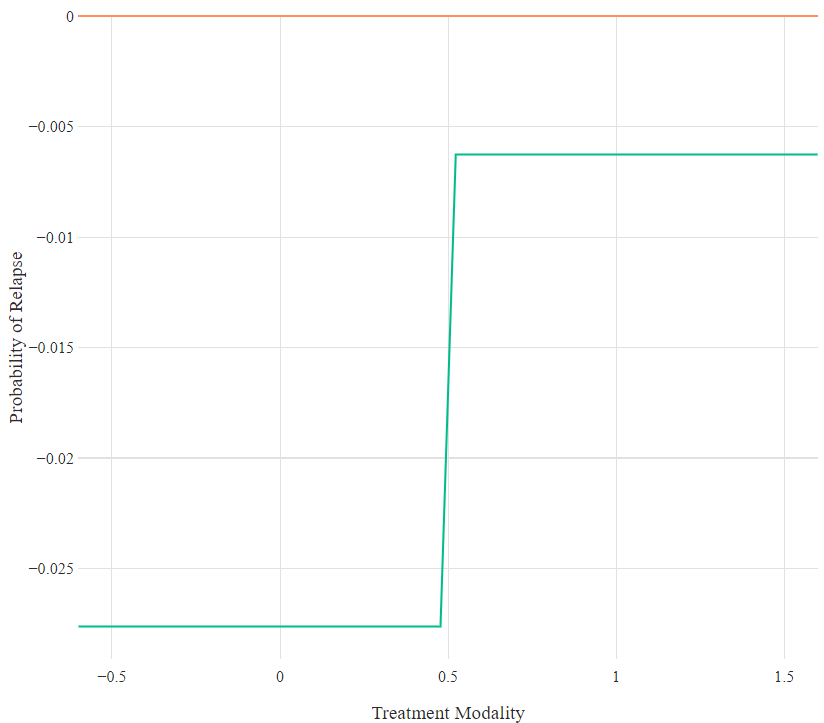
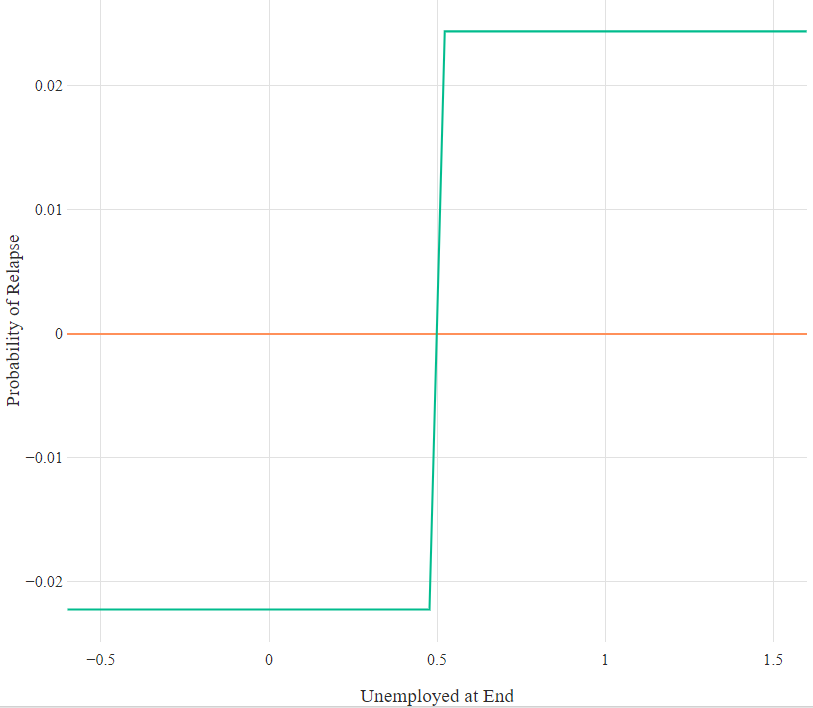


Figure G28. Partial dependence plot for the relationship between unemployment at end and relapse in Secondary Model. X-axis values: ‘0’=employment, ‘1’=unemployment.



**APPENDIX H**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table H1**  *Reasons Behind Participants’ Perceptions of Relapse Classifications, and the Risk Factors and Coping Strategies Discussed by Each Participant* | | | | |
| Cases | Relapse? | Reasons for Relapse Viewpoint | Risk Factors | Coping Strategies |
| P1 | Agree | Worse symptoms, impaired functioning | Don’t know;  C&B (poor self-efficacy) | ES (further healthcare) |
| P3 | Agree | Worse symptoms | P (work);  C&B (limited activities) | IS (non-health related distraction) |
| P5 | Agree | Worse symptoms | C&B (limited activities) | ES (further healthcare) |
| P6 | Agree | Worse symptoms | Don’t know | N/A |
| P9 | Agree | Worse symptoms | Don’t know;  C&B (poor self-efficacy) | ES (further healthcare) |
| P14 | Agree | Worse symptoms, specific life factors | P (work) | ES (further healthcare) |
| P17 | Agree | Specific life factors | P (work; living arrangements);  SE (unanticipated);  C&B (poor self-efficacy) | ES (further healthcare; work support);  IS (relax more) |
| P21 | Agree | Worse symptoms, impaired functioning, specific life factors | P (work; relationships);  SE (anticipated) | ES (further healthcare; social support; work support);  IS (LiCBT techniques; relax more; exercise and health) |
| P22 | Agree | Worse symptoms, temporary deterioration, specific life factors | P (relationships);  SE (unanticipated) | ES (end of stressful event);  IS (non-health related distraction) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P23 | Agree | Specific life factors | P (work);  SE (anticipated);  C&B (substance use) | ES (further healthcare; social support) |
| P26 | Agree | Impaired functioning, specific life factors | P (work); SE (anticipated) | ES (further healthcare) |
| P28 | Agree | Worse symptoms | SE (unanticipated);  C&B (stopped LiCBT skills) | ES (further healthcare);  IS (LiCBT techniques) |
| P33 | Agree | Worse symptoms, specific life factors | P (work; living arrangements; education);  SE (unanticipated; anticipated) | ES (further healthcare) |
| P34 | Agree | Worse symptoms, specific life factors | P (work; relationships);  SE (anticipated) | N/A |
| P35 | Agree | Specific life factors | P (relationships; education);  SE (unanticipated; anticipated);  C&B (substance use) | ES (end of stressful event);  IS (exercise and health; continue as doing) |
| P36 | Agree | Specific life factors | P (relationships);  SE (unanticipated) | ES (end of stressful event);  IS (continue as doing) |
| P38 | Agree | Worse symptoms, specific life factors | P (relationships);  SE (unanticipated) | IS (LiCBT techniques) |
| P39 | Agree | Worse symptoms, temporary deterioration, specific life factors | P (work) | ES (further healthcare) |
| P40 | Agree | Worse symptoms | P (work; education);  SE (anticipated);  C&B (stopped LiCBT skills) | ES (further healthcare; social support);  IS (LiCBT techniques; relax more; exercise and health) |
| P41 | Agree | Worse symptoms, impaired functioning, specific life factors | P (work) | ES (further healthcare);  IS (relax more) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P44 | Agree | Worse symptoms | P (work);  SE (unanticipated) | ES (further healthcare) |
| P2 | Unsure | Specific life factors | P (work) | ES (seek social support);  IS (LiCBT techniques) |
| P8 | Unsure | Worse symptoms, temporary deterioration | N/A | ES (further healthcare) |
| P10 | Unsure | Temporary deterioration, partial relapse, specific life factors | SE (unanticipated) | ES (end of stressful event) |
| P11 | Unsure | Temporary deterioration | P (relationships);  SE (anticipated) | ES (further healthcare; social support);  IS (relax more) |
| P12 | Unsure | Specific life factors | P (work; relationships) | N/A |
| P13 | Unsure | Worse symptoms, specific life factors | P (work);  SE (unanticipated) | ES (further healthcare);  IS (exercise and health) |
| P16 | Unsure | Worse symptoms, potential medical issue | P (work) | ES (further healthcare) |
| P18 | Unsure | Worse symptoms, temporary deterioration | P (work; living arrangements; education);  SE (anticipated) | ES (end of stressful event);  IS (relax more) |
| P20 | Unsure | Worse symptoms; partial relapse, specific life factors | P (work) | ES (further healthcare) |
| P24 | Unsure | Potential medical issue | SE (unanticipated) | N/A |
| P25 | Unsure | Worse symptoms, temporary deterioration, specific life factors | SE (anticipated) | ES (end of stressful event);  IS (continue as doing) |
| P27 | Unsure | Worse symptoms, temporary deterioration | P (living arrangements);  SE (unanticipated) | ES (further healthcare) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P29 | Unsure | Worse symptoms, temporary deterioration | P (work; relationships; education) | ES (social support) |
| P31 | Unsure | Partial relapse | N/A | ES (further healthcare) |
| P32 | Unsure | Specific life factors | N/A | N/A |
| P4 | Disagree | Temporary deterioration, other specific life factors | P (work) | IS (continue as doing) |
| P7 | Disagree | Temporary deterioration, other specific life factors | P (work);  SE (anticipated) | ES (end of stressful event);  IS (LiCBT techniques) |
| P15 | Disagree | Temporary deterioration | N/A | ES (further healthcare);  IS (relax more; exercise and health) |
| P19 | Disagree | Temporary deterioration, other specific life factors | P (work) | ES (social support) |
| P30 | Disagree | Temporary deterioration | N/A | ES (further healthcare) |
| P37 | Disagree | Temporary deterioration | P (relationships) | IS (LiCBT techniques) |
| P42 | Disagree | Other specific life factors | P (work; relationships) | N/A |
| P43 | Disagree | Temporary deterioration, other specific life factors | P (relationships; living arrangements) | ES (further healthcare);  IS (LiCBT techniques; exercise and health) |
| P45 | Disagree | Other specific life factors | P (work; relationships);  SE (unanticipated) | IS (relax more; exercise and health) |
| Abbreviations: C&B, cognitive and behavioural factors; ES, external strategies; IS, internal strategies; P, psychosocial factors; SE, stressful events | | | | |

**APPENDIX I**

|  |  |
| --- | --- |
| **Table I1**  *Potential Predictors Input into XGBoost Model and Glossary for Predictors* | |
| Predictor | Definition |
| Age | The age of the patient. |
| Baseline GAD-7 | GAD-7 score at assessment. |
| Baseline PHQ-9 | PHQ-9 score at assessment. |
| Baseline WSAS | WSAS score at assessment. |
| Change in GAD-7 | Change in GAD-7 over treatment. |
| Change in PHQ-9 | Change in PHQ-9 over treatment. |
| Change in WSAS | Change in WSAS over treatment. |
| Diagnosis (Affective Disorder) | Primary diagnosis of an affective disorder, recorded by clinician. |
| Diagnosis (Mixed Anxiety/Depression) | Primary diagnosis of mixed anxiety and depression, recorded by clinician. |
| Diagnosis (Anxiety Disorder) | Primary diagnosis of an anxiety disorder, recorded by clinician. |
| Diagnosis (OCD) | Primary diagnosis of OCD, recorded by clinician. |
| Diagnosis (PTSD) | Primary diagnosis of PTSD, recorded by clinician. |
| Diagnosis (Other) | Primary diagnosis of a disorder other than those above, recorded by clinician. |
| Dropped-Out | Patients dropping out of treatment or not. |
| Ethnicity | The ethnicity of the patient (White British vs other). |
| GAD-7 at End | GAD-7 score at the end of treatment. |
| Gender | The gender of the patient (male vs female). |
| Intensity of Treatment at End of Episode | Patients receiving low- or high- intensity treatment at the end of their treatment episode. |
| Long-Term Condition | Patients having a long-term physical condition or not. |
| Neighbourhood Deprivation | Level of deprivation of patient’s neighbourhood. |
| Number of Treatment Sessions Attended | The number of sessions patients received. |
| Number of Treatment Sessions Cancelled | The number of sessions patients cancelled. |
| Number of Treatment Sessions Not Attended | The number of sessions patients did not attend. |
| PHQ-9 at End | PHQ-9 score at end of treatment. |
| Taking Medication at Beginning | Patients taking medication at assessment or not. |
| Taking Medication at End | Patients taking medication at end of treatment or not. |
| Time Between Assessment and Discharge | The number of days between assessment appointment and treatment discharge. |
| Time Between Referral and Assessment | The number of days between referral to service and their assessment appointment. |
| Unemployment at Beginning | Patients unemployed at assessment or not. |
| Unemployment at End | Patients unemployment at end of treatment of not. |
| WSAS at End | WSAS score at end of treatment. |
| Abbreviations: GAD-7, Generalized Anxiety Disorder scale; OCD, obsessive compulsive disorder; PHQ-9, Patient Health Questionnaire; PTSD, post-traumatic stress disorder; WSAS, Work and Social Adjustment Scale. | |

**APPENDIX J**

The partial dependence plots below are organized with regards to the two developed models, and displayed in order of relative importance. The x-axis of each plot represents the values of a specific predictor, while the y-axis represents probabilities of relapse (log odds ratio). Values above zero on the y-axis indicate an increased risk for relapse according to the respective developed model, while values below zero indicate a higher probability for remaining in-remission.

**Primary Model**

Figure J1. Partial dependence plot for the relationship between ethnicity and relapse in Primary Model. X-axis values: *‘0’=non-White, ‘1’=White British, ‘2’=missing.*

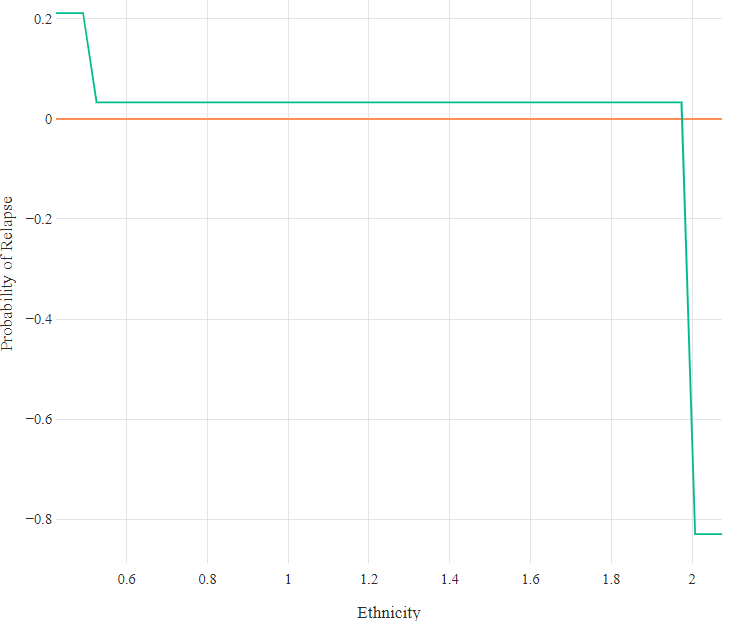


Figure J2. Partial dependence plot for the relationship between long-term conditions and relapse in Primary Model. X-axis values: *‘0’=no LTC, ‘1’=LTC.*

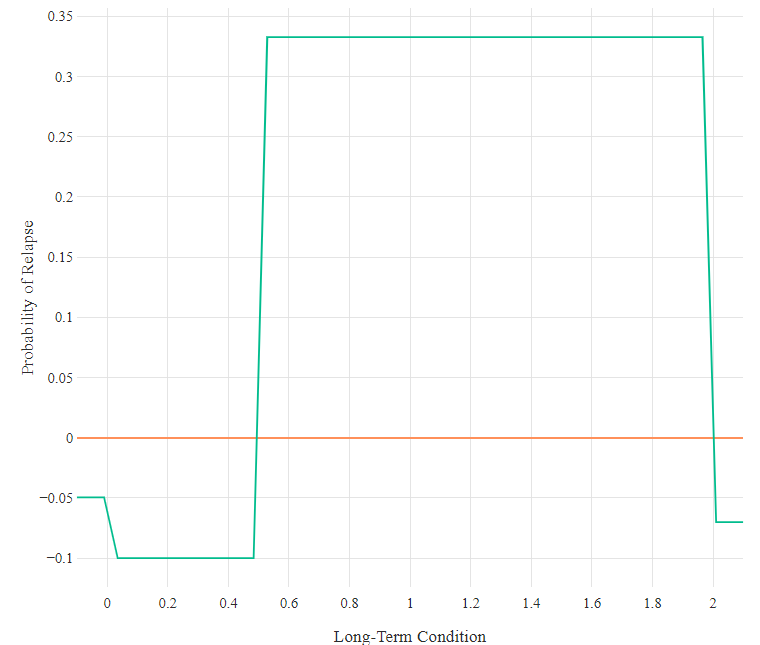


Figure J4. Partial dependence plot for the relationship between age and relapse in Primary Model.

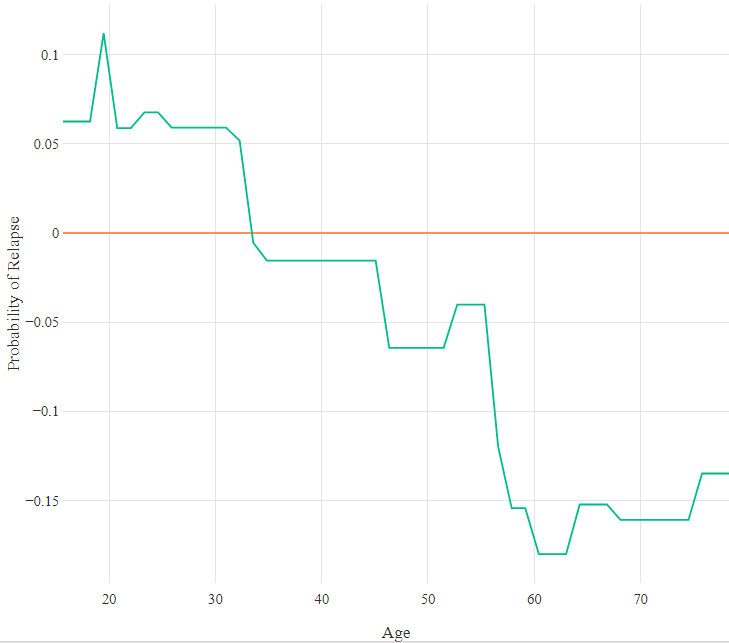


Figure J5. Partial dependence plot for the relationship between baseline GAD-7 and relapse in Primary Model.

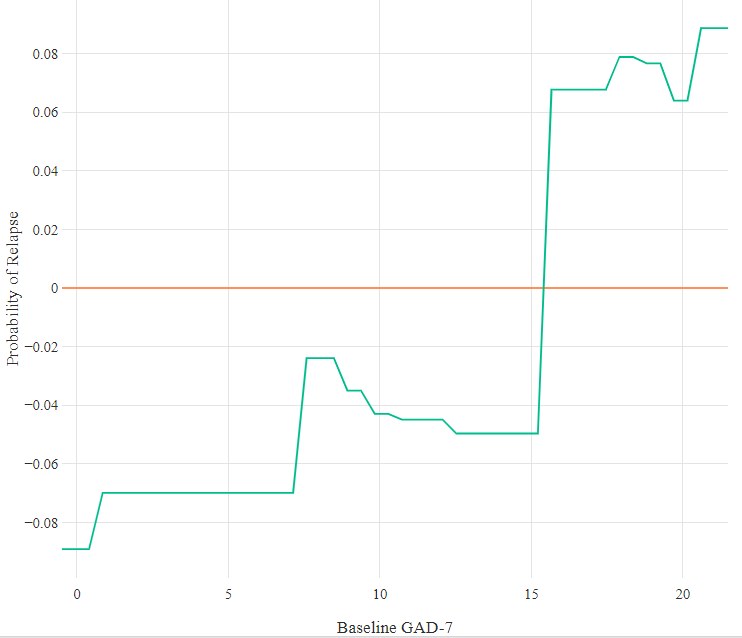


Figure J3. Partial dependence plot for the relationship between PHQ-9 at end of treatment and relapse in Primary Model.

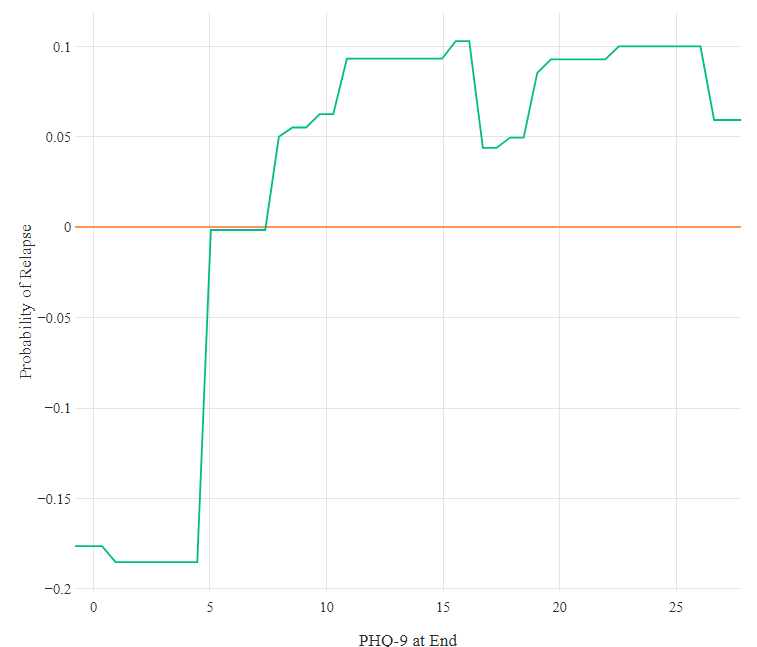


Figure J6. Partial dependence plot for the relationship between number of attended treatment sessions and relapse in Primary Model.

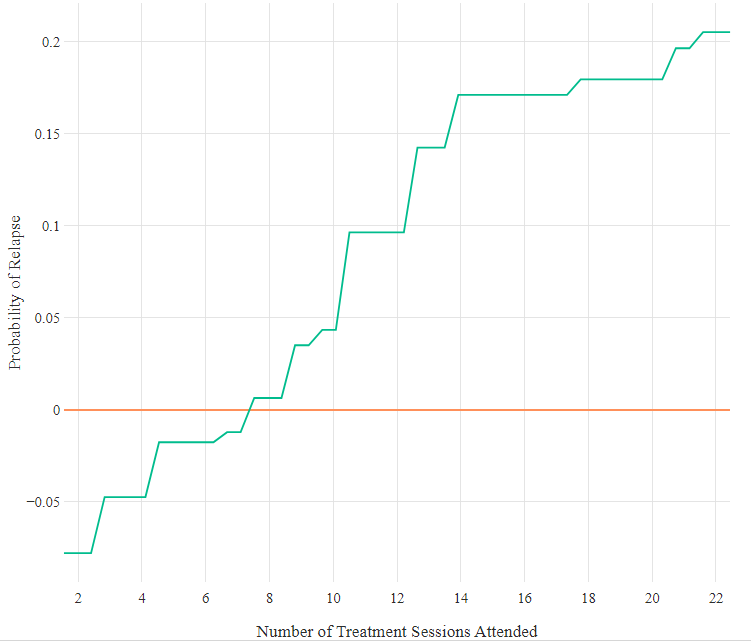


Figure J7. Partial dependence plot for the relationship between change in PHQ-9 over treatment and relapse in Primary Model.

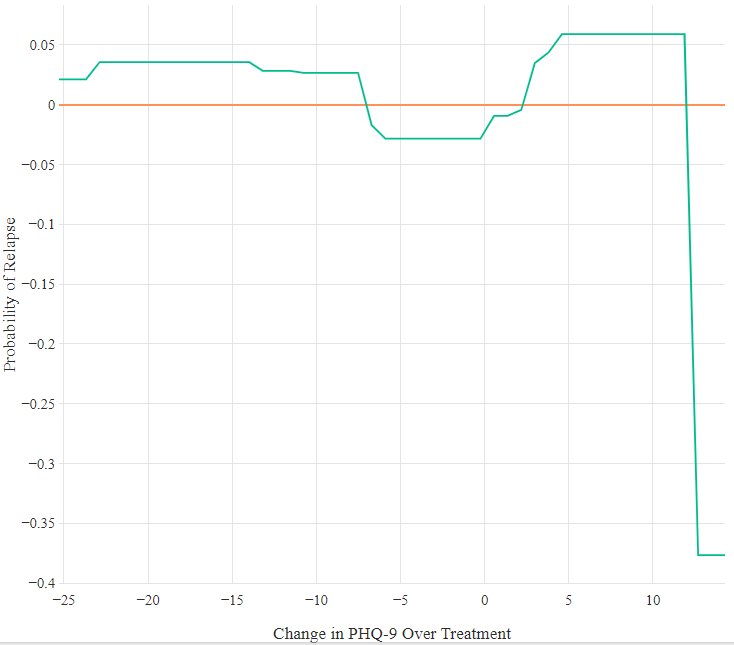


Figure J8. Partial dependence plot for the relationship between number of cancelled treatment sessions and relapse.

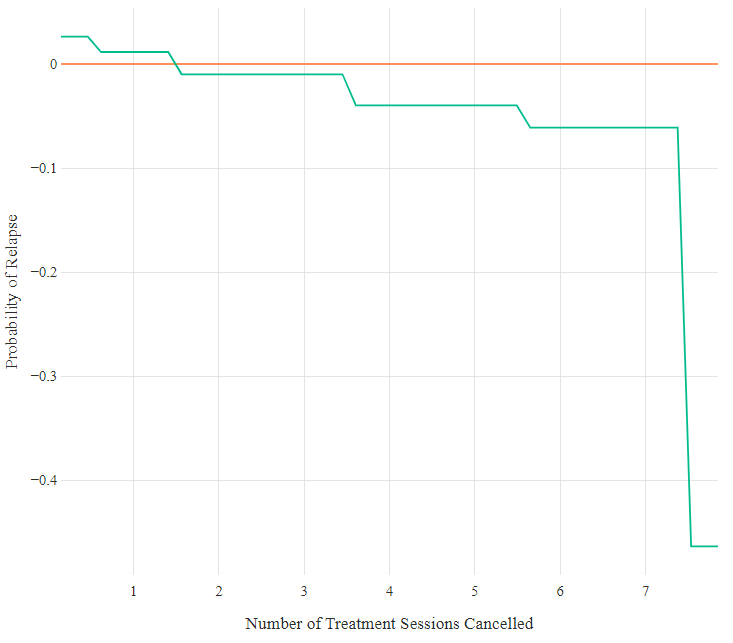


Figure J9. Partial dependence plot for the relationship between neighbourhood deprivation and relapse in Primary Model.

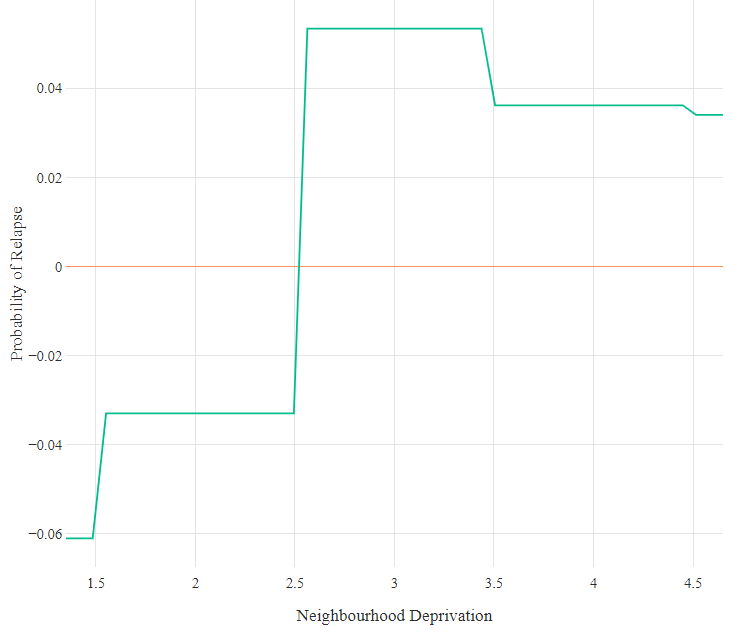


Figure J10. Partial dependence plot for the relationship between intensity of treatment at end of treatment and relapse in Primary Model. X-axis values: *‘2’=low, ‘3’=high.*

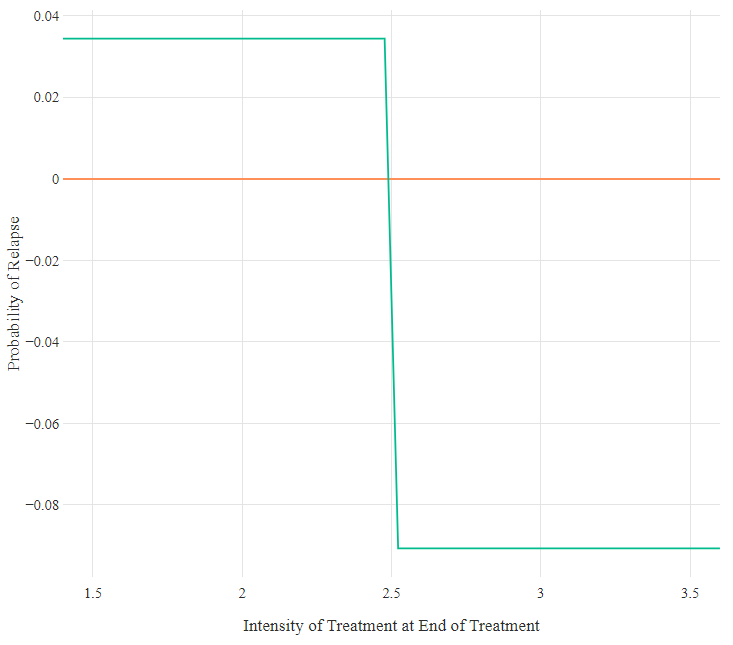


Figure J11. Partial dependence plot for the relationship between primary problem of anxiety and relapse in Primary Model. X-axis values; *‘0’=not anxiety, ‘1’=anxiety.*

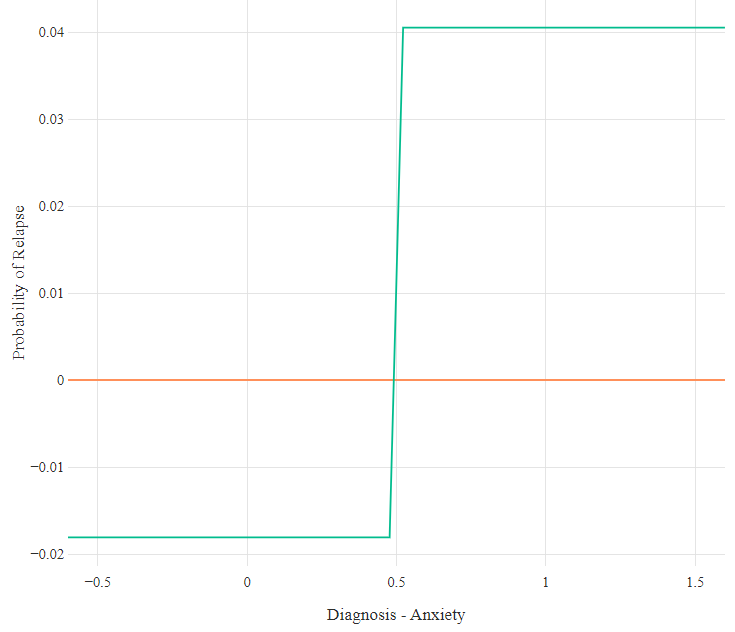


Figure J12. Partial dependence plot for the relationship between taking medication at end of treatment and relapse in Primary Model. X-axis values: *‘0’=not taking, ‘1’=taking.*

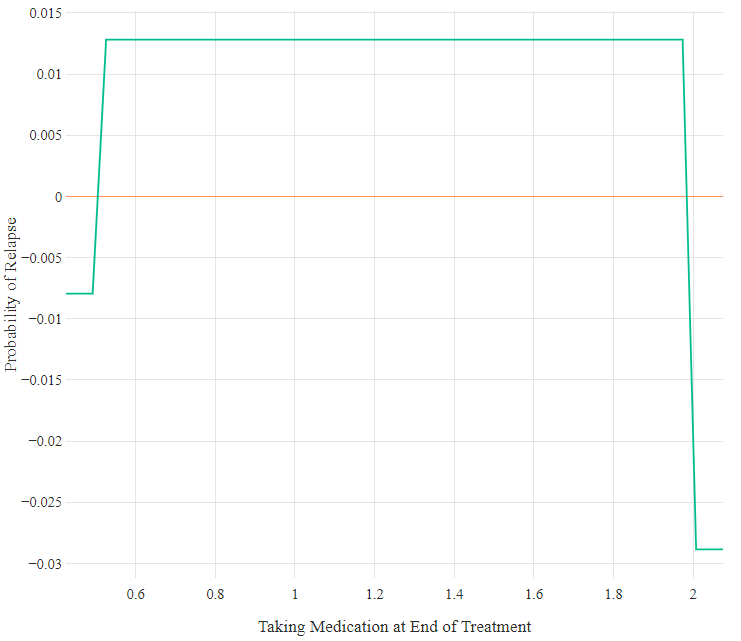


Figure J13. Partial dependence plot for the relationship between primary problem of PTSD and relapse in Primary Model. X-axis values: *‘0’=not PTSD, ‘1’=PTSD.*

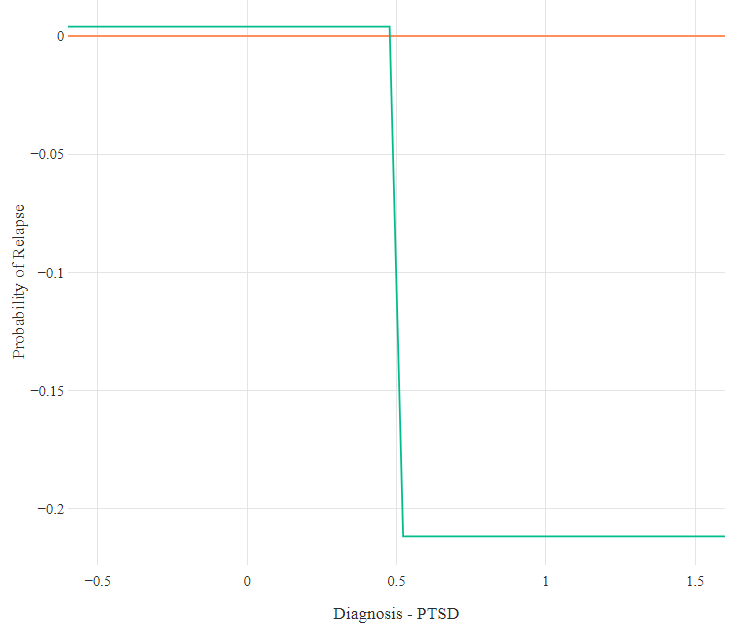


Figure J14. Partial dependence plot for the relationship between primary problem of mixed anxiety/depression and relapse in Primary Model. X-axis values: *‘0’=not mixed anxiety/depression, ‘1’=mixed anxiety/depression.*

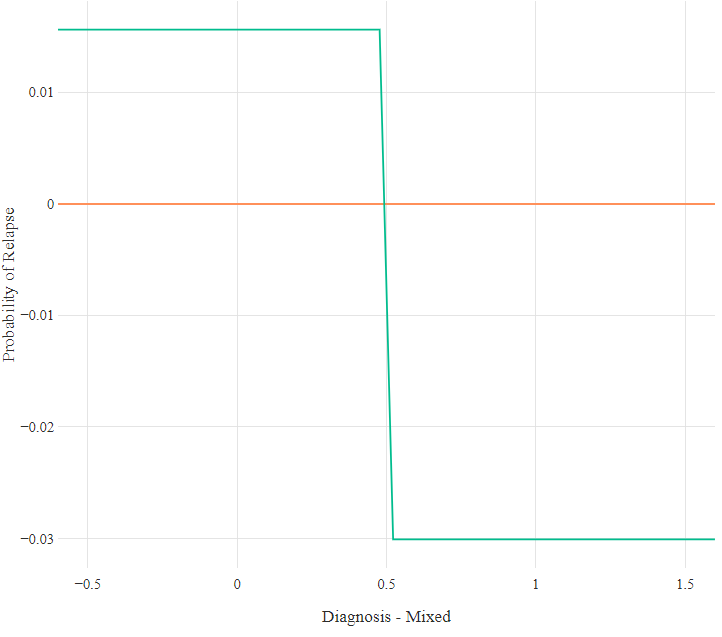
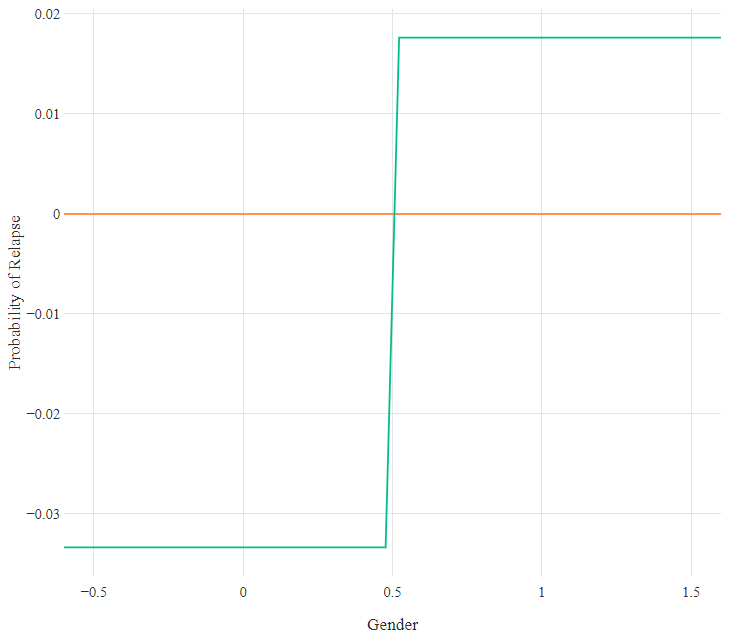


Figure J15. Partial dependence plot for the relationship between gender and relapse in Primary Model. X-axis values: *‘0’=male, ‘1’=female.*



**Secondary Model**

Figure J16. Partial dependence plot for the relationship between wait for assessment appointment and relapse in Primary Model.

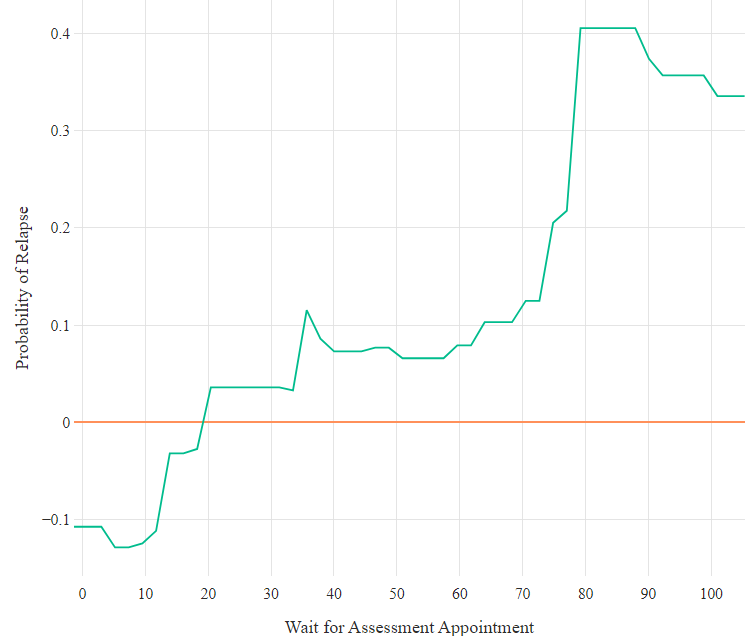


Figure J17. Partial dependence plot for the relationship between having a long-term condition and relapse in Primary Model. X-axis values: *‘0’=no LTC, ‘1’=LTC.*

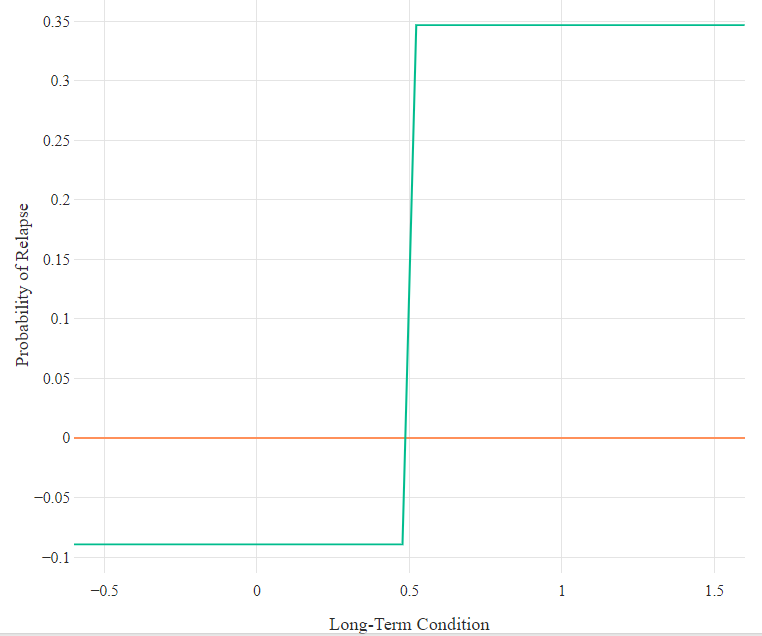


Figure J18. Partial dependence plot for the relationship between age and relapse in Primary Model.

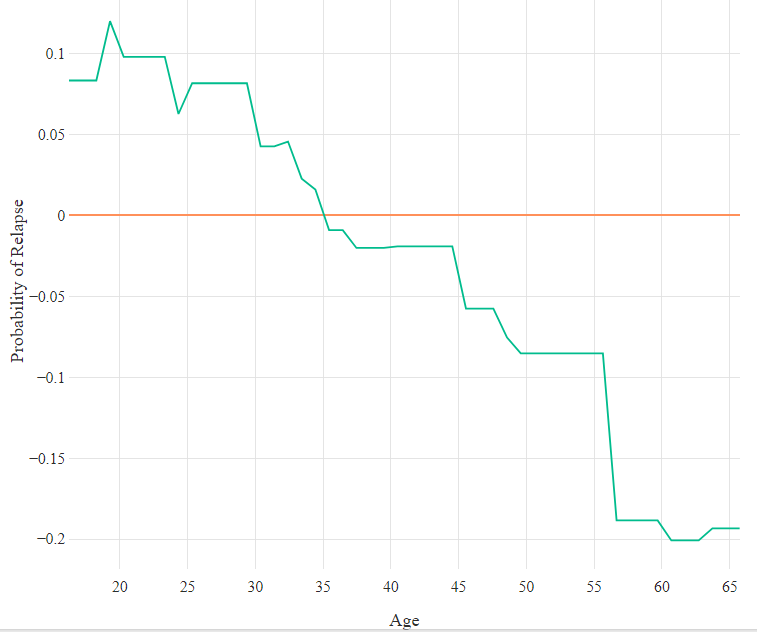


Figure J19. Partial dependence plot for the relationship between baseline PHQ-9 and relapse in Primary Model.

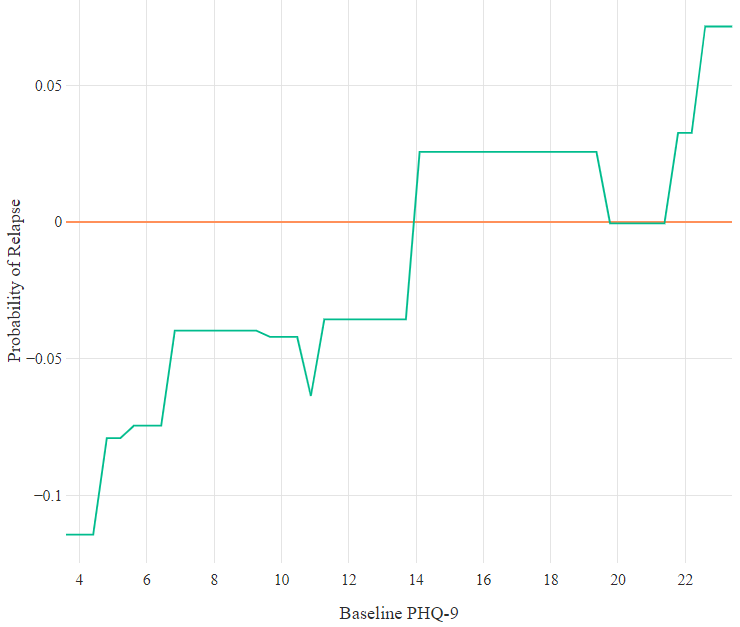


Figure J20. Partial dependence plot for the relationship between number of attended treatment sessions and relapse in Primary Model.

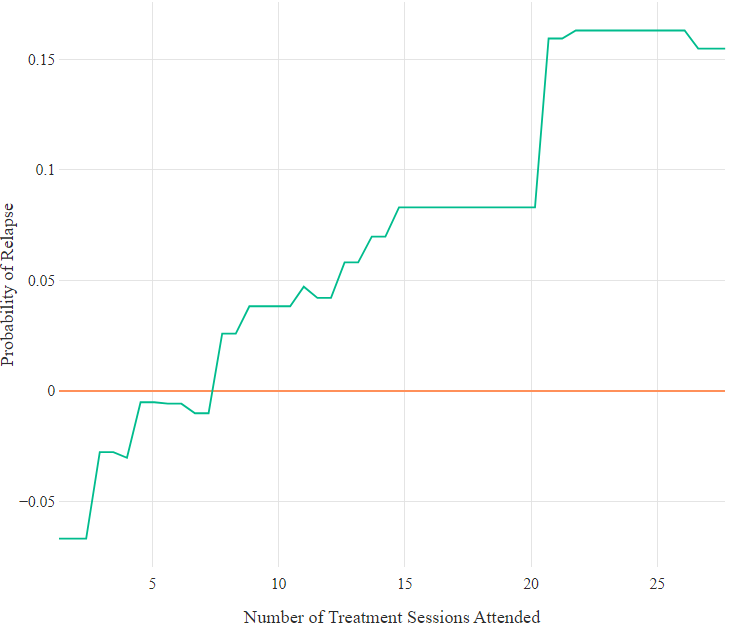


Figure J21. Partial dependence plot for the relationship between PHQ-9 at end of treatment and relapse in Primary Model.

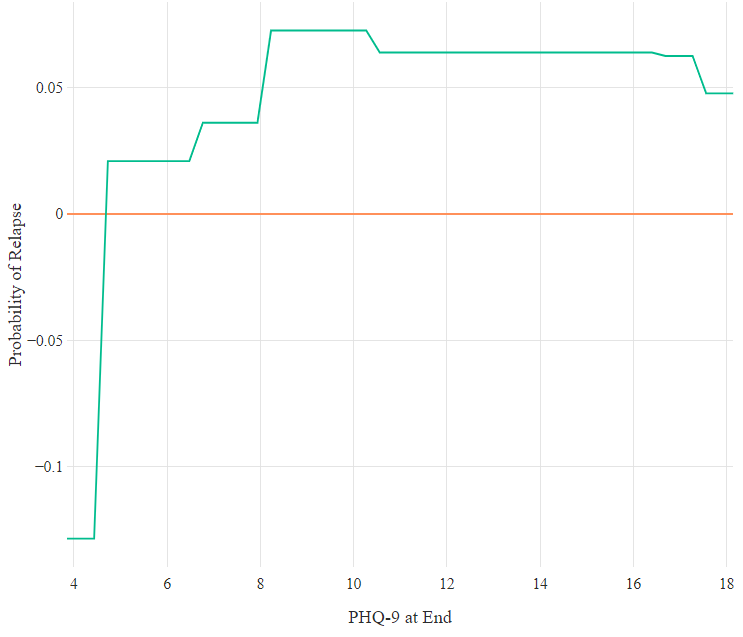


Figure J22. Partial dependence plot for the relationship between ethnicity and relapse in Primary Model. X-axis values: *‘0’=non-White, ‘1’=White British.*

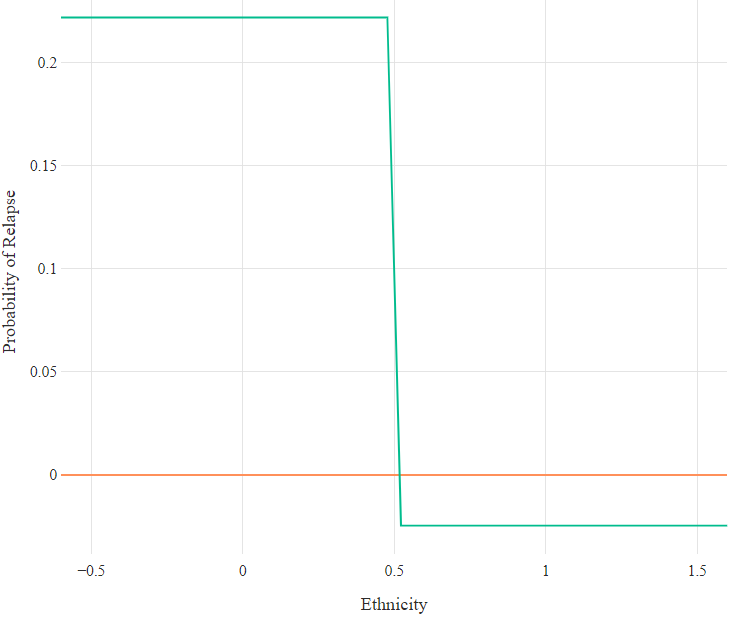


Figure J23. Partial dependence plot for the relationship between number of cancelled treatment sessions and relapse in Primary Model.

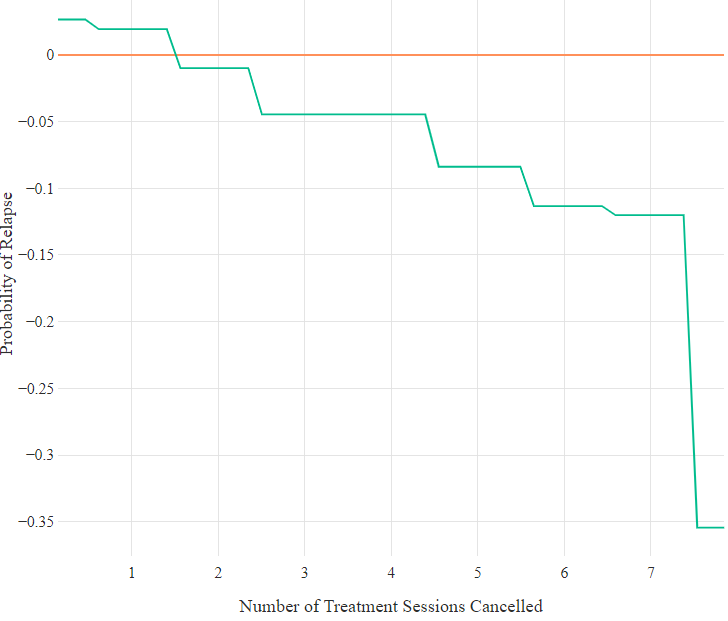


Figure J24. Partial dependence plot for the relationship between neighbourhood deprivation and relapse in Primary Model.

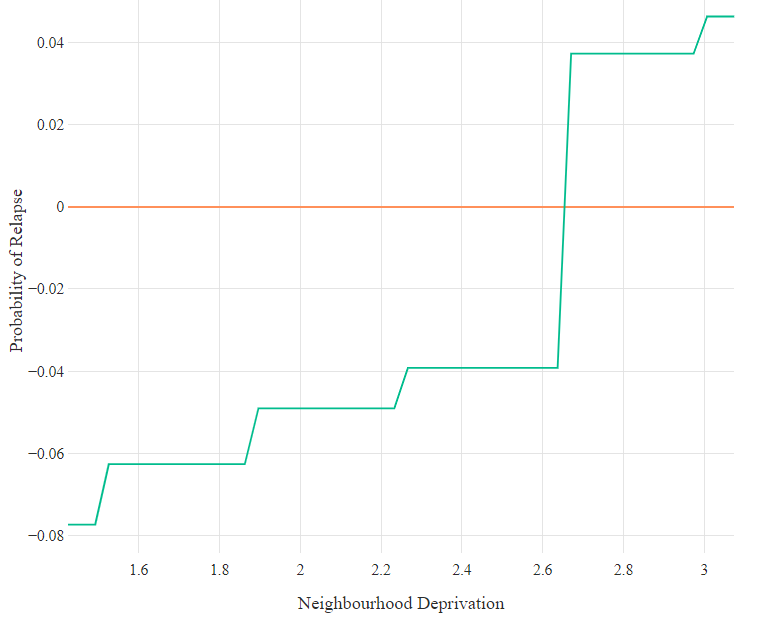


Figure J25. Partial dependence plot for the relationship between taking medication at start of treatment and relapse in Primary Model. X-axis values: *‘0’=not taking, ‘1’=taking.*

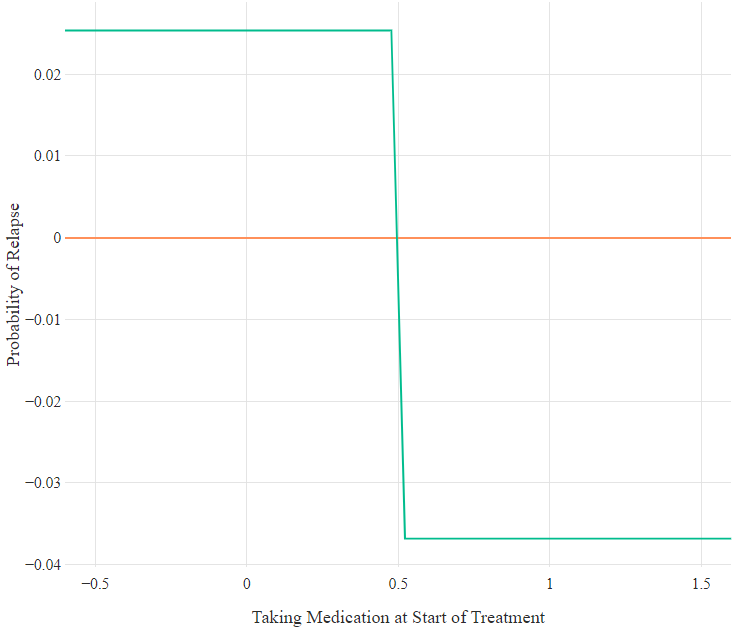
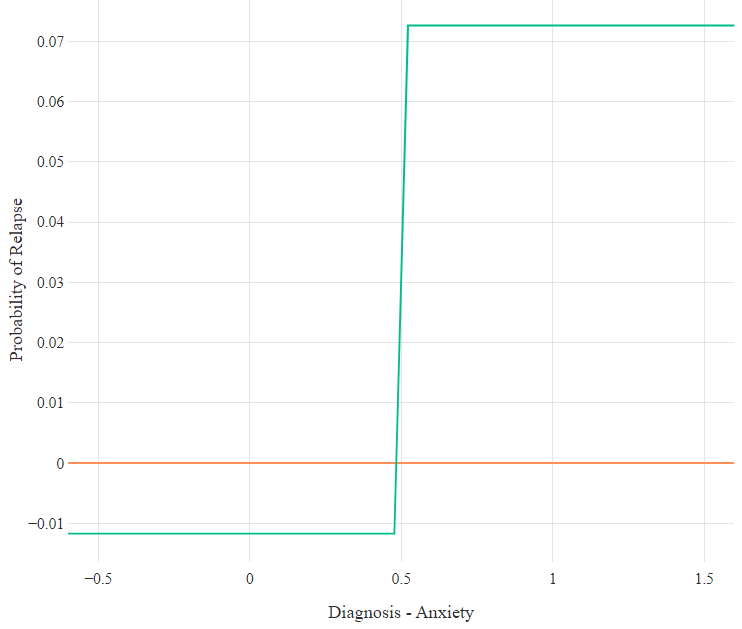


Figure J26. Partial dependence plot for the relationship between primary problem of anxiety and relapse in Primary Model. X-axis values: *‘0’=not anxiety, ‘1’=anxiety.*



1. The third section of the checklist (i.e. Section C – ‘Will the results help locally?’) was not used for this review, as it was not considered relevant to the review’s objective. [↑](#footnote-ref-1)
2. Cohen’s kappa was interpreted using Landis and Koch’s (1977) guidelines: <0 no agreement; 0-0.20 slight agreement; 0.21-0.40 fair agreement; 0.41-0.60 moderate agreement; 0.61-0.80 substantial agreement; and 0.81-1 almost perfect agreement. [↑](#footnote-ref-2)
3. Fava et al. (1995) used the revised, third edition of the DSM (American Psychiatric Association [APA], 1987), while Fava et al. (2001a, 2001b) used the fourth edition (APA, 1994). [↑](#footnote-ref-3)
4. Although this was not explicitly stated in Fava et al. (1995), it was stated in Fava et al. (2001b). [↑](#footnote-ref-4)
5. It should be noted that *n*=12 participants were classified as having relapsed in the original WYLOW study due to them returning to mental healthcare to receive further support (either pharmacological support or support through IAPT), and not because of changes in their outcome measures. [↑](#footnote-ref-5)
6. The AUC statistics reported in this chapter differ slightly from the values reported in the published journal article (Lorimer et al., 2021). This is due to the AUCs in this chapter being calculated through comparisons of all predicted probabilities within the total sample, while the AUCs reported in the journal article represented ‘average AUCs’ that were identified through five-fold internal cross-validation. This change was implemented to ensure consistency across thesis chapters. [↑](#footnote-ref-6)
7. It was initially planned to also investigate relapse following LiCBT and group-based interventions. However, the follow-up reviews that were delivered following these interventions were recorded too inconsistently, with different practitioners from those who delivered the initial treatment episode often conducting the follow-up reviews. This prevented it from being possible to determine from the data whether patients received a follow-up review or not. [↑](#footnote-ref-7)
8. Specifically, *n*=41 patients were excluded for not being above clinical thresholds at start of treatment, *n*=6 patients were excluded for having missing symptom scores at start of treatment, *n*=86 patients were excluded for being above clinical thresholds at end of treatment, and *n*=1 patients was excluded for being below clinical thresholds at start of treatment and being above thresholds at end of treatment. [↑](#footnote-ref-8)
9. Most patients (84.0%) completed their episode of treatment less than 12 months before the end of the study observation period (May 2021). [↑](#footnote-ref-9)
10. Demographic statistics calculated using only those cases with full, relevant information. [↑](#footnote-ref-10)
11. Sixteen patients who were not followed-up received two separate high-intensity treatment episodes during the study period. For the purposes of investigating potential sources of selection bias, only the first episodes received by these patients were included. [↑](#footnote-ref-11)
12. PHQ-9 score at the final treatment session and unemployment at the beginning of treatment were chosen as the bases for the two ‘noise variables’, as they have previously been indicated as being predictive of relapse (e.g., Chapter 3). [↑](#footnote-ref-12)
13. Of the *n*=2031 patients who completed an episode of high-intensity psychotherapy but did not attend a follow-up review appointment, *n*=107 received internet-enabled CBT administered by digital health company IESO Digital Health. Patients who received treatment in this pathway were not followed-up generally. [↑](#footnote-ref-13)
14. If applying the reliable change indices as applied in Chapter 3 (i.e., >= 5 for both the PHQ-9 and GAD-7), an additional *n*=3 patients would be classed as having relapsed. This would result in a new total of *n*=53 relapse cases, representing a new relapse rate of 20.2%. [↑](#footnote-ref-14)
15. Although many predictors related to residual symptoms were identified in Chapter 3, only residual depressive (PHQ-9) symptoms at the final treatment session was explored. This was due to there being no missing values for this variable, and also that when the sample was divided between patients with and without PHQ-9 scores ≥ 5, the resulting subgroups both had sufficient, relatively balanced numbers of participants (*n*=18 vs *n*=29). [↑](#footnote-ref-15)
16. Issues related to spouses were indexed as being both ‘family issues’ and ‘romantic/sexual relationship issues’. Consequently, two participants who discussed issues associated with their husbands were indexed as discussing both forms of social relationship issues. [↑](#footnote-ref-16)
17. A treatment episode was determined to have occurred if a patient was recorded as having at least two appointments with the service. This is due to the fact that the first recorded appointment involves an initial assessment to determine suitability for psychological treatment. Patients with two or more recorded appointments, therefore, accessed at least one therapy session after the initial assessment. [↑](#footnote-ref-17)
18. Demographic statistics calculated using only those cases with full, relevant information. [↑](#footnote-ref-18)
19. Reported descriptive statistics were calculated using cases with full, relevant information. [↑](#footnote-ref-19)
20. This number is smaller than the *n*=24838 reported in the previous paragraph, as this number did not include those patients who were discharged from treatment in 2015, still in treatment by 1st July 2015, or excluded for likely data inputting errors related to treatment return. [↑](#footnote-ref-20)
21. If applying the reliable change indices (RCI) applied in Chapter 3 (i.e., >= 5 for both the PHQ-9 and GAD-7), *n*=10 fewer patients would be classed as having relapsed. This would result in a new total of *n*=1054 relapse cases, representing a new relapse rate of 36.9%. Meanwhile, if applying the RCI applied in Chapter 4 (i.e., >= 6 for the PHQ-9 and >= 5 for the GAD-7), *n*=32 fewer patients would be classed as having relapsed. This would result in a new total of *n*=1032 relapse cases, and a new relapse rate of 36.1%. [↑](#footnote-ref-21)