PHD

Precision Medicine in Depression and Anxiety: Measuring Clinically Meaningful Change, Assessing Symptom-Specific Response in Psychotherapy, and Exploring Stratified Treatment Selection

(Alternative Format Thesis)

Bauer-Staeb, Clarissa

Award date:
2023

Awarding institution:
University of Bath

Precision Medicine in Depression and Anxiety: Measuring Clinically Meaningful Change, Assessing Symptom-Specific Response in Psychotherapy, and Exploring Stratified Treatment Selection

Clarissa Maria Mercedes Bauer-Stäb

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Psychology

September 2022

Copyright Notice

Attention is drawn to the fact that copyright of this thesis rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as licenced, permitted by law or with the consent of the author or other copyright owners, as applicable.

Access to this thesis in print or electronically is restricted until

_____

Signed on behalf of the Doctoral College

_____

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally. Individual chapters are accompanied by an authorship statement indicating my contribution to the work.

Conflict of Interest Statement

The candidate declares no conflict of interest.

Table of Contents

1 ACKNOWLEDGEMENTS 6

2 GENERAL ABSTRACT 7

3 GENERAL INTRODUCTION 9

3.1 Depression and Anxiety 9

3.2 Treatment for Depression and Anxiety 11

3.3 Precision/Personalised Medicine 12

3.4 Traditional Approaches to Personalised Psychotherapy 14

3.5 Improving Access to Psychological Therapies (IAPT) Programme 25

3.6 More Recent Approaches to Personalised Psychotherapy 30

3.7 Aims and Objectives 33

3.8 References 36

4 EFFECTIVE DOSE 50 METHOD AS THE MINIMAL CLINICALLY IMPORTANT DIFFERENCE: EVIDENCE FROM DEPRESSION TRIALS 46

4.1 Statement of Authorship 46

4.2 Abstract 48

4.3 What's New 49

4.4 Introduction 50

4.5 Methods 51

4.6 Results 53

4.7 Discussion 55

4.8 Tables 60

4.9 References 67

4.10 Declarations of interest 71

4.11 CRediT Author Statement 71

4.12 Acknowledgements 71

4.13 Trial Registration 71

5 PERSONALISED PSYCHOTHERAPY IN PRIMARY CARE: AN EVALUATION OF DATA-DRIVEN TREATMENT ALLOCATION TO COGNITIVE BEHAVIOURAL THERAPY VS. COUNSELLING FOR DEPRESSION 72

5.1 Statement of Authorship 72

5.2 Abstract 73

5.3 Introduction 74
5.4 Method 75
5.5 Results 81
5.6 Discussion 83
5.7 Tables 89
5.8 References 94
5.9 Declaration of Interests 99
5.10 CRediT Author Statement 99
5.11 Acknowledgements 99
5.12 Funding 99
5.13 Ethics Statement 99
6 TRAJECTORIES OF DEPRESSION AND GENERALISED ANXIETY SYMPTOMS OVER THE COURSE OF COGNITIVE BEHAVIOURAL THERAPY IN PRIMARY CARE: AN OBSERVATIONAL, RETROSPECTIVE COHORT 100
6.1 Statement of Authorship 100
6.2 Abstract 102
6.3 Introduction 103
6.4 Methods 105
6.5 Results 108
6.6 Discussion 111
6.7 Tables and figures 117
6.8 References 122
6.1 Declaration of Interests 128
6.2 CRediT Author Statement 128
6.3 Acknowledgements 128
6.4 Funding 128
7 THE EARLY IMPACT OF COVID-19 ON PRIMARY CARE PSYCHOLOGICAL THERAPY SERVICES: A DESCRIPTIVE TIME SERIES OF ELECTRONIC HEALTHCARE RECORDS 129
7.1 Statement of Authorship 129
7.2 Abstract 131
7.3 Research in Context 132
7.4 Introduction 133
7.5 Methods 134

7.6 Results 136
7.7 Discussion 140
7.8 Tables and Figures 146
7.9 References 155
7.10 Declaration of Interests 159
7.11 CRediT Author Statement 159
7.12 Acknowledgements 159
7.13 Funding 159
8 GENERAL DISCUSSION 160
8.1 Summary and conclusions of main findings 160
8.2 Strengths & limitations/challenges 164
8.3 Future directions for personalised psychotherapy 170
8.4 Conclusion 176
8.5 References 177
9 SUPPLEMENTARY MATERIAL 183
9.1 Supplementary Material A: Chapter 3 183
9.2 Supplementary Material B: Chapter 4 190
9.3 Supplementary Material C: Chapter 5 203
9.4 Supplementary Material D: Chapter 6 215
9.5 Supplementary Material E: Chapter 7 230
9.6 Supplementary Material F: Data Access Statement 239

1 Acknowledgements

First and foremost, I am extremely grateful to my supervisory team Katherine Button, Julian Faraway, and Emma Griffith. Kate - thank you for pushing me to finish my PhD and publish my work even when I thought I couldn't and for supporting my decision to find a job during my PhD. Julian – thank you for your patience when I was learning R, all your invaluable statistical advice, and your words of encouragement that kept me grounded. Emma – thank you for providing an invaluable clinical perspective to all my work, for reading my drafts so thoroughly, and for always being so positive throughout my PhD.

I would also like to say a huge thank you to Chris Eldridge and everyone at Mayden who supported me with my work and completed the data extraction. I am also incredibly grateful to the NHS trusts that agreed to participate in my research, as well as the patients whose data I used, without which this PhD would not have been possible.

Finally, I am incredibly grateful to my family and friends, especially my parents and my brother. Mom, thank you for your unwavering support in any situation. Dad, thank you for inspiring me to do a PhD in the first place - I miss you. Frederick, thank you for all that you do for me.

2 General Abstract

Depression and anxiety are amongst the most common mental health problems worldwide. While multiple psychotherapies exist, treatment response is varied and leaves room for improvement. There has been a growing interest in personalised psychotherapy, which aims to better match psychotherapies to patients in order to maximise treatment outcomes. The aim of my PhD was to explore different approaches that have the potential to contribute to personalised psychotherapy for depression and anxiety.

A critical component of evaluating if a treatment is effective, including any efforts made in personalised psychotherapy, is an outcome metric against which success can be measured. Many outcome metrics rely solely on statistical change on outcome questionnaires and ignore the patient perspective. In study 1, I developed a method to approximate the minimal clinically important difference (MCID), which is the smallest difference in outcome questionnaires that are tangible to patients. The patient's subjective experiences of improvement were mapped onto the changes in depression and anxiety questionnaire scores. The MCID was defined as the point at which there was a 50% probability of feeling better. The results were in line with previous literature showing that there was strong baseline dependency – how much change is associated with patients feeling better depends on their baseline severity, with patients experiencing more severe symptomatology requiring larger changes. As such, the MCID was personalised to each level of baseline severity. Metrics like the MCID can be used in clinical practice and research to support the evaluation of treatments.

In study 2, I conducted an analysis that used a prescriptive algorithm to assess whether assigning patients to two commonly available psychotherapies for depression, based on baseline characteristics, could improve clinical outcomes. The analysis was performed using a retrospective, observational cohort of patients receiving cognitive behavioural therapy or counselling for depression, which was derived from electronic healthcare records. The results showed that no individual characteristic was important in determining a differential response to the two treatments. However, when considering all characteristics collectively, small improvements of 4-10% were seen across the sample if patients were assigned to the treatment that was indicated as optimal according to the algorithm.

In study 3, I explored the impact of cognitive behavioural therapy on the individual symptoms of depression and anxiety. Electronic healthcare records were used to generate a retrospective, observational cohort of patients that received cognitive behavioural therapy and the individual symptom trajectories were modelled. Each symptom trajectory was compared to the average trajectory of all other symptoms. While all symptoms improved across cognitive behavioural therapy, the results suggested that *low mood/hopelessness* and *guilt/worthlessness* improved at a faster pace relative to other depressive symptoms, with *sleeping problems, appetite changes,* and *psychomotor retardation/agitation* improving relatively slower. The anxiety symptoms that improved fastest were *uncontrollable worry* and *too much worry,* with *irritability* and *restlessness* improving at a slower pace. Understanding variability in the response of individual symptoms to treatment has the potential to inform personalised psychotherapy. It potentially allows patients to be matched to psychotherapies in the future based on symptom profiles and/or informing where an augmentation of treatment may be necessary to enhance the response of symptoms that improve slower.

The results of my PhD, and the wider literature, suggest that small improvements may be possible through personalised psychotherapy. However, the clinical magnitude of these remains debatable, with no "magic bullet" found to date. While the current approaches in personalised psychotherapy are unlikely to lead to a profound impact at the individual level, small improvements across the population may nonetheless be relevant from a public health perspective if such approaches were adopted in national healthcare systems in the future. When interpreting the results of my PhD, the observational nature of the data must be considered, which makes inferring causality difficult. Further research with more robust research designs is required to validate the findings and develop the ideas presented further. Ground-breaking advances in the field of personalised psychotherapy are likely dependent on establishing a robust understanding of the mechanisms underlying depression and anxiety as well as the psychotherapies themselves. This may allow for more robust matching of psychotherapies to clinical presentations based on empirical evidence of the mechanisms.

3 General Introduction

Depression and anxiety are leading causes of disability, having a profound impact on people's lives as well as societal implications (James et al., 2018). Pharmacotherapy and psychotherapy are treatment options for depression and anxiety. While various psychotherapies exist, the evidence suggests they broadly exhibit equal effectiveness (Cuijpers, Karyotaki, de Wit, & Ebert, 2020; Cuijpers, Sijbrandij, et al., 2014). However, they are only effective for approximately half of patients (Cuijpers et al., 2021). This leaves room for improvement, given that many people don't respond. There has been a focus on personalised medicine to maximise clinical outcomes and reduce the total burden of disease. In England, Improving Access to Psychological Therapies (IAPT) services deliver psychological interventions for depression and anxiety in primary care settings (Clark, 2011). Electronic healthcare records from IAPT provide an invaluable dataset for exploratory analyses to examine if personalised psychotherapy can improve clinical outcomes. Below, I will define key terms and concepts and describe previous literature for context.

3.1 Depression and Anxiety

3.1.1 Definition

According to the American Psychiatric Association, depression is characterised by nine symptoms, which include 1) low mood, 2) anhedonia, 3) weight/appetite changes, 4) changes in sleep, 5) psychomotor agitation/retardation, 6) fatigue/low energy, 7) feelings of guilt/worthlessness, 8) concentration problems, and 9) thoughts of death (American Psychiatric Association, 2013). For a diagnosis of Major Depressive Disorder, at least five symptoms must be present in a 2-week period. One must include what are considered core symptoms - low mood or anhedonia (American Psychiatric Association, 2013).

Concerning anxiety, this thesis will focus on general anxiety or Generalised Anxiety Disorder (GAD). According to the American Psychiatric Association, GAD is characterised by 1) excessive worry and 2) uncontrollable worry, which must be present for at least six months (American Psychiatric Association, 2013). Furthermore, for a diagnosis of GAD, patients' worry must be associated with at least 3 of the following: 3) restlessness, 4) fatigue, 5) concentration problems, 6) irritability, 7) muscle tension, and/or 8) sleep problems (American Psychiatric Association, 2013).

3.1.2 Prevalence and impact

Depression and anxiety are amongst the most common mental health problems worldwide. The global prevalence and incidence of depressive disorders (in thousands) were estimated at 264 455.6 (95% CI: 246 380.1 to 286 312.0) and 258 164.5 (95% CI: 238 280.7 to 281 665.5), respectively in 2017 (James et al., 2018). There appears to be an increasing trend over time - the increase in years lived with disability for depressive disorders between 2007 and 2017 has been estimated at 14.3% (95% CI: 13.1 to 15.6) (James et al., 2018). Estimates suggest 1.45 million people in England will experience depression by 2026 (McCrone, Dhanasiri, Patel, Knapp, & Lawton-Smith, 2007). The global prevalence and incidence (in thousands) of anxiety disorders were 284 360.1 (95% CI: 265 607.5 to 304 531.7) and 42 340.0 (95% CI: 39 597.5 to 45 199.5) respectively in 2017 (James et al., 2018). Similar to depression, there appears to be an increasing trend over time with the increase in years lived with disability for anxiety disorders between 2007 and 2017 having been estimated at 12.8% (95% CI: 11.7 to 14.0) (James et al., 2018). Estimates suggest 2.56 million people in England will experience anxiety by 2026 (McCrone et al., 2007).

Depression and anxiety can have a profound impact on individuals, with qualitative research of people with lived experience suggesting a sense of overwhelming and constant suffering, characterised by complex and multidimensional pain incorporating physical, emotional, and social aspects (Weitkamp, Klein, & Midgley, 2016; Woodgate, Tennent, Barriage, & Legras, 2020). In addition to the impact that depression and anxiety have on people's lives, they also have far-reaching implications for health services and the economy. The World Health Organisation has attributed 37.1% of all years lived with disability to mental ill-health (Prince et al., 2007). Mental ill-health is one of the most significant causes of disability in the United Kingdom, accounting for 22.8% of the total burden of disease (Department of Health and Social Care, 2011). The total cost of depression in England in 2007 was estimated to be £7.5 billion, with £1.68 billion attributed to service costs and £5.82 billion attributed to lost earnings (McCrone et al., 2007). The projected costs of depression in 2026 are estimated at £12.15 billion, with £2.96 billion resulting from service costs and £9.19 billion from lost earnings (McCrone et al., 2007). The cost of anxiety in England in 2007 was estimated to be £8.94 billion, with £1.24 billion attributed to service costs and £7.7 billion to lost earnings (McCrone et al., 2007). The projected costs of depression in 2026 are estimated at £14.19

billion, with £2.04 billion resulting from service costs and £12.15 billion from lost earnings (McCrone et al., 2007).

3.2 Treatment for Depression and Anxiety

Two primary approaches to treating depression and anxiety are psychotherapy and pharmacotherapy. Meta-analytic evidence suggests that patients show a three-fold preference for psychotherapy relative to pharmacotherapy (McHugh, Whitton, Peckham, Welge, & Otto, 2013). The National Institute of Health and Care Excellence (NICE) recommends various psychological therapies for new episodes of less severe depression: guided self-help, group or individual cognitive behavioural therapy, group or individual behavioural activation, group exercise, group mindfulness and meditation, interpersonal psychotherapy, counselling, short-term psychodynamic psychotherapy, and behavioural couples therapy for depression (National Institute for Health and Care Excellence, 2022). For new episodes of more severe depression, NICE recommends guided self-help, individual cognitive behavioural therapy, individual behavioural activation, individual problem-solving, counselling, short-term psychodynamic psychotherapy, interpersonal psychotherapy, group exercise, and behavioural couples therapy for depression (National Institute for Health and Care Excellence, 2022). Psychological interventions may be offered in combination with antidepressant medication depending on depression severity and patient preference (National Institute for Health and Care Excellence, 2022). For chronic depression, cognitive behavioural therapy and antidepressant medication are recommended (National Institute for Health and Care Excellence, 2022). For generalised anxiety, NICE recommends unguided or guided self-help, psychoeducation, cognitive behavioural therapy, or applied relaxation (National Institute for Health and Care Excellence, 2020). As in the case of depression, antidepressant medication may also be offered for anxiety (National Institute for Health and Care Excellence, 2020). In England, psychological therapy services (Improving Access to Psychological Therapies) offer these NICE-recommended psychological therapies in primary care settings. While numerous psychotherapies exist, the evidence suggests they show broadly comparable efficacy (Barth et al., 2013; Berg & Høie, 2010; Cuijpers et al., 2020; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Hunot, Churchill, Silva de Lima, & Teixeira, 2007).

While psychotherapies are effective, meta-analytic evidence suggests that they are not effective for everyone; roughly 30-60% of patients recover or improve after receiving psychotherapy for depression and anxiety disorders in randomised controlled trials (Cuijpers

et al., 2021; Hansen, Lambert, & Forman, 2002). In clinical practice settings such as IAPT, approximately 60% of patients improve, and 40-50% recover (Clark et al., 2018; NHS Digital, 2020). However, there is enormous variability in clinical outcomes between service providers in clinical practice, with recovery rates ranging from approximately 20% to 60% (Clark et al., 2018). Personalised medicine has gained traction in an attempt to understand variability in clinical outcomes and improve the overall response to treatments.

3.3 Precision/Personalised Medicine

3.3.1 Definition

Evidence-based medicine is the principle of using the best available empirical evidence to inform clinical decisions in healthcare (Pencina & Peterson, 2016). Traditionally, this evidence is derived from randomised controlled trials or a systematic synthesis of multiple randomised controlled trials (Ahuja, 2019; Akobeng, 2005). Randomised controlled trials are considered the gold standard of evidence as they contain a central element – randomisation (Akobeng, 2005). Patients who enter into randomised controlled trials are randomly allocated to receive an intervention of interest or into a control (or comparator) arm. After treatment, the average outcomes in the intervention group are compared to the control group. Randomisation allows for the inference of causality as, when completed successfully, potentially confounding variables are more likely to be distributed equally between the intervention and the control group (Akobeng, 2005). As such, any differences in outcome between the groups are more likely to result from the treatment effect than any confounding factors. On average, treatments are deemed effective if the intervention group has better outcomes than the control group. However, as previously noted, treatments are rarely 100% effective for all patients, and this approach has been criticised for taking a "one-size-fits-all" perspective (Pencina & Peterson, 2016). While more patients in the intervention group may have better outcomes, this is not the case for all patients.

There has been a growing interest in precision medicine in efforts to explore heterogeneity in treatment response and improve clinical outcomes for all patients. Various terms have been coined to define the field, including stratified, targeted, personalised, and precision medicine, to name a few. While some argue there are subtle differences between the terms, they fundamentally refer to very similar principles and are frequently used interchangeably in research, as for the purposes of this thesis. Precision medicine can be broadly defined as "the use of combined knowledge (genetics, or otherwise) about a person to predict disease

susceptibility, disease prognosis or treatment response and thereby improve that person's health" (Redekop & Mladsi, 2013, p.S5).

While the focus is on the individual, these approaches are not tailored to every specific patient but rather (to date) focus on the identification of a subset of characteristics/traits that may provide insights into which patients develop a disease, how a disease progresses in specific groups of patients, or who will respond best to which treatment (a stratified approach). Thus, going beyond considering only the broad phenotypic grouping of symptoms that traditional diagnostic manuals provide. Precision medicine approaches are person-centred by taking specific patient characteristics into account to maximise clinical outcomes for subgroups of individuals. Precision medicine appears to be becoming an international priority, with the Obama administration launching the Precision Medicine Initiative (Fox, 2015) and NHS England making it a priority in its Five Year Forward Plan (Keogh, 2015). However, the concept is not novel in medicine. It can perhaps be traced as far back as 1909 when Archibald Garrod's work focused on the individuality and variation in human chemical processes (Perlman & Govindaraju, 2016).

While the idea is not new, advances in technology have undoubtedly contributed to our ability to pursue any efforts of precision medicine. An initial focus of precision medicine was genetic data, facilitated by advancements such as the Human Genome Project (Brittain, Scott, & Thomas, 2017). Such developments have allowed greater transparency in the identification of subgroups of diseases and their underlying mechanisms amongst disorders which have similar manifestations (Redekop & Mladsi, 2013). Advancements in data capture and infrastructure, as well as greater computing power, have allowed increasingly multifaceted and more extensive data to be used in precision medicine, including (but not limited to) clinical, environmental, social, and biological characteristics.

3.3.2 Successful examples

Some successful advances in precision medicine using genetics have been discovered in oncology regarding disease susceptibility, prognosis, and treatment. Women with an inherited mutation in the breast cancer genes BRCA1 and BRCA2 have a higher likelihood of developing breast and ovarian cancer (King, Marks, & Mandell, 2003). It has been estimated that the lifetime risk of developing breast and ovarian cancer amongst those who carry the BRCA1 mutation is 80% and 23%, respectively. For women carrying a mutation in BRCA2,

the lifetime risk of ovarian cancer is approximately 50% (King, Marks & Mandell, 2003). In terms of treatment, results from the MINDAT trial suggested that amongst women with early-stage breast cancer who have high clinical risk but a low genomic risk (as identified via the 70-gene signature test), the 5-year rate of survival without distant metastasis did not meaningfully differ between those who received chemotherapy and those who did not (1.5% difference) (Cardoso et al., 2016). Thus, suggesting approximately 45% of women who would be considered to have a high clinical risk may not require chemotherapy as they potentially have a low genomic risk (Cardoso et al., 2016). A further example is phenylketonuria, a genetic disorder that causes a build-up of the amino acid phenylalanine in the blood and brain due to a deficiency in the enzyme that breaks down phenylalanine, resulting in irreversible brain damage (National Institute for Health and Care Excellence, 2021). Such clear-cut prognostic markers in oncology can ultimately contribute to clinical decision-making to maximise the patient's health, such as potentially not delivering chemotherapy to women with low genomic risk, given the substantial impact of chemotherapy but no substantial clinical benefit. For phenylketonuria, people with this condition must follow a strict diet that is limited in protein and aspartame as well as NICE recommending the medication sapropterin (a cofactor for the enzyme that breaks down phenylalanine) for those under the age of 21 and pregnant women (National Institute for Health and Care Excellence, 2021).

3.4 Traditional Approaches to Personalised Psychotherapy

3.4.1 Predictors and moderators of treatment outcomes

Efforts to better understand variation in treatment response in psychiatry, thereby potentially opening an avenue for more personalised treatments, have focused on examining predictors and moderators. Predictors (or prognostic variables) are characteristics that influence a patient's prognosis. These characteristics address the question of who is likely to get better and who may be less likely to get better. Predictors can broadly be classified into non-specific and specific. Non-specific predictors examine the prognosis regardless of the intervention (potentially including no intervention), whereas specific predictors examine the prognosis in response to a specific treatment. Moderators (or prescriptive variables) are characteristics that determine a differential response to two (or more) treatments. These characteristics are prescriptive in that they address the question of who responds better to which intervention. Baseline characteristics are often considered the most clinically valuable predictors and moderators, as these are known when clinical decisions about treatment allocation are made.

Traditionally, predictors and particularly moderators have been examined in randomised controlled trials. Due to the extensive literature on psychotherapy randomised controlled trials, many predictor and moderator analyses have been performed. While predictors are helpful in terms of identifying patients who might have a poor prognosis and require additional care, moderators are clinically most useful for the purposes of personalised psychotherapy. Moderators can support clinical decision-making by providing evidence-based recommendations on treatment allocation. The benefit is two-fold, ensuring patient outcomes are maximised as well as potentially increasing cost-effectiveness. Fewer resources, such as further treatment, will likely be required if patients are offered a treatment that works best for them in the first instance. Therefore, I focus on presenting a brief overview of meta-analyses that examine baseline moderators of treatment outcomes in psychotherapy for depression and GAD below. The methods for study identification for the overview can be found in supplementary material A.

3.4.2 An overview of meta-analyses of moderators of treatment outcomes for psychotherapy

Table 1. *Characteristics of meta-analyses examining moderators for psychotherapeutic interventions included in the narrative overview*

#	Reference	Intervention	Comparator	Studies (n)	Sample size (n)
DEPRESSION					
1	(Cuijpers, Weitz, et al., 2014)	Cognitive behavioural therapy	Pharmacotherapy	13	1766
2	(Weitz et al., 2015)	Cognitive behavioural therapy	Pharmacotherapy	16	1700
3	(Santoft et al., 2019)	Cognitive behavioural therapy	Control/ pharmacotherapy/ psychotherapy	34	5358
4	(Karyotaki et al., 2017)	Web-delivered cognitive behavioural therapy	Control	13	3876
5	(Karyotaki et al., 2018)	Guided internet-based interventions	Control	24	4889
6	(Cuijpers et al., 2012)	Pharmacotherapy or Psychotherapy	Psychotherapy/ combined	52	4734

7	(Cuijpers, Ebert, Acarturk, Andersson, & Cristea, 2016)	Psychotherapy	Any psychotherapy	41	2741
8	(Driessen, Cuijpers, Hollon, & Dekker, 2010)	Psychotherapy	Control	132	10,134
9	(Gould, Coulson, & Howard, 2012)	Cognitive behavioural therapy	Control/ pharmacotherapy/ psychotherapy	23	NR
10	(Braun, Gregor, & Tran, 2013)	Psychotherapy	Psychotherapy	53	3,965
11	(Cuijpers, Koole, et al., 2014)	Psychotherapy	Control	74	Subclinical: 1913 & MDD: NR
12	(Gellatly et al., 2007)	Self-help	Control	34	NR
13	(Cuijpers et al., 2008)	Psychotherapy	Psychotherapy	91	NR

14	(Barth et al., 2013)	Psychotherapy	NR	198	NR
ANXIETY					
15	(Kishita & Laidlaw, 2017)	Cognitive behavioural therapy	Control	15	770
16	(Cuijpers, Sijbrandij, et al., 2014)	Psychotherapy	Control/ pharmacotherapy/ psychotherapy	41	2132
17	(Hanrahan, Field, Jones, & Davey, 2013)	Cognitive Therapy	Control/psychotherapy	15	907
18	(Carl et al., 2020)	Psychotherapy or pharmacotherapy	Control	79	11,002
DEPRESSION & ANXIETY					
19	(Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016)	Cognitive behavioural therapy	Control	144	NR

**MDD: Major Depressive Disorder, GAD: Generalised Anxiety Disorder, PAD: Panic Disorder, SAD: Social Anxiety Disorder*

I present an overview of nineteen relevant meta-analyses that examined moderators of treatment outcomes in psychotherapy vs. other psychotherapies, pharmacotherapy, or control groups (Table 1). Of the 19 meta-analyses, 14 focused on depression, four specifically on GAD and one examined both. Multiple characteristics were only examined by one or two meta-analyses, including ethnicity, employment status, education, previous episodes of depression and anxiety, medication status, and physical health comorbidities (Table 2). Thus, providing limited confidence due to a lack of consistently replicated findings across meta-analyses. Some of the potentially effect modifying characteristics were examined by multiple meta-analyses. Here, the evidence seemed somewhat conflicting, leaving almost no unambiguous characteristics when interventions were considered collectively. Diagnosis was the only variable examined numerous times as a moderator, where no evidence of effect modification was found across the board. Fewer meta-analyses (four) focused specifically on GAD or had a subgroup analysis on GAD (one). Much of the previous literature using meta-analytic techniques have examined a range of anxiety disorders combined, despite their presentation varying and constituting distinct diagnoses (Haug, Nordgreen, Ost, & Havik, 2012; Keefe, McCarthy, Dinger, Zilcha-Mano, & Barber, 2014; Pearl & Norton, 2017; Springer, Levy, & Tolin, 2018). Overall, no clear, overwhelming evidence indicated a given characteristic was a moderator (Table 2). However, it should be noted that the included evidence is unlikely to be exhaustive. Furthermore, the meta-analyses are likely to contain overlapping randomised controlled trials and are therefore not independent of one another.

Table 2. *Summary of moderators of clinical outcomes in psychotherapy for depression and anxiety derived from meta-analyses*

Variable	Potential Moderator	No Difference
Age	5,6,7,18	4,6,7,8,9,10,15,16,17
Gender	8,10	1,4,5,18
Ethnicity	5	
Employment status		4,5
Education		4,5
Relationship status	5	4,5
Settings	3,6,8,12	5,6,8,9,10,13,16
Baseline severity	5,8,11	2,3,4,5,10,18
Diagnosis		3,5,8,9,10,12,13,14
Previous episodes		5
Medication	9	5
Target population	7,8	5,8,10,11,13,14, 19
Depression type	6	6,8
Physical health comorbidity		7, 18
Mental health comorbidity	7,10	4,5,18

**Numbers denote study number from table 1.*

3.4.2.1 Heterogeneity

Identifying clear moderators is also complicated by the heterogeneity of randomised controlled trials included within the meta-analyses. While particular moderators were examined across multiple meta-analyses, such as age, these were often explored amongst different interventions and comparators. Greater confidence in potential moderators may be expected if the same moderator was found to have a similar effect across multiple meta-analyses when the same intervention and comparator are examined, as moderating effects may be expected to differ for different treatments. Effect sizes are most commonly used in meta-analyses. Effect sizes are helpful as they provide a standardised unit of measurement

that can be easily compared. However, they are challenging to interpret and provide little information on the clinical importance of improvements (Cuijpers, Karyotaki, et al., 2014). Some meta-analyses report other outcome metrics such as response, remission, or post-treatment scores, which place a greater focus on clinically relevant improvements rather than statistical change alone. While all outcome metrics attempt to examine clinical improvements, they measure success differently. Different outcome metrics may therefore also partially contribute to mixed findings. In addition, the outcome metrics are based on different outcome questionnaires, which include self-report questionnaires, clinician-reported questionnaires, or a composite of the two. Some evidence suggests that self-report questionnaires are associated with smaller effect sizes (Cuijpers, Sijbrandij, et al., 2014). Furthermore, the methods of analyses varied for randomised controlled trials, with some reporting primarily intention-to-treat analyses and others providing complete-case analyses. As such, meta-analyses relied on the provided analysis and combined different methods of analysis (Cuijpers, Sijbrandij, et al., 2014).

3.4.2.2 The usefulness of comparators in moderator analyses

Almost all meta-analyses performed moderator analyses on randomised controlled trials that compared an intervention to control conditions composed of inactive controls (such as a waitlist control) as well as potentially active controls (such as treatment as usual, TAU). While moderator analyses comparing interventions to inactive control groups are useful in terms of identifying potential sources of heterogeneity between trials as well as identifying groups where specific treatments may be particularly useful, they lack the ability to be prescriptive and offer a meaningful alternative for treatment. As such, they are more akin to predictors of treatment. Patients who do not fall into specific subgroups that may benefit more from the examined treatment will still receive treatment in clinical practice. It could be considered unethical to withhold treatment where interventions are generally regarded as effective. More clinically meaningful or practical moderators are those that directly compare interventions. Potentially active control groups such as TAU are slightly more prescriptive. However, they have a caveat of ambiguity when no clear measurement and specification of what TAU entails is provided. Research has shown that TAU in depression and anxiety trials varies considerably, ranging from psychotherapy and/or pharmacotherapy to minimal contact interventions such as monitoring and information provision (Watts, Turnell, Kladnitski, Newby, & Andrews, 2015). Unfortunately, the present meta-analyses did not investigate

moderators of treatments for each type of control group used but rather combine all control conditions – potentially active and inactive - into one.

3.4.2.3 Methodological Rigour

Various methods have been used to examine potential moderators in the reviewed meta-analyses. The formal method to examine treatment moderators requires a test for interaction between a potential moderator and treatment to formally test whether clinical outcomes in different treatments vary by a particular characteristic. An interaction term is considered a hallmark criterion of rigour when examining moderators (Pincus et al., 2011). This can be done at the study level or individual patient-level (Karyotaki et al., 2018). However, other methods have been reported in meta-analyses to identify potential moderators. One method is to examine the efficacy of two interventions in a specific and targeted population (Cuijpers, Ebert, et al., 2016; Cuijpers et al., 2012). For example, examining the efficacy of cognitive behavioural therapy vs. pharmacotherapy amongst older adults speaks to a potential moderating effect of age in these treatments but lacks a formal comparison with other age groups such as adult patients. A further method reported is to examine potential moderators as a predictor of a difference (Driessen et al., 2010). However, the implications of using different methods to examine effect modification and how these contribute to mixed findings are still unclear and remain to be formally evaluated.

3.4.2.4 Statistical Power

Randomised controlled trials are designed to evaluate the efficacy or effectiveness of treatments and are therefore powered for such analyses. Moderators of treatment all implicitly or explicitly examine interactions between characteristics and treatment. Such tests have significantly lower statistical power than when a main effect is examined alone, as in an evaluation of efficacy (Greenland, 1983). Estimates from the literature suggest that sample sizes should be 4-fold to 16-fold of those required to test for main effects to have sufficient power for interactions (Gelman, 2018; Leon & Heo, 2009). In more practical terms, the evidence suggests that prescriptive models should, subject to assumptions, preferably have at least 500 patients per treatment arm in depression (Luedtke, Sadikova, & Kessler, 2019).

Despite the meta-analyses containing much larger samples than individual randomised controlled trials, statistical power may still be an issue. Cuijpers and colleagues report that power is typically small to detect differences between subgroups of patients (Cuijpers,

Sijbrandij, et al., 2014). Some characteristics are rarer and therefore have small numbers of patients/studies investigating these. As such, some effects being based on single studies may have contributed to a difficulty in identifying between-group differences (Cuijpers, Ebert, et al., 2016). Equally, where study-level data is used, there is a reliance on mean values, which may not sufficiently capture the variability in the baseline characteristics (Driessen et al., 2010). Furthermore, there can be considerable associations between different characteristics (Cuijpers, Sijbrandij, et al., 2014). As such, moderator analyses are unstable and susceptible to the addition or removal of one study and should always be considered with caution (Cuijpers, Sijbrandij, et al., 2014).

The issue of statistical power appears particularly pertinent to psychotherapy. In most instances, there were too few randomised controlled trials to directly compare two specific psychotherapies, which would be of most benefit in informing personalised treatment allocation in psychological therapy services. Where psychotherapies were compared, it was most frequently reported as a specific intervention, such as cognitive behavioural therapy, being compared to all other psychotherapies rather than a comparison being drawn to another specific psychotherapy (Braun et al., 2013; Cuijpers, Ebert, et al., 2016). This may obscure the benefits of a given psychotherapy over another, assuming that different psychotherapies have specific components. Furthermore, it provides limited clinical utility in terms of treatment allocation if the effect modification does not favour the particular psychotherapy being investigated, as the question of which treatment the patient should be assigned to remains unanswered. This may partially reflect an inherent difficulty in conducting psychotherapy trials, particularly where two or more psychotherapies are compared. Due to the nature of psychotherapy, conducting large-scale trials comparing multiple psychotherapies requires considerable effort, funding, and time.

Few meta-analyses examined statistical power formally. However, Cuijpers and colleagues performed power analyses in two meta-analyses to examine the confidence in potential moderators. In 2016, Cuijpers and colleagues reported that insufficient studies where two psychotherapies are compared had been completed to have sufficient statistical power to identify moderating effects of 27 potential characteristics (Cuijpers, Ebert, et al., 2016). Only approximately 6% of studies that would be required to find moderate effect sizes ($g = 0.50$), and 2% of studies to find smaller effect sizes ($g = 0.24$), had been conducted to date (Cuijpers, Ebert, et al., 2016). In terms of time, it was estimated that it would take

approximately 326 years to find medium effect sizes and 1372 years to find smaller effect sizes if studies are conducted at the same pace as they have been conducted to date (Cuijpers, Ebert, et al., 2016). Collectively, this highlights that more powerful alternatives are needed for precision psychiatry, particularly in the context of comparing psychotherapies with each other (Cuijpers, Ebert, et al., 2016).

This also highlights the difference between psychiatry and the previous cancer example of the BRCA1 and 2 genes. In the latter, the presence and absence of a gene variant strongly dictates prognosis and treatment response. Much smaller samples are required to find such strong effect sizes. Effects in psychiatry appear to be more subtle. There is no substantial evidence to suggest that an individual characteristic by itself appears to produce a strong differential effect on treatment outcomes which is also consistently found across studies. The absence of strong evidence in current meta-analyses suggests that any possible moderating effects may be much smaller. As such, more nuanced approaches to personalised psychotherapy may be required.

3.4.2.5 Quality

A further concern that limits the conclusions that can be drawn from the current evidence is the quality of research—15 out of the 20 meta-analyses reported at least some concern regarding the quality of the randomised controlled trials. For example, Santoft and colleagues reported that only 9% of included randomised controlled trials had a low risk of bias (Santoft et al., 2019). There is some evidence to suggest that study quality can affect effect-estimates in randomised controlled trials, with lower quality studies appearing to overestimate treatment effects (Hempel et al., 2011). Additionally, 10 out of 17 studies that reported a test for publication bias found evidence of a potential publication bias.

3.4.2.6 Generalisability

Randomised controlled trials are considered the gold standard of clinical research and are paramount to evaluating the efficacy and safety of treatments (Booth & Tannock, 2014). Well-conducted trials minimise bias such as confounding and can therefore address research questions in a causal manner (Booth & Tannock, 2014). Despite their inherent benefits, randomised controlled trials nonetheless have limitations. Randomised controlled trials are often conducted in stringent settings to achieve high internal validity (Booth & Tannock, 2014). Randomised controlled trials commonly have very specific inclusion criteria, with

treatment delivered in a consistent and protocolised manner. These stringent contexts are often not representative of real-world settings, both in terms of patients who are included in randomised controlled trials as well as how the treatment is delivered, making generalisability to wider contexts in real-world practice somewhat limited (Booth & Tannock, 2014; Kennedy-Martin, Curtis, Faries, Robinson, & Johnston, 2015).

3.4.2.7 Clinical Utility

Traditional approaches, such as those examined in the present overview, typically examine moderators in isolation. While these investigations are clearly worthwhile, taking such an approach has limitations in terms of clinical utility – it remains unclear how potential moderators may interact with one another. If reliable moderators were identified, the influence that other patient characteristics would have on them remains unknown. However, the overall evidence suggests that there are no clear-cut moderators that strongly dictate treatment response in psychiatry, as may be the case with the oncology example described previously. More nuanced approaches in larger datasets may be required for personalised psychotherapy to identify more subtle and/or complex effects.

3.5 Improving Access to Psychological Therapies (IAPT) Programme

IAPT is an English public-health initiative set up to provide greater access to evidence-based psychological interventions for depression and anxiety in primary care settings (Clark, 2011). The electronic healthcare records recorded as part of routine clinical practice provide an invaluable resource for research with information on patients, their treatment, and clinical outcomes. Datasets of such scale containing information on psychological treatment are difficult to find elsewhere. Below I briefly describe the overall model and structure of IAPT and its origins to provide context and historical background.

3.5.1 Overview

NICE provide evidence-based guidelines on treatment. NICE reviews the available evidence commonly consisting of randomised controlled trials to make clinical recommendations (Rawlins, 2015). IAPT offer a range of NICE-recommended interventions based on a stepped-care model (Clark, 2011). Stepped-care models are based on the principle that the least restrictive interventions are the first line of treatment (Clark, 2018). In IAPT, low-intensity therapy (LIT, often referred to as step 2) is delivered in the first instance (Clark, 2018). High-intensity therapy (HIT, often referred to as step 3) is delivered where there is an

insufficient response to LIT or where there is a clinical need for HIT, such as higher symptom severity (Clark, 2018).

LITs are usually shorter, less resource-intensive therapies, including therapist-guided or unguided self-help interventions (National Institute for Health and Care Excellence, 2011). HITs are commonly longer, more resource-intensive therapies. While an initial focus in IAPT was placed on cognitive behavioural therapy, the choice of psychological therapies has expanded over time (Clark, 2011). For example, a report by the British Association for Counselling and Psychotherapy showed that amongst 114 audited services, 95% of services offered cognitive behavioural therapy, 92% offered counselling, 13% offered brief psychodynamic psychotherapy, 36% offered interpersonal psychotherapy, and 47% offered couples therapy (Perfect, Jackson, Pybis, & Hill, 2016). However, the availability of psychological therapies per service varies – 13% of services offered only one HIT, and only 0.9% of the services provided all five therapies (Perfect et al., 2016).

3.5.2 Rationale for IAPT

The benefits of offering increased access to psychological therapies were twofold – humanitarian and economic (Layard, Clark, Knapp, & Mayraz, 2007). There is a clear humanitarian argument for providing adequate provision of mental healthcare. Medication was the primary form of treatment offered to those experiencing common mental health problems, such as depression and anxiety, in the community, with only a small proportion of patients receiving psychological interventions at the time (Layard et al., 2007; McManus, Meltzer, Brugha, Bebbington, & Jenkins, 2009). Yet, there is a 3-fold patient preference for psychological interventions compared to medication, with a large proportion of those preferring psychological therapies opting for no treatment rather than medication (Layard, 2006; McHugh et al., 2013). Given the established efficacy of psychological interventions, Layard and colleagues described this discrepancy between the treatment that was delivered and the patient preference as the most significant gap between best and actual practice in the NHS (Layard et al., 2007).

There was also an economic argument for increasing access to psychological therapies in England, predominantly driven by Richard Layard and David Clark (Clark, 2018). Treatment for depression and anxiety would not only reduce the substantial suffering to individuals, but it would also result in a reduction of public costs and an increase in revenue via means of

reducing incapacity benefits and medical costs as well as increased economic activity and productivity (Clark, 2018; Layard et al., 2007). As such, the original argument stated that increasing access to psychological therapies would pay for itself (Layard et al., 2007). The cost of cognitive behavioural therapy was estimated as a one-off of approximately £750 per treatment episode (Layard et al., 2007). Whereas the savings to society were estimated to be £4,700 and the savings to the Exchequer were estimated to be £1,200 (Layard et al., 2007). As such, any initiative to increase psychological therapy was argued to come at no net cost - the estimated savings would far exceed the initial expenditure (cost of treatment).

3.5.3 Development of IAPT over time

Given the accumulating evidence for a benefit of increased access to psychological interventions, the Department of Health funded the first implementation of the IAPT initiative at two primary care trust demonstration sites in 2006 (Clark et al., 2009). They received £1.3-1.5 million in funding to implement these demonstration sites, offering NICE-recommended treatment in the form of a stepped-care model (Clark et al., 2009). In 2008, following the successful completion of the pilots, the Department of Health announced the National Implementation Plan to roll IAPT out nationally (Department of Health and Social Care, 2008).

A large part of the implementation programme concerned the training of a new workforce. The primary barrier to implementing the NICE-recommended psychological therapies was insufficient qualified mental health professionals (Clark, 2018). The aim was to train 10,500 new therapists by 2021 and deploy these in services to meet the demand for psychological therapy (Clark, 2018). The IAPT workforce comprises low-intensity practitioners, called psychological wellbeing practitioners (PWPs), who deliver LITs (Clark, 2018). Furthermore, there are high-intensity practitioners such as high-intensity trained therapists, counsellors, or clinical psychologists, who deliver HITs (Clark, 2018).

In 2008/2009, the first year of the IAPT initiative, 35 services were established (Gyani, Shafran, Layard, & Clark, 2013). The expansion of services was rapid, with over 150 services being set up in the first three years (Gyani et al., 2013). By 2019/20, IAPT services were available nationally, receiving 1.69 million referrals annually (NHS Digital, 2020). There is a continued commitment to expanding the capacity of IAPT with the NHS Mental Health Implementation Plan aiming to increase access to 1.9 million adults and older adults by

2023/24 (NHS England, 2019). While there is an inherent difficulty in finding reliable published figures for the total funding to date, estimates suggest that the total cost of the IAPT initiative up to 2018 was approximately £1 billion (Marks, 2018).

3.5.4 Evaluating the impact of IAPT

In efforts to evaluate the success of IAPT, a distinctive feature of routine data collection was introduced into IAPT. Detailed records of patient characteristics, their treatment, and their clinical outcomes are recorded. Clinical outcome questionnaires are recorded on a session-by-session basis (Clark et al., 2018). This is preferable to pre- and post-treatment measures being taken as it increases complete-case recording of clinical progress when patients drop out of treatment. Due to this session-by-session monitoring of outcomes, approximately 98% of patients have a pre- and post-treatment score in IAPT (Clark et al., 2018). The standard measure of depression in IAPT is the 9-item Patient Health Questionnaire (PHQ-9) and the 7-item Generalised Anxiety Disorder Scale (GAD-7) is used for GAD or general anxiety (Johnson, Ulvenes, Oktedalen, & Hoffart, 2019; Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, Williams, & Lowe, 2006). However, disorder-related outcome questionnaires are used for specific anxiety disorders such as social anxiety, obsessive-compulsive disorder, health anxiety, panic disorder, and post-traumatic stress disorder. Summary statistics of IAPT are published nationally on a monthly basis (Clark et al., 2018).

IAPT measures its success in terms of three outcome metrics: recovery, reliable change, and reliable recovery (National Collaborating Centre for Mental Health, 2019). Recovery is assessed in terms of caseness. Patients are defined as being at caseness if either their depression and/or anxiety scores are at or above a clinical threshold at baseline (National Collaborating Centre for Mental Health, 2019). For example, these thresholds are 10 and 8 points for the PHQ-9 and GAD-7 (National Collaborating Centre for Mental Health, 2019). Patients who are at caseness are counted as being recovered if both their depression and anxiety measures are below these clinical thresholds. Reliable change is measured by the Jacobsen and Traux index (Jacobson & Truax, 1991). Patients are counted as having achieved reliable change if their pre-post treatment change is equal to or greater than the measurement error for either their depression or anxiety questionnaires (National Collaborating Centre for Mental Health, 2019). For example, this threshold for change is 6 and 4 points for the PHQ-9 and GAD-7, respectively. Patients are counted as reliably recovered if they achieve recovery and reliable change metrics (National Collaborating Centre for Mental Health, 2019).

3.5.5 Limitations of outcome metrics in IAPT

While IAPT is one of the first of its kind to measure outcomes routinely and report these nationally, the outcome metrics have limitations. Recovery metrics can only be applied to patients who are at caseness; however, IAPT's remit is to treat mild to moderate depression and anxiety. As such, measuring recovery amongst patients with milder symptoms is impossible. An additional limitation of threshold approaches, such as the recovery metric, is that patients who score very close to the threshold at baseline may only require a single point change to fall below it. This, however, may be insufficient for clinically meaningful improvements. Conversely, patients with high baseline scores may experience a large drop in scores that may not quite reach the threshold. As such, patients may not be classed as recovered; however, they may nonetheless have substantially benefited from the intervention. Furthermore, threshold approaches such as recovery do not account for baseline dependency (Button et al., 2015). Baseline dependency is the phenomenon where patients with higher baseline severity require larger changes to feel better. Reliable change addresses some of the limitations of recovery metrics, such as the issue of patients falling very close to or far away from thresholds at baseline by measuring symptom change. Such metrics are more frequently used to measure clinically meaningful change. However, the reliable change metric entails a change that exceeds the measurement error and is therefore derived from the statistical spread of the data. This approach ignores patients' perceptions which are arguably imperative to clinically meaningful improvements, particularly where subjective experiences such as depression and anxiety are being targeted (McGlothlin & Lewis, 2014). Furthermore, similar to recovery, the reliable change metric does not account for baseline dependency. As such, there is a lack of outcome metric that captures clinical change that is meaningful to patients and captures baseline dependency. My first study, therefore, focuses on developing a metric that potentially addresses some of these limitations.

3.5.6 Clinical Outcomes in IAPT

Despite the limitations of the outcome metrics, routine data recording and national reporting are invaluable for assessing clinical outcomes nationally. In efforts to examine variation in clinical outcomes, Clark and colleagues conducted a large-scale analysis of IAPT outcomes at a national level using data from the financial years 2014-16 (Clark et al., 2018). The data showed that approximately 60% of patients improve, and approximately 40% recover. However, variation between services was considerable, with recovery rates ranging from

approximately 20% to 60% and reliable improvement ranging from approximately 30% to 80% (Clark et al., 2018). More recently, in the financial year 2019/20, recovery rates were at 51%, and reliable improvement rates were at 67% (NHS Digital, 2020). This evidence suggests there is room for improvement in clinical outcomes, to which personalised psychotherapy may be able to contribute.

3.6 More Recent Approaches to Personalised Psychotherapy

There has been a growing interest in personalised medicine to improve clinical outcomes. More recent developments in personalised psychotherapy have moved beyond the traditional approach of examining individual predictors or moderators in isolation, taking more comprehensive and varied approaches. The routine recording of electronic clinical records and the rapid expansion of IAPT have partially contributed to this through the availability of large-scale psychotherapy datasets, which have more power to explore other approaches.

3.6.1 Prediction

Some research in IAPT has focused on examining multiple baseline characteristics to create a composite risk score, as demonstrated by Delgadillo and colleagues, who developed the Leeds Risk Index (LRI) (Delgadillo, Moreea, & Lutz, 2016). The LRI is a patient profiling tool that identifies patients who are potentially at risk of poor outcomes (Delgadillo et al., 2016). The evidence suggested that disability, unemployment, younger age, higher functional impairment, higher baseline depression, and lower outcome expectancy are associated with worse prognosis after therapy (Delgadillo et al., 2016). Specifically, amongst patients with a low-risk index, approximately 55% showed a clinically relevant improvement, whereas only about 25% of patients with a high-risk index showed improvements (Delgadillo et al., 2016). An extension of examining baseline characteristics to predict clinical outcomes was examined by Bone and colleagues, who developed the oracle algorithm (Bone et al., 2021).

The oracle algorithm goes beyond traditional, static prediction models, which rely on baseline data to predict clinical outcomes to incorporate information collected throughout treatment (Bone et al., 2021). Thereby providing a dynamic model that updates prognosis as patients progress through treatment. Empirical evidence suggests that the oracle algorithm accounts for approximately 38 - 47% of the variability in treatment outcomes and has a high prediction accuracy by seven appointments (Bone et al., 2021). Such approaches give insight into patient prognoses, which can provide clinicians with evidence of which patients are

(un)likely to respond, allowing for appropriate action to be taken at the earliest possible stage, such as switching treatment or providing augmentation with other interventions. However, focusing solely on prediction has limitations as it is potentially possible to identify patients who may be less likely to respond. Still, prediction does not indicate whether a different treatment would work better or if augmentation with another intervention would lead to better outcomes.

3.6.2 Differential treatment allocation

Compared to prediction alone, moderator analyses can provide information on whether differential allocation to one treatment over another may result in better outcomes. Compared to traditional moderator analyses, a greater focus has been placed on models that combine multiple variables that examine both predictors and moderators to predict differential response to treatment (multivariable prescriptive models). These take more of an actuarial risk approach where various variables are used to build a more comprehensive clinical prediction for different treatments. Such an example is the Personalised Advantage Index (PAI), which uses multiple characteristics to identify the benefit of one intervention over another (DeRubeis et al., 2014). For example, DeRubeis and colleagues showed that when comparing cognitive behavioural therapy to pharmacotherapy for depression, approximately 60% of patients benefitted from applying a PAI (DeRubeis et al., 2014). Amongst the patients, those who received the model-indicated optimal treatment scored 3.6 points lower on the Hamilton Rating Depression Scale after treatment compared to those who received the model-indicated suboptimal treatment (Hamilton, 1960). While first developed in randomised controlled trials where power is lower, such approaches have been applied in IAPT data, such as by Deisenhofer and colleagues. They developed a prescriptive algorithm to support clinician decision-making in allocating patients to trauma-focused cognitive behavioural therapy or Eye-Movement Desensitization and Reprocessing using the PAI (Deisenhofer et al., 2018). The results showed that 63% of patients who were assigned to their optimal treatment showed clinically meaningful improvements, whereas only 34% of patients improved when assigned to their suboptimal treatment (Deisenhofer et al., 2018). The focus of my second study will adopt a similar approach, exploring the benefit of an actuarial model for differential treatment allocation in depression.

3.6.3 Use of longitudinal data

Prediction and moderation frequently focus on baseline characteristics as this provides information to clinicians at the beginning of treatment. However, other approaches have taken advantage of longitudinal data. For example, Saunders and colleagues examined different groups of response trajectories over therapy and patient characteristics associated with these (Saunders et al., 2019). They identified four distinct trajectories for depression and five for anxiety (Saunders et al., 2019). Most patients' clinical trajectories could likely be identified by the third appointment, but one group showed a slow initial response with a later rapid improvement after the sixth appointment (Saunders et al., 2019). Lower baseline severity of symptoms, higher social functioning, and the absence of phobic anxiety were associated with better response trajectories (Saunders et al., 2019). Alternatively, Robinson and colleagues investigated the dose-response relationship of therapy for patients with different clinical presentations (Robinson, Kellett, & Delgadillo, 2020). They found that 95% of patients improve within seven appointments in LIT and 14 appointments in HIT but that patients with post-traumatic stress disorder, social anxiety disorder, and obsessive-compulsive disorder require higher doses of treatment (Robinson et al., 2020). Such approaches use more rich data collected across the course of therapy to understand heterogeneity in response further and inform clinical decision-making.

3.6.4 Symptom heterogeneity

While a significant focus has been placed on patient characteristics and what occurs during treatment, a different approach has also gained traction recently. Here, researchers focus on the heterogeneity of the disorders themselves. Both depression and anxiety are heterogenous disorders, with their mechanisms being poorly understood. For example, the DSM criteria for depression are composed of nine symptoms, of which two are core symptoms; this does not include various non-DSM symptoms that are often considered part of depression (American Psychiatric Association, 2013). As such, depression can present quite differently between individual patients. Researchers examined symptom heterogeneity between patients and found approximately 1000 unique symptom profiles in one study alone, with the most common profile being shared by only approximately 2% of patients (Fried & Nesse, 2015a). Furthermore, symptoms appear to have different risk factors, different associations with functional impairment, and different biological underpinnings (Fried & Nesse, 2015b). These findings align with opinions that question the validity of diagnoses and focus more on transdiagnostic approaches (Dalgleish, Black, Johnston, & Bevan, 2020). Such perspectives

have implications for psychotherapy research. According to the traditional perspective, sum scores have been used in outcome research to measure depression severity where individual symptoms are interchangeable. However, researchers have recently begun examining psychotherapy's effects on specific symptoms. This not only has the potential to explain some of the heterogeneity of clinical outcomes and potentially begin to shed light on the poorly understood mechanisms of how psychotherapy works but also provides an avenue to tailor treatments to clinical presentations of individual patients. For example, Boschloo and colleagues examined the symptom-specific effects of an internet-delivered depression intervention. They found it to have direct effects on sleeping problems, feelings of guilt/worthlessness, difficulty concentrating, and tiredness/fatigue (Boschloo et al., 2019). A better understanding of the symptom specific effects of treatments could lead to more personalised psychotherapy as therapies could be matched to patients based on their symptom profile and the symptoms that are most pertinent to patients. However, this research has been completed in randomised controlled trials to date. Similar to moderator analyses, randomised controlled trials are powered to examine the efficacy of treatment for overall depression severity, not individual symptoms. To my knowledge, research has yet to be applied in datasets with larger samples and greater statistical power. My third study takes advantage of longitudinal data, as discussed in point 3.6.3, and adopts an approach which focuses on looking at how individual symptoms of depression and anxiety behave during treatment with cognitive behavioural therapy.

Overall, the above evidence suggests some potentially promising developments have been made in personalised psychotherapy. IAPT data has been an invaluable resource to many of these developments by providing a large-scale resource of psychotherapy data. However, some approaches have yet to benefit from the availability of larger patient datasets, such as IAPT data.

3.7 Aims and Objectives

The present PhD aims to explore whether the clinical response in treating depression and anxiety can be improved through personalised psychotherapy. Exploratory analyses focused on applying statistical models to secondary data from randomised controlled trials and electronic health care records to better understand how to measure clinically meaningful change, whether stratified approaches to treatment selection can enhance clinical outcomes,

and exploring variation in treatment response in more detail by looking at symptom-specific changes. Specifically, the aims of this PhD were:

1. To develop an outcome metric for changes on clinical questionnaires which is rooted in the patient experience (rather than relying on statistical change alone) as well as accounting for baseline dependency, given its importance in quantifying subjective improvements.
    a. To achieve this, a method of estimating the minimal clinically important difference or MCID was developed. The MCID is the smallest difference that is perceivable by patients. The study's approach conceptualised the MCID as the effective dose 50 (ED50). The ED50 represents a change in clinical questionnaires where there is a 50% probability of patients feeling better. Data from two high-quality randomised controlled trials were used to map subjective patient improvements onto changes in questionnaire scores for two of the most widely used questionnaires for depression and anxiety. From this model, the ED50 was personalised to each level of baseline severity.

2. To explore whether a prescriptive algorithm based on baseline patient characteristics can support clinical decision-making by showing who may benefit most from which therapy before patients start treatment for depression.
    a. To achieve this, secondary data from IAPT was used to generate a retrospective cohort of patients receiving two of the most widely delivered therapies for depression in IAPT - cognitive behavioural therapy and counselling for depression. An actuarial approach was taken to predict differential response between therapies based on multiple patient baseline characteristics and evaluated in a held-out dataset.

3. To explore the symptom-specific effects of cognitive behavioural therapy for depression and anxiety. With most previous research focusing on sum scores of depression and anxiety measures, the present approach aims to examine and understand a more nuanced response to treatment.
    a. To achieve this, a retrospective cohort of patients receiving cognitive behavioural therapy in IAPT was generated. Trajectories of individual

depression and anxiety symptoms were modelled across the course of cognitive behavioural therapy and compared relative to one another.

4. COVID-19 note – during my PhD, the COVID-19 pandemic occurred. As such, I had the timely opportunity to examine how access to mental health services changed and how services responded and adapted their service provision to the pandemic.
    a. To achieve this, a retrospective cohort of all referrals and appointments before and during the early stages of the pandemic was generated. I used descriptive analyses to examine the changes in overall referrals to IAPT and the sociodemographic characteristics of these referrals. Furthermore, I examined how the overall number of appointments changed and the medium through which these appointments were delivered.

3.8 References

Ahuja, A. S. (2019). Should RCT's be used as the gold standard for evidence based medicine? *Integrative Medicine Research, 8*(1), 31-32. https://doi.org/10.1016/j.imr.2019.01.001

Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood, 90*(8), 840-844. https://doi.org/10.1136/adc.2004.058222

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Barth, J., Munder, T., Gerger, H., Nuesch, E., Trelle, S., Znoj, H., . . . 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

Berg, R. C., & Høie, B. (2010). Effectiveness of psychotherapy for adults with depression: a systematic review of the best available evidence. *Procedia - Social and Behavioral Sciences, 5*, 2194-2200. https://doi.org/10.1016/j.sbspro.2010.07.435

Bone, C., Simmonds-Buckley, M., Thwaites, R., Sandford, D., Merzhvynska, M., Rubel, J., . . . Delgadillo, J. (2021). Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. *The Lancet Digital Health, 3*(4), e231-e240. https://doi.org/10.1016/S2589-7500(21)00018-2

Booth, C. M., & Tannock, I. F. (2014). Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *British Journal of Cancer, 110*(3), 551-555. https://doi.org/10.1038/bjc.2013.725

Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., & Klein, J. P. (2019). Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: The potential of network estimation techniques. *Behaviour Research and Therapy, 122*, 103440. https://doi.org/10.1016/j.brat.2019.103440

Braun, S. R., Gregor, B., & Tran, U. S. (2013). Comparing bona fide psychotherapies of depression in adults with two meta-analytical approaches. *PloS One, 8*(6), e68135. https://doi.org/10.1371/journal.pone.0068135

Brittain, H. K., Scott, R., & Thomas, E. (2017). The rise of the genome and personalised medicine. *Clinical Medicine, 17*(6), 545-551. https://doi.org/10.7861/clinmedicine.17-6-545

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., . . . Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychological Medicine, 45*(15), 3269-3279. https://doi.org/10.1017/s0033291715001270

Cardoso, F., van't Veer, L. J., Bogaerts, J., Slaets, L., Viale, G., Delaloge, S., . . . Piccart, M. (2016). 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *New England Journal of Medicine, 375*(8), 717-729. https://doi.org/10.1056/NEJMoa1602253

Carl, E., Witcraft, S. M., Kauffman, B. Y., Gillespie, E. M., Becker, E. S., Cuijpers, P., . . . Powers, M. B. (2020). Psychological and pharmacological treatments for generalized anxiety disorder (GAD): a meta-analysis of randomized controlled trials. *Cognitive Behaviour Therapy, 49*(1), 1-21. https://doi.org/10.1080/16506073.2018.1560358

Clark, D. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry, 23*(4), 318-327. https://doi.org/10.3109/09540261.2011.606803

Clark, D. (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., 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., 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

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., Ebert, D. D., Acarturk, C., Andersson, G., & Cristea, I. A. (2016). Personalized Psychotherapy for Adult Depression: A Meta-Analytic Review. *Behavior Therapy, 47*(6), 966-980. https://doi.org/10.1016/j.beth.2016.04.007

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*(3), 288-299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Karyotaki, E., de Wit, L., & Ebert, D. D. (2020). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research, 30*(3), 279-293. https://doi.org/10.1080/10503307.2019.1649732

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies 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., Koole, S. L., van Dijke, A., Roca, M., Li, J., & Reynolds, C. F. (2014). Psychotherapy for subclinical depression: meta-analysis. *British Journal of Psychiatry, 205*(4), 268-274. https://doi.org/10.1192/bjp.bp.113.138784

Cuijpers, P., Reynolds, C. F., 3rd, Donker, T., Li, J., Andersson, G., & Beekman, A. (2012). Personalized treatment of adult depression: medication, psychotherapy, or both? A systematic review. *Depression and Anxiety, 29*(10), 855-864. https://doi.org/10.1002/da.21985

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clinical Psychology Review, 34*(2), 130-140. https://doi.org/10.1016/j.cpr.2014.01.002

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology, 76*(6), 909-922. https://doi.org/10.1037/a0013075

Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., . . . Hollon, S. D. (2014). Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: an "individual patient data" meta-analysis. *Depression and Anxiety, 31*(11), 941-951. https://doi.org/10.1002/da.22328

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology, 88*(3), 179-195. https://doi.org/10.1037/ccp0000482

Deisenhofer, A.-K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic

stress disorder. *Depression and Anxiety, 35*(6), 541-550. https://doi.org/10.1002/da.22755

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79*, 15-22. https://doi.org/10.1016/j.brat.2016.02.003

Department of Health and Social Care. (2008). *Improving access to psychological therapies implementation plan: national guidelines for regional delivery*. https://www.mhinnovation.net/sites/default/files/downloads/innovation/tools/IAPT-Implementation-Plan.pdf

Department of Health and Social Care. (2011). *No health without mental health: A cross-Government mental health outcomes strategy for people of all ages*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/138253/dh_124058.pdf

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS One, 9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology, 78*(5), 668–680. https://doi.org/10.1037/a0020570

Fox, J. L. (2015). Obama catapults patient-empowered Precision Medicine. *Nature Biotechnology, 33*(4), 325. https://doi.org/10.1038/nbt0415-325

Fried, E. I., & Nesse, R. M. (2015a). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 172*, 96-102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., & Nesse, R. M. (2015b). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Medicine, 13*(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S., & Lovell, K. (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine, 37*(9), 1217-1228. https://doi.org/10.1017/S0033291707000062

Gelman, A. (2018). *You need 16 times the sample size to estimate an interaction than to estimate a main effect*. https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/

Gould, R. L., Coulson, M. C., & Howard, R. J. (2012). Cognitive behavioral therapy for depression in older people: a meta-analysis and meta-regression of randomized controlled trials. *Journal of the American Geriatrics Society, 60*(10), 1817-1830. https://doi.org/10.1111/j.1532-5415.2012.04166.x

Greenland, S. (1983). Tests for interaction in epidemiologic studies: a review and a study of power. *Statistics in Medicine, 2*(2), 243-251. https://doi.org/10.1002/sim.4780020219

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597-606. https://doi.org/10.1016/j.brat.2013.06.004

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56-62. https://doi.org/10.1136/jnnp.23.1.56

Hanrahan, F., Field, A. P., Jones, F. W., & Davey, G. C. (2013). A meta-analysis of cognitive therapy for worry in generalized anxiety disorder. *Clinical Psychology Review, 33*(1), 120-132. https://doi.org/10.1016/j.cpr.2012.10.008

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343. https://doi.org/10.1093/clipsy.9.3.329

Haug, T., Nordgreen, T., Ost, L. G., & Havik, O. E. (2012). Self-help treatment of anxiety disorders: a meta-analysis and meta-regression of effects and potential moderators. *Clinical Psychology Review, 32*(5), 425-445. https://doi.org/10.1016/j.cpr.2012.04.002

Hempel, S., Suttorp, M. J., Miles, J. N. V., Wang, Z., Maglione, M., Morton, S., . . . Shekelle, P. G. (2011). AHRQ Methods for Effective Health Care. In *Empirical Evidence of Associations Between Trial Quality and Effect Size*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Hunot, V., Churchill, R., Silva de Lima, M., & Teixeira, V. (2007). Psychological therapies for generalised anxiety disorder. *Cochrane Database of Systematic Reviews, 2007*(1), Cd001848. https://doi.org/10.1002/14651858.CD001848.pub4

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

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., . . . Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 392*(10159), 1789-1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Johnson, S. U., Ulvenes, P. G., Oktedalen, T., & Hoffart, A. (2019). Psychometric Properties of the General Anxiety Disorder 7-Item (GAD-7) Scale in a Heterogeneous Psychiatric Sample. *Frontiers in Psychology, 10*, 1713. https://doi.org/10.3389/fpsyg.2019.01713

Karyotaki, E., Ebert, D. D., Donkin, L., Riper, H., Twisk, J., Burger, S., . . . Cuijpers, P. (2018). Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. *Clinical Psychology Review, 63*, 80-92. https://doi.org/10.1016/j.cpr.2018.06.007

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., . . . Cuijpers, P. (2017). Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. *JAMA Psychiatry, 74*(4), 351-359. https://doi.org/10.1001/jamapsychiatry.2017.0044

Keefe, J. R., McCarthy, K. S., Dinger, U., Zilcha-Mano, S., & Barber, J. P. (2014). A meta-analytic review of psychodynamic therapies for anxiety disorders. *Clinical Psychology Review, 34*(4), 309-323. https://doi.org/10.1016/j.cpr.2014.03.004

Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials, 16*(1), 495. https://doi.org/10.1186/s13063-015-1023-4

Keogh, B. (2015). *Personalised Medicine Strategy*. https://www.england.nhs.uk/wp-content/uploads/2015/09/item5-board-29-09-15.pdf

King, M. C., Marks, J. H., & Mandell, J. B. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science, 302*(5645), 643-646. https://doi.org/10.1126/science.1088759

Kishita, N., & Laidlaw, K. (2017). Cognitive behaviour therapy for generalized anxiety disorder: Is CBT equally efficacious in adults of working age and older adults? *Clinical Psychology Review, 52*, 124-136. https://doi.org/10.1016/j.cpr.2017.01.003

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

Layard, R. (2006). *The depression report: A new deal for depression and anxiety disorders*. https://cep.lse.ac.uk/pubs/download/special/depressionreport.pdf

Layard, R., Clark, D., Knapp, M., & Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *National Institute Economic Review, 202*, 90-98. https://doi.org/10.1177/0027950107086171

Leon, A. C., & Heo, M. (2009). Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model. *Computational Statistics & Data Analysis, 53*(3), 603-608. https://doi.org/10.1016/j.csda.2008.06.010

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample Size Requirements for Multivariate Models to Predict Between-Patient Differences in Best Treatments of Major Depressive Disorder. *Clinical Psychological Science, 7*(3), 445-461. https://doi.org/10.1177/2167702618815466

Marks, D. F. (2018). IAPT under the microscope. *Journal of Health Psychology, 23*(9), 1131-1135. https://doi.org/10.1177/1359105318781872

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2007). *Paying the price: the cost of mental health care in England to 2026*. https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal Clinically Important Difference: Defining What Really Matters to Patients. *JAMA, 312*(13), 1342-1343. https://doi.org/10.1001/jama.2014.13128

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *Journal of Clinical Psychiatry, 74*(6), 595-602. https://doi.org/10.4088/JCP.12r07757

McManus, S., Meltzer, H., Brugha, T., Bebbington, P. E., & Jenkins, R. (2009). *Adult psychiatric morbidity in England: results of a household survey*. https://files.digital.nhs.uk/publicationimport/pub02xxx/pub02931/adul-psyc-morb-res-hou-sur-eng-2007-rep.pdf

National Collaborating Centre for Mental Health. (2019). *The Improving Access to Psychological Therapies Manual - Appendices and helpful resources*. https://www.rcpsych.ac.uk/docs/default-source/improving-care/nccmh/iapt/nccmh-iapt-manual-appendices-helpful-resources-v2.pdf?sfvrsn=a607ef5_4

National Institute for Health and Care Excellence. (2011). *Common mental health disorders: identification and pathways to care*. https://www.nice.org.uk/guidance/cg123

National Institute for Health and Care Excellence. (2020). *Generalised anxiety disorder and panic disorder in adults: management Clinical guideline [CG113]*. https://www.nice.org.uk/guidance/cg113

National Institute for Health and Care Excellence. (2021). *Sapropterin for treating hyperphenylalaninaemia in phenylketonuria*. https://www.nice.org.uk/guidance/ta729

National Institute for Health and Care Excellence. (2022). *Depression in adults: treatment and management NICE guideline [NG222]*. https://www.nice.org.uk/guidance/ng222

NHS Digital. (2020). *Psychological Therapies - Annual report on the use of IAPT services, England 2019-20.* https://files.digital.nhs.uk/B8/F973E1/psych-ther-2019-20-ann-rep.pdf

NHS England. (2019). *NHS Mental Health Implementation Plan 2019/20–2023/24*. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/07/nhs-mental-health-implementation-plan-2019-20-2023-24.pdf

Pearl, S. B., & Norton, P. J. (2017). Transdiagnostic versus diagnosis specific cognitive behavioural therapies for anxiety: A meta-analysis. *Journal of Anxiety Disorders, 46*, 11-24. https://doi.org/10.1016/j.janxdis.2016.07.004

Pencina, M. J., & Peterson, E. D. (2016). Moving From Clinical Trials to Precision Medicine: The Role for Predictive Modeling. *JAMA, 315*(16), 1713-1714. https://doi.org/10.1001/jama.2016.4839

Perfect, D., Jackson, C., Pybis, J., & Hill, A. (2016). *Choice of therapies in IAPT - An overview of the availability and client profile of step 3 therapies*. https://www.bacp.co.uk/media/1977/bacp-choice-of-therapies-in-iapt.pdf

Perlman, R. L., & Govindaraju, D. R. (2016). Archibald E. Garrod: the father of precision medicine. *Genetics in Medicine, 18*(11), 1088-1089. https://doi.org/10.1038/gim.2016.5

Pincus, T., Miles, C., Froud, R., Underwood, M., Carnes, D., & Taylor, S. J. (2011). Methodological criteria for the assessment of moderators in systematic reviews of

randomised controlled trials: a consensus study. *BMC Medical Research Methodology, 11*, 14. https://doi.org/10.1186/1471-2288-11-14

Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *Lancet, 370*(9590), 859-877. https://doi.org/10.1016/s0140-6736(07)61238-0

Rawlins, M. D. (2015). National Institute for Clinical Excellence: NICE works. *Journal of the Royal Society of Medicine, 108*(6), 211-219. https://doi.org/10.1177/0141076815587658

Redekop, W. K., & Mladsi, D. (2013). The faces of personalized medicine: a framework for understanding its meaning and scope. *Value in Health, 16*(6 Suppl), S4-9. https://doi.org/10.1016/j.jval.2013.06.005

Robinson, L., Kellett, S., & Delgadillo, J. (2020). Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. *Depression and Anxiety, 37*(3), 285-294. https://doi.org/10.1002/da.22999

Santoft, F., Axelsson, E., Ost, L. G., Hedman-Lagerlof, M., Fust, J., & Hedman-Lagerlof, E. (2019). Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. *Psychological Medicine, 49*(8), 1266-1274. https://doi.org/10.1017/S0033291718004208

Saunders, R., Buckman, J. E. J., Cape, J., Fearon, P., Leibowitz, J., & Pilling, S. (2019). Trajectories of depression and anxiety symptom change during psychological therapy. *Journal of Affective Disorders, 249*, 327-335. https://doi.org/10.1016/j.jad.2019.02.043

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Springer, K. S., Levy, H. C., & Tolin, D. F. (2018). Remission in CBT for adult anxiety disorders: A meta-analysis. *Clinical Psychology Review, 61*, 1-8. https://doi.org/10.1016/j.cpr.2018.03.002

Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M., & Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. *Journal of Affective Disorders, 175*, 152-167. https://doi.org/10.1016/j.jad.2014.12.025

Weitkamp, K., Klein, E., & Midgley, N. (2016). The Experience of Depression: A Qualitative Study of Adolescents With Depression Entering Psychotherapy. *Global Qualitative Nursing Research, 3,* 1-12. https://doi.org/10.1177/2333393616649548

Weitz, E. S., Hollon, S. D., Twisk, J., van Straten, A., Huibers, M. J., David, D., . . . Cuijpers, P. (2015). Baseline Depression Severity as Moderator of Depression Outcomes Between Cognitive Behavioral Therapy vs Pharmacotherapy: An Individual Patient Data Meta-analysis. *JAMA Psychiatry, 72*(11), 1102-1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Woodgate, R. L., Tennent, P., Barriage, S., & Legras, N. (2020). The lived experience of anxiety and the many facets of pain: A qualitative, arts-based approach. *Canadian Journal of Pain, 4*(3), 6-18. https://doi.org/10.1080/24740527.2020.1720501

4 Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials

4.1 Statement of Authorship

This declaration concerns the article entitled:			
Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials			
Publication status **Draft manuscript** [ ] **Submitted** [ ] **In review** [ ] **Accepted** [ ] **Published** [X]			
Publication details	Bauer-Staeb, C., Kounali, D. Z., Welton, N. J., Griffith, E., Wiles, N. J., Lewis, G., ... & Button, K. S. (2021). Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials. *J Clin Epidemiol*, *137*, 200-208. https://doi.org/10.1016/j.jclinepi.2021.04.002		
Copyright status (tick the appropriate statement) The material has been published with a CC-BY license [X] The publisher has granted permission to replicate the material included here [ ]			
Candidate's contribution to the paper	The candidate predominantly executed the formulation of ideas (70%) and the design of methodology (75%). The candidate predominantly conducted the formal analysis (80%) and executed the presentation of data in journal format (80%).		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	[redacted]	**Date**	28/08/2022

Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials

Clarissa Bauer-Staeb,[a] Daphne-Zacharenia Kounali,[b] Nicky J. Welton,[b] Emma Griffith,[a,c] Nicola J. Wiles,[b] Glyn Lewis,[d] Julian J. Faraway,[e] Katherine S. Button[a]

[a] Department of Psychology, University of Bath, Bath, UK

[b] Bristol Medical School, University of Bristol, Bristol, UK

[c] Avon and Wiltshire Mental Health Partnership NHS Trust, Bath, UK

[d] Division of Psychiatry, University College London, London, UK

[e] Department of Mathematical Sciences, University of Bath, Bath, UK

4.2 Abstract

Objective. Previous research on the minimal clinically important difference (MCID) for depression and anxiety is based on population averages. The present study aimed to identify the MCID across the spectrum of baseline severity.

Study Design and Settings. The present analysis used secondary data from two randomised controlled trials for depression (n =1,122) to calibrate the Global Rating of Change with the PHQ-9 and GAD-7. The MCID was defined as a change in scores corresponding to a 50% probability of patients "*feeling better*", given their baseline severity, referred to as the Effective Dose 50 (ED50).

Results. MCID estimates depended on baseline severity and ranged from no change for very mild symptom severity to 14 points (52%) on the PHQ-9 and up to 10 points (48%) on the GAD-7 for very high severity. The average MCID estimates were 3.7 points (23%) and 3.3 (28%) for the PHQ-9 and GAD-7, respectively.

Conclusion. The ED50 method generates MCID estimates across the spectrum of baseline severity, offering greater precision but at the cost of greater complexity relative to population average estimates. This has implications for evaluations of treatments and clinical practice where users can employ these results to tailor the MCID to specific populations according to baseline severities.

4.3 What's New

- Previous minimal clinically important difference (MCID) estimates are based on typical populations.
- It may be insufficient to have one MCID estimate when evaluating clinically meaningful change across a population with varying degrees of symptom severity – more tailored approaches may be beneficial.
- The present findings provide estimates inclusive of the full spectrum of baseline severity, allowing for a precise MCID to be tailored to specific populations of interest.

4.4 Introduction

Depression and anxiety are the most common mental health problems worldwide.[1] In the absence of objective tests, self-report questionnaires are frequently used to measure symptom change. However, uncertainty remains about how much change on these questionnaires is clinically meaningful. A first step towards conceptualising clinically meaningful improvement has been to define minimal clinically important differences (MCID) - the smallest difference in scores that are of perceived benefit to patients.[2] While various methods of estimating meaningful differences on questionnaires exist, it is imperative to include patients' perceptions to define clinically meaningful change,[2-4] particularly where subjective experiences, such as depression and anxiety, are targeted. Anchor-based approaches, which anchor questionnaire outcomes in patients' reports of subjective improvement, are truly patient-centred by incorporating the patients' experiences.[2]

Early work estimating the MCID using these methods for the Beck Depression Inventory-II demonstrated baseline dependency.[5,6] Patients with a higher baseline severity require larger changes to experience a subjective improvement. Various methods exist to address this problem (see supplementary material B). Two commonly used methods include examining the standardised mean differences amongst those who report slight improvements compared to those who feel the same or, alternatively, measuring the proportionate change - percentage change in symptoms relative to baseline.[5-9] Recent research has explored the MCID for depression and anxiety on the Patient Health Questionnaire (PHQ-9) and Generalised Anxiety Disorder Scale (GAD-7).[10-12] Collectively, the research suggests that the MCID can be defined as approximately 20% improvement from baseline.[5,9] While providing a good rule-of-thumb, they are unable to fully capture baseline dependency equally well across all patients – the 20% estimate applies less well to patients with lower baseline severity or patients with treatment-resistant depression with higher baseline severity.[5,9] This is substantiated by research demonstrating a 51% disagreement when comparing the 20% MCID to patient self-reported improvement.[13] Standardised effect sizes have been criticised for being difficult to interpret and providing little clinical information.[14] As such, there is a need to further address baseline dependency when estimating the MCID. In light of the above, we present a novel approach to estimating a baseline-dependent MCID for widely used measures of depression and anxiety – the PHQ-9 and GAD-7.[10,11]

4.5 Methods

4.5.1 The sample

The present study used data from two multi-centre randomised controlled trials (RCTs): PANDA and CoBalT.[15,16] PANDA (*n*=653) compared sertraline vs. placebo in primary care patients where there was clinical equipoise about the benefits of antidepressant medication.[15] CoBalT (*n*=469) compared cognitive behavioural therapy (CBT) as an adjunct to usual care (pharmacotherapy) to usual care alone in primary care patients with treatment-resistant depression.[16] The data were pooled across RCTs, resulting in more observations at each level of baseline severity and therefore increasing the precision of analyses. Data from all treatment arms were used as we assume stability between change in symptoms and subjective improvement, irrespective of how the change in symptoms is brought about. Pooling data from both RCTs and across treatment arms also increases the generalisability of the results.

4.5.2 The 9-Item Patient Health Questionnaire (PHQ-9) and the 7-Item Generalised Anxiety Disorder Scale (GAD-7)

The PHQ-9 and GAD-7 are two very widely used self-report questionnaires assessing the severity of depression and anxiety symptoms over the past 2 weeks.[10,11] The range for the PHQ-9 is 0-27 and 0-21 on the GAD-7, with higher scores indicating greater symptom severity. The PHQ-9 and GAD-7 were completed at baseline, 2, 6, and 12 weeks in PANDA.[15] In CoBalT, the PHQ-9 was measured at baseline, 3, 6, 9, and 12 months whereas the GAD-7 was collected at baseline, 6 and 12 months.[16]

4.5.3 Global Rating of Change (GRC)

PANDA and CoBalT included the 1-item GRC asking patients how they felt compared to when they were last seen.[15-18] The GRC was measured at all follow-up time points. CoBalT patients could respond: *"I feel better", "I feel about the same",* and *"I feel worse".*[16] In PANDA, patients could answer: "*I feel a lot better", "I feel slightly better", "I feel about the same", "I feel slightly worse",* and *"I feel a lot worse".*[15] For all models, groups were dichotomised into *feeling better* and *not feeling better,* as the aim was to estimate the point at which patients experience an improvement. The category *not feeling better* consisted of patients who felt the same or worse.

4.5.4 Statistical analysis

An extensive methodological justification can be found in supplementary material B. All analyses were performed in the R statistical programming language.[19]

4.5.4.1 Modelling changes across time

Change across multiple follow-ups was calculated from the previous timepoint (a rolling baseline), so at time *t* the change is: $x_{(t-1)} - x_t$, where x_t is the follow-up score at timepoint *t*. Negative scores indicate deteriorations in symptoms whereas positive scores indicate improvements in symptoms.

To establish that the GRC is an appropriate anchor, Spearman rank correlation coefficients were estimated, examining the association between the categorical GRC and change scores. Correlation coefficients ≥0.30 have been deemed as appropriate.[20] This threshold was exceeded across studies and time points ranging from -0.32 to -0.52 (supplementary table B1).

4.5.4.2 Generalised additive mixed models (GAMMs)

GAMMs provide a flexible approach to model complex, interacting relationships while maximising model fit. A logistic GAMM was fitted, specifying the binary GRC (*better vs. not better*) as the outcome using the "mgcv" package.[21] Change in symptom scores and baseline severity were classed as predictors with an interaction term given the established importance of baseline dependency.[4,5,12] Due to the repeated measurement, a random intercept was included for patients.[15,16] There is a natural variation in GRC responses between individuals – different patients will be more or less likely to respond feeling *better* or *not better* even when accounting for baseline severity and change. Random effects can account for the correlation between repeated observations of the same individual. In order to deal with the intrinsic correlation between change and baseline scores as well as the bounded nature of the scales thin plate splines with a monotonicity constraint were used to model the combined effect of change and baseline severity on the response.[21] As the data were obtained from two separate studies and collected over multiple follow-up periods, a further model evaluated the effects of time and study by adding these as covariates. Model summaries and 95% confidence intervals can be found in supplementary table B2 and supplementary figures B1 and B2.

4.5.4.3 Effective Dose 50 (ED50)

In the present study we applied the ED50 as a new method to estimate the MCID. The ED50 is an interpretable and well-validated measure used in drug safety and pharmaceutical research to determine minimum thresholds for effective therapeutic doses.[22] Applied to the current context, the ED50 is the change in scores where there is a 50% probability of patients reporting *feeling better*. The ED50 has face validity as an MCID as it identifies the smallest point where a patient might be marginally more likely to *feel better* than *not*. Further face validity is added to the concept of using the ED50 as a MCID given that the lowest bound of response to treatment is often defined as a 50% improvement.[23] Here, this principle is applied to the subjective experience of improvement rather than the symptom measure itself. From the GAMMs, we predicted the probability of response and identified the change in scores associated with 50% probability of *feeling better*. A limit of 0 change was set, as it would be clinically unacceptable to classify symptom deteriorations as improved. The absolute MCIDs were converted to a percent change from baseline. The ED25 and ED75 - the point at which there is a 25% and 75% probability that the patient reports *feeling better* - were also calculated as interval estimates, providing an index of variability of *feeling better* under different clinically acceptable probabilities. Furthermore, the sensitivity and specificity of the ED50 as well as the agreement between the MCID and patient-reported improvement were estimated.

4.5.4.4 (Standardised) Mean Difference (SMD)

To allow for comparisons with more traditional methods, the crude and standardised mean difference between those "*feeling slightly better*" and those "*feeling about the same*" were examined in Panda using the "TableOne" package.[7-9,24] These data were not available in CoBalT.

4.6 Results

4.6.1 Sample Characteristics

Baseline sociodemographic and clinical characteristics of all patients recruited into the RCTs are presented in Table 1. Patients in PANDA had a lower clinical severity at baseline, with moderate symptoms of depression and mild anxiety. Patients in CoBalT had higher scores across all measures with severe depression and moderate anxiety scores. Table 2 shows the mean change associated with GRC responses, stratified by study and follow-up time.

4.6.2 GAMM

We found statistically significant effects of study and time on the probability of feeling better. However, as might be expected, these made little difference to the MCID estimates and were therefore omitted from the final model for interpretability and generalisability (see supplementary tables B3 and B4). Of note, the effects of study on probability of feeling better appear to be driven by the differing baseline severities of the two samples at time point 1 due to their differing selection criteria (see supplementary tables B3 and B4). Combining the datasets is advantageous as it provides rich data across the distribution of baseline scores and the model produces a weighted average that accounts for the number of observations in each study.

4.6.3 ED50

Table 3 shows the ED estimates for both questionnaires. Across the PHQ-9 and GAD-7, patients with minimal symptoms at baseline need no change to have at least a 50% probability of feeling better; however, as severity increases the ED estimates increase in incremental steps. However, this is not a uniform, linear pattern, demonstrating the complexity of the effect change and baseline severity have on the probability of feeling better.

The ED50 score averaged over patients coincides with moderate depression (PHQ-9) and mild anxiety (GAD-7). However, there was a large range of MCID estimates, from 0 points (0%) up to 14 points (52%) on the PHQ-9, and up to 10 points (48%) on the GAD-7. Larger changes are needed on the GAD-7 than the PHQ-9 to feel *better*.

The models could not predict higher probabilities of feeling better amongst patients with very low baseline severity on the GAD-7, given the marginal ability to improve in symptoms. Patients would have to change more than is possible to obtain high probabilities of improvement. For clinical interpretation, equating these to 100% change is reasonable.

4.6.4 Sensitivity and Specificity

Table 4 demonstrates that the ED50 estimates show adequate sensitivity and specificity, providing a reasonable estimate for the smallest change in scores needed to *feel better.* The specificity of the ED50 was generally higher than the sensitivity and did not fall below 0.70,

which could be deemed a clinically acceptable threshold. The disagreement between GRC and improvements based on the ED50 was 28.4% on the PHQ-9 and 28.9% on the GAD-7.

4.6.5 (S)MD

Table 5 shows that the mean difference between those *feeling the same* and those *feeling slightly better* was ~ 2 points on both questionnaires. The SMD on the PHQ-9 was ~0.6 and ~0.5 on the GAD-7.

4.7 Discussion

A patient-centred approach was taken to estimate the MCID for widely used measures of depression and anxiety. The MCID was defined in a novel way as the change in scores that reflects at least a 50% probability that patients report *feeling better*. We produced MCID estimates stratified by severity scores, which increased with baseline severity in a non-linear manner, ranging from no change for very mild baseline severity up to 14 points (52%) on the PHQ-9 and up to 10 points (48%) on the GAD-7 for high severity. Across the sample, the average MCID estimates were 3.7 points (23%) and 3.3 (28%) for the PHQ-9 and GAD-7 respectively. For comparative purposes, the (standardised) mean difference method was applied to PANDA yielding estimates of ~0.6 and ~0.5 for the PHQ-9 and GAD-7 respectively.[7-9]

Previous research modelling proportionate change suggests the MCID is ~ 20-30% improvement for moderately-severe populations for depression and anxiety respectively.[5,9] Specifically, for patients with a moderate baseline severity a MCID of 21% change on the PHQ-9 and a 27% change from baseline on the GAD-7 were previously reported, which translates into a 1.7 and 1.5 point improvement, respectively and standardized mean differences of ~0.5.[9] This is consistent with other medical fields where MCIDs defined as effect sizes range from 0.3-0.5.[5,12] Primary care services providing psychological therapy for depression and anxiety in England currently a use a 6- and 4-point change for the PHQ-9 and the GAD-7 respectively to capture improvement, which are based on Jacobsen and Traux's Reliable Change Index.[25,26]

The MCID is a concept; it is not mathematically defined. There are various methods by which it can be estimated, each with different modelling assumptions and inferential

objectives, meaning any comparisons between estimates are indirect and crude. However, the flexibility of the present method allows different levels of the probability of response to be modelled, contextualising where previous methods may lie on the spectrum of probability of *feeling better*. The mean difference method, applied in the less severe PANDA sample, suggests an MCID of ~2 points or a SMD ~0.5-0.6, which is comparable to previous research.[9,12] We advocate for the ED50 to be used at each level of baseline severity as the mean will vary from study to study based on the severity of the sample. Indeed, when we include the more severe CoBalT sample, we find the mean of the ED50 estimates across patients yields somewhat higher MCID estimates (~3.5) in absolute terms. However, the averages of the present changes (~20-30%) is very similar to previous estimates, as might be expected given that proportional change accounts for some of the observed baseline severity.[12] The ED estimates suggest that previous methods in research settings appear to define the MCID as a probability of *feeling better* that lies somewhere between 25% and 50%. The 6- and 4-point PHQ-9 and GAD-7 estimates used in clinical practice appear to fall within 50% to 75% probability of response. [25,26] Given the ambiguity of what can be defined as a clinically acceptable probability of response, the current method also affords flexibility to the user to determine which level of probability is appropriate in a given context.

Interestingly, patients with very low baseline severities do not appear to require an improvement in scores to have a 50% probability of *feeling better*. This initially appears to contrast our previous research, which used Bayesian hierarchical regression models and derived parameters to calculate the optimum sensitivity and specificity on a Receiver Operator Characteristics (ROC) curve and found patients with low baselines severity needed larger changes proportionately to *feel better*. [9] However, at very low baselines no change versus a 1-point improvement translates into a large difference in proportionate change of 0% or 100%, respectively, for those with a baseline score of 1. Therefore, this seeming discrepancy is essentially two sides of the same coin, reflecting problems of estimation at the lower end of the scale which manifest differently according to the method used. This is supported by the observation that at low baseline severity the agreement between MCID and GRC responses appears lower.[9] It may be difficult for patients to discern a precise point at which they experience an improvement when there is such little scope to change in questionnaire scores. This suggests that the measures used may not be sufficiently sensitive for the lower ranges of severity,

highlighting a need for further exploration of how to evaluate interventions in subclinical populations where conventional scales are at the limit of their operability.

Importantly, the present research also highlights a large range of MCID estimates which suggests that previous MCID estimates may be well suited for typical/average populations but may not capture the MCID across all patients equally well. Previous approaches provide an easy to implement guide but come at a cost of 51% disagreement between the MCID and patient reports of improvement. [10] The present approach is more specific, with ~ 23% better agreement, but at a cost of greater complexity to implement by providing an MCID for each level of baseline severity.

4.7.1 Strengths and limitations

The present study used data from two high-quality RCTs resulting in a large sample with clinically distinct populations, which is critical given that the MCID is baseline dependent. The GRC has clear face validity providing a useful patient-centred anchor.[17]

The use of difference scores was a limitation as it ignores the measurement error; however, these effects are largely mitigated by the use of smoothing parameters in the statistical analysis. Furthermore, the GRC is subjective in nature - the concept of recovery is complex and unique to each patient. Clinical questionnaires commonly focus solely on symptoms. Responses to the GRC may incorporate wider (mental) health and psychosocial influences, such as comorbidities, life events, or quality of life, that may not be captured by depression symptoms alone.[27] Further adjustment of predictors may improve the accuracy of the MCID estimates. However, these influences are likely to be wide and varied and would therefore require very large samples and could not be completed in the present analysis due to sample size limitations. It is also noteworthy that we assumed the relationship between changes in outcome questionnaires and subjective improvements, was not affected by treatment. Future research could examine this relationship more closely and how it may be affected by different treatments and research design characteristics such as blinding. The secondary use of data resulted in further limitations. PANDA and CoBalT had different follow-up time points potentially resulting in time-dependent confounding. However, random effects were introduced to account for repeated measurements and the effects of time were not practically meaningful for estimating the MCID. The two studies also had differing levels of granularity of the GRC scales which meant we could only estimate the differences

in mean change between those feeling *slightly better* and *the same* in PANDA. We used all of the data in our GAMM model, combining *same* and *worse* into a single *not better* category to keep in line with our previous research.[5,12] Our MCID estimates may be over-estimated as a consequence relative to methods which exclude those who feel worse (see supplementary material B). Although patients in both RCTs experienced depression and anxiety to varying degrees, the results indicate that greater changes are needed on the GAD-7 to feel better than the PHQ-9. Both studies recruited patients on the basis of depression as the primary problem. Changes in depression may have been perceived of greater relative importance, requiring smaller changes to *feel better*. As such, findings may not generalise to populations experiencing anxiety as their primary or only problem.

4.7.2 Implications

Despite the limitations, providing estimates to measure clinically meaningful change has important implications for research as well as clinical practice. In the analysis of results from clinical trials, the MCID could be applied to each patient within the treatment arms, allowing for comparisons between treatments on the number of patients who scored a change equal to or greater than the MCID. In a similar notion, the MCID could inform evaluations in clinical practice bringing greater face validity to experiences of symptomatic improvement in conceptualisations of clinical recovery. Equally, the within-subject change could be applied to examine between-treatment differences. While the MCID might be relevant to superiority and equivalence trials, it may be particularly pertinent to non-inferiority trials where an alternative treatment is cheaper, less resource-intensive, or simpler to implement. Here, the MCID could be used to ascertain that the difference in treatment effects does not exceed the MCID; thereby, allowing for evaluations of cost-effectiveness that assure a newer or cheaper treatment is not of less benefit to patients. The ED50 MCID can inform sample size calculations by providing mean estimates of the expected change where at least 50% of patients would experience an improvement. However, they cannot inform the variance part of such calculations which will require wider considerations of the population studied. Baseline variability in outcome scores, however, is the major driver of patient heterogeneity and population level estimates of variance are easily obtainable.

4.7.3 Conclusion

The MCID contributes to our ability to assess clinically meaningful change rather than statistical significance alone. However, the research highlights the difficulty of calibrating

patient experiences with structured questionnaires, such as the need to account for baseline severity. Here, we present an approach where the MCID is tailored to baseline severity to fully capture the entire spectrum of severity. Such approaches come at the cost of greater complexity but offer greater precision. The development and triangulation of different methods will advance our understanding of how abstract concepts can be defined mathematically and contextualise what different MCID approaches are measuring.

4.8 Tables

Table 1. *Sociodemographic and clinical characteristics, stratified by study*

	PANDA	**CoBalT**
n	*653*	*469*
Age (years)	39.70 (14.93)	49.59 (11.70)
Female	384 (59%)	339 (72%)
White*	579 (89%)	459 (98%)
Marital Status*		
Married or living as married	255 (39%)	248 (53%)
Single	296 (45%)	89 (19%)
Separated, divorced, or widowed	101 (15%)	132 (28%)
In paid employment *	433 (66%)	206 (44%)
Highest educational qualification*†		
A level, higher grade or above	450 (69%)	217 (47%)
GCSE, standard grade or other	169 (26%)	130 (28%)
No formal qualifications	33 (5%)	116 (25%)
Financial difficulty*		
Living comfortably or doing alright	364 (56%)	167 (36%)
Just about getting by	204 (31%)	174 (37%)
Finding it difficult or very difficult to make ends meet	84 (13%)	128 (27%)
Number of life events in past 6 months	1.22 (1.19)	1.25 (1.15)
12-Item Short Form Survey mental health subscale	32.47 (11.04)	28.60 (9.14)
12-Item Short Form Survey physical health subscale	52.07 (9.70)	43.45 (13.47)
Patient Health Questionnaire-9	12.00 (5.80)	16.59 (5.67)
Generalised Anxiety Disorder Scale-7	9.43 (5.28)	11.75 (5.05)

**Data missing for one person in Panda. † Data missing for six people in CoBalT*

Data are presented as mean (standard deviation) for continuous variables and n (%) for categorical variables.

Table 2. *Mean change in outcome questionnaires, stratified by Global Rating of Change, study, and follow-up*

		Baseline to Follow-up 1					Follow-up 1 to Follow-up 2				
	Global Rating of Change	**Mean Baseline (SD)**	**n**	**%**	**Mean Change**	*SD*	**Mean Baseline (SD)**	**n**	**%**	**Mean Change**	*SD*
Patient Health Questionnaire-9											
PANDA	A lot better	11.89 (5.76)	35	6	**6.51**	*4.85*	10.04 (5.68)	110	21	**5.11**	*4.74*
	Slightly better		164	29	**3.57**	*4.52*		168	32	**2.70**	*4.46*
	About the same		291	51	**0.95**	*3.62*		172	32	**0.42**	*3.35*
	Slightly worse		63	11	**-0.19**	*3.58*		65	12	**-1.80**	*4.00*
	A lot worse		15	3	**-5.60**	*4.93*		16	3	**-4.75**	*3.96*
CoBalT	Better	16.48 (5.69)	214	49	**6.51**	*6.00*	12.59 (6.12)	202	49	**4.04**	*5.40*
	Same		168	38	**2.19**	*5.10*		140	34	**0.61**	*5.17*
	Worse		57	13	**-1.70**	*6.08*		69	17	**-3.07**	*6.28*
Generalised Anxiety Disorder Scale-7											
PANDA	A lot better	9.27 (5.29)	35	6	**6.00**	*5.57*	7.79 (5.35)	109	21	**4.34**	*4.50*
	Slightly better		163	29	**2.96**	*4.55*		166	31	**2.28**	*3.69*
	About the same		292	51	**0.59**	*3.62*		171	32	**0.58**	*3.61*
	Slightly worse		63	11	**0.06**	*4.26*		65	12	**-1.09**	*4.03*
	A lot worse		15	3	**-4.80**	*4.28*		16	3	**-4.25**	*4.22*
*CoBalT**	Better	-	**-**	-	**-**	-	11.64 (5.02)	205	49	**6.31**	*5.21*
	Same		**-**	-	**-**	-		142	34	**1.34**	*4.15*
	Worse		**-**	-	**-**	-		72	17	**-0.68**	*4.90*

		Follow-up 2 to Follow-up 3					Follow-up 3 to Follow-up 4				
	Global Rating of Change	**Mean Baseline (SD)**	**n**	**%**	**Mean Change**	*SD*	**Mean Baseline (SD)**	**n**	**%**	**Mean Change (SD)**	*SD*
Patient Health Questionnaire-9											
PANDA	A lot better	8.25 (5.68)	118	23	**3.14**	*4.31*	-	-	-	-	-
	Slightly better		143	28	**2.10**	*3.55*		-	-	-	-
	About the same		174	34	**0.22**	*3.24*		-	-	-	-
	Slightly worse		66	13	**-2.39**	*4.79*		-	-	-	-
	A lot worse		17	3	**-6.65**	*5.44*		-	-	-	-
CoBalT	Better	10.81 (6.88)	171	45	**2.23**	*4.96*	10.42 (6.74)	174	48	**2.34**	*5.03*
	Same		121	32	**0.52**	*4.89*		124	34	**-0.38**	*4.71*
	Worse		86	23	**-3.42**	*5.94*		62	17	**-2.89**	*6.26*
Generalised Anxiety Disorder Scale-7											
PANDA	A lot better	6.16 (5.17)	117	23	**2.16**	*3.91*	-	-	-	-	-
	Slightly better		143	28	**1.69**	*3.71*		-	-	-	-
	About the same		172	33	**-0.01**	*3.01*		-	-	-	-
	Slightly worse		66	13	**-1.73**	*3.71*		-	-	-	-
	A lot worse		17	3	**-4.24**	*4.96*		-	-	-	-
*CoBalT**	Better	-	-	-	-	-	8.12 (5.86)	186	48	**2.27**	*4.39*
	Same		-	-	-	-		135	35	**-0.22**	*4.38*
	Worse		-	-	-	-		65	17	**-2.88**	*4.48*

**Generalised Anxiety Disorder Scale-7 data was not collected at follow-up one and three. Baseline and change scores are derived from previous follow-up. SD - Standard deviation. Data reported for patients with complete Global Rating of Change and change scores on each respective outcome questionnaires.*

Table 3. *The Minimal Clinically Important Difference at each level of baseline severity*

		Patient Health Questionnaire -9				**Generalised Anxiety Disorder Scale -7**			
Baseline Score	**Clinical Cut-Off**	**ED25**	**ED50**	**ED50 (%)**	**ED75**	**ED25**	**ED50**	**ED50 (%)**	**ED75**
1	Minimal	0	0	0	1	0	0	0	N.A.
2		0	0	0	2	0	0	0	2
3		0	0	0	2	0	0	0	3
4		0	0	0	2	0	0	0	3
5	Mild	0	0	0	3	0	1	20	4
6		0	0	0	3	0	2	33	5
7		0	1	14	4	0	2	29	5
8		0	1	13	4	0	3	38	6
9		0	2	22	5	0	4	44	7
10	Moderate	0	3	30	5	0	4	40	7
11		0	3	27	6	0	5	45	8
12		0	4	33	6	1	5	42	8
13		0	4	31	7	2	6	46	9
14		1	5	36	7	2	6	43	9
15	Severe	1	5	33	8	3	7	47	10
16		2	5	31	9	3	7	44	11

17	2	6	35	9	4	8	47	11
18	3	7	39	10	5	8	44	12
19	3	7	37	11	5	9	47	12
20	4	8	40	12	6	10	50	13
21	4	9	43	13	6	10	48	13
22	5	10	45	14	-	-	-	-
23	6	11	48	14	-	-	-	-
24	7	11	46	15	-	-	-	-
25	7	12	48	16	-	-	-	-
26	8	13	50	17	-	-	-	-
27	9	14	52	18	-	-	-	-
Average across sample	1.2	3.7	23.3	6.4	1.0	3.3	28.0	6.1

**ED25 - Effective Dose 25; ED50 - Effective Dose 50; ED75 - Effective Dose 75; N.A - Not available.*

Table 4. *Sensitivity and specificity of the Minimal Clinically Important Difference for the overall sample and stratified by study*

	Patient Health Questionnaire -9		**Generalised Anxiety Disorder Scale -7**	
	Sensitivity	Specificity	Sensitivity	Specificity
Overall	0.65	0.77	0.67	0.75
PANDA	0.69	0.73	0.65	0.72
CoBalT	0.61	0.83	0.70	0.81

Table 5. *Standardised Mean Difference based on subgroups of the Global Rating of Change, stratified by time in PANDA*

	Feeling Slightly Better	**Feeling the Same**		
Patient Health Questionnaire -9	***Change (SD)***	***Change (SD)***	***Crude Difference***	***Standardised Mean Difference***
Baseline to Follow-up 1	3.57 (4.52)	0.95 (3.62)	2.62	0.64
Follow-up 1 to Follow-up 2	2.70 (4.46)	0.42 (3.35)	2.28	0.58
Follow-up 2 to Follow-up 3	2.10 (3.55)	0.22 (3.24)	1.88	0.55
Generalised Anxiety Disorder Scale-7				
Baseline to Follow-up 1	2.96 (4.55)	0.59 (3.62)	2.37	0.58
Follow-up 1 to Follow-up 2	2.28 (3.69)	0.58 (3.61)	1.7	0.47
Follow-up 2 to Follow-up 3	1.69 (3.71)	-0.01 (3.01)	1.7	0.50

**SD - Standard deviation*

4.9 References

[1] World Health Organisation. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017. Available from: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf

[2] McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA*. 2014;312(13):1342–1343. https://doi.org/10.1001/jama.2014.13128

[3] Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Res Ther.* 2015;17(1):143. https://doi.org/10.1186/s13075-015-0658-6

[4] Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7(5):541–546. https://doi.org/10.1016/j.spinee.2007.01.008

[5] Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, Ades AE, Lewis G. Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychol Med.* 2015;45(15):3269–3279. http://doi.org/10.1017/S0033291715001270

[6] Beck A, Steer RA, Brown GK. Beck Depression Inventory – Second Edition: Manual. The Psychological Corporation: San Antonio, TX, 1996.

[7] Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch Intern Med*. 1993 Jun;153(11):1337-1342.

[8] Borenstein M. Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC (Eds.). The Handbook of research synthesis and meta-analysis (2nd Ed.). New York: Russell Sage Foundation; 2009. p. 222-236.

[9] Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol.* 2017;82:128–136. http://doi.org/10.1016/j.jclinepi.2016.11.016

[10] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a Brief depression Severity Measure. *J Gen Intern Med*. 2001;16(9):606–13. http://doi.org/10.1046/j.1525-1497.2001.016009606.x

[11] Spitzer RL, Kroenke K, Williams JB, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med.* 2006;166(10):1092–1097. http://doi.org/10.1001/archinte.166.10.1092

[12] Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, Duffy L, Lanham P, Peters TJ, Wiles N, Lewis G. How Much Change is Enough? Evidence from a longitudinal study on depression in UK Primary Care. *Psychol Med*. 2020:52(10):1875–1882. http://doi.org/10.1017/S0033291720003700

[13] Hobbs C, Lewis G, Dowrick C, Kounali D, Peters TJ, Lewis G. Comparison between self-administered depression questionnaires and patients' own views of changes in their mood: a prospective cohort study in primary care. *Psychol Med*. 2020:51(5):853–860. http://doi.org/10.1017/S0033291719003878

[14] Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *J Affect Disord*. 2014;159:118–126. https://doi.org/10.1016/j.jad.2014.02.026

[15] Lewis G, Duffy L, Ades A, Amos R, Araya R, Brabyn S, Button KS, Churchill R, Derrick C, Dowrick C, Gilbody S. The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *Lancet Psychiatry.* 2019;6(11):903–914. https://doi.org/10.1016/S2215-0366(19)30366-9

[16] Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, Garland A, Hollinghurst S, Jerrom B, Kessler D, Kuyken W. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *Lancet*. 2013;381(9864):375–384. https://doi.org/10.1016/S0140-6736(12)61552-9

[17] Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther.* 2009;17(3):163–170. https://doi.org/10.1179/jmt.2009.17.3.163

[18] Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–415. https://doi.org/10.1016/0197-2456(89)90005-6

[19] R Core Team. A language and environment for statistical computing. Vienna, Austria; 2016. Available from: https://R-project.org/

[20] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

[21] Wood SN. Generalized Additive Models: An Introduction with R (2nd Ed). New York: C Chapman and Hall/CRC; 2017. https://doi.org/10.1201/9781315370279

[22] Dimmitt S, Stampfer H, Martin JH. When less is more–efficacy with less toxicity at the ED50. *Br J Clin Pharmacol.* 2017;83: 1365–1368. https://doi.org/10.1111/bcp.13281

[23] Macher JP, Crocq MA. Treatment goals: response and nonresponse. *Dialogues Clin Neurosci.* 2004;6(1):83–91. https://doi.org/10.31887/DCNS.2004.6.1/jpmacher2

[24] Panos A, Mavridis D. TableOne: an online web application and R package for summarising and visualising data. *Evid Based Ment Health*. 2020;23(3):127–130. http://dx.doi.org/10.1136/ebmental-2020-300162

[25] Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*.1991;59(1):12–19. https://doi.org/10.1037/0022-006X.59.1.12

[26] NHS England. Improving Access to Psychological Therapies Manual: Appendices and helpful resources; 2018. Available from: https://www.england.nhs.uk/wp-content/uploads/2018/06/iapt-manual-appendices-and-helpful-resources-v3.pdf

[27] Robinson J, Khan N, Fusco L, Malpass A, Lewis G, Dowrick C. Why are there discrepa ncies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open* . 2017;7(4):e014519. http://dx.doi.org/10.1136/bmjopen-2016-014519

4.10 Declarations of interest

None.

4.11 CRediT Author Statement

Clarissa Bauer-Staeb: Conceptualisation, Methodology, Formal Analysis, Writing – Original Draft, Writing – Review and Editing, and Visualisation

Daphne Kounali: Writing – Review and Editing

Nicola Wiles: Writing – Review and Editing, Resources, Funding Acquisition

Glyn Lewis: Writing – Review and Editing, Resources, Funding Acquisition

Nicky Welton: Writing – Review and Editing, Funding Acquisition

Emma Griffith: Writing – Review and Editing, Supervision

Julian Faraway: Conceptualisation, Methodology, Formal Analysis, Writing – Review and Editing, Supervision

Katherine Button: Conceptualisation, Methodology, Writing – Review and Editing, Supervision

4.12 Acknowledgements

PANDA was funded by the National Institute for Health Research Programme Grant for Applied Research (RP-PG-0610-10048). CoBalT was funded by National Institute for Health Research Health Technology Assessment (Project Number 06/404/02). The research was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust, the University College London Hospital NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the Sponsor, NHS, NIHR or Department of Health and Social Care. The funder had no role in the study design, data collection, data analysis, interpretation of data or writing of the present manuscript. We would like to thank all involved in participating, conducting, or otherwise supporting both RCTs and are grateful to all co-applicants of the RCTs who were not involved in drafting the present manuscript.

4.13 Trial Registration

Panda and CoBalT were registered with the Controlled Trials ISRCTN Registry: PANDA (ISRCTN84544741) and CoBalT (ISRCTN38231611).

5 Personalised Psychotherapy in Primary Care: An Evaluation of Data-Driven Treatment Allocation to Cognitive Behavioural Therapy vs. Counselling for Depression

5.1 Statement of Authorship

This declaration concerns the article entitled:			
Personalised Psychotherapy in Primary Care: An Evaluation of Data-Driven Treatment Allocation to Cognitive Behavioural Therapy vs. Counselling for Depression			
Publication status **Draft manuscript** [ ] **Submitted** [ ] **In review** [ ] **Accepted** [X] **Published** [ ]			
Publication details	Bauer-Staeb, C., Griffith, E., Faraway, J. J., & Button, K. S. Personalised Psychotherapy in Primary Care: An Evaluation of Data-Driven Treatment Allocation to Cognitive Behavioural Therapy vs. Counselling for Depression. *Manuscript accepted at BJPsych Open.*		
Copyright status (tick the appropriate statement) The material has been published with a CC-BY license [ ] The publisher has granted permission to replicate the material included here [X]			
Candidate's contribution to the paper	The candidate predominantly executed the formulation of ideas (90%) and the design of methodology (80%). The candidate predominantly made arrangements for access to the secondary data used (80%) and conducted the formal analysis (90%). The candidate predominantly executed the presentation of data in journal format (95%).		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	[redacted]	**Date**	22/02/2023

5.2 Abstract

Background. Various effective psychotherapies exist for the treatment of depression, however only approximately half of patients recover after treatment. In efforts to improve clinical outcomes, a greater emphasis is being placed on personalised psychotherapy - an attempt to match patients to treatments they are most likely to respond to.

Aim. The aim of the present research was to evaluate the benefit of a data-driven model to support clinical decision-making in differential treatment allocation to Cognitive Behavioural Therapy vs. Counselling for Depression.

Method. The present analysis used electronic healthcare records from primary care psychological therapy services for patients receiving CBT (n= 14,544) and CFD (n= 4725). A linear regression with baseline sociodemographic and clinical characteristics was used to differentially predict post-treatment Patient Health Questionnaire (PHQ-9) scores between the two treatments. The benefit of precision prescription was evaluated in a held-out validation set.

Results. On average, patients who received their model-indicated optimal treatment saw a 1.78 PHQ-9 points greater improvement. This translated into 4-10% more patients achieving clinically meaningful changes. However, for individual patients, the estimated differences in benefits of treatments were small and rarely met the threshold for minimal clinically important differences.

Conclusion. Precision prescription of psychotherapy based on sociodemographic and clinical characteristics is unlikely to produce large benefits for individual patients. However, when applied at scale, the benefits would be meaningful from an aggregate public health perspective.

5.3 Introduction

A range of psychotherapies are recommended by the National Institute of Health and Care Excellence for the treatment of depression,[1] with a large body of evidence suggesting psychotherapies are equally effective.[2] However, treatment effects remain modest.[3, 4] In the absence of superior treatments, personalised medicine has focused on identifying who responds best to which treatment.[5] Traditionally, such efforts have been explored with secondary data from randomised controlled trials (RCTs), which suffer from sample size limitations.[6] Further methodological limitations include a lack of validation in external samples and the examination of individual characteristics in isolation.[7] More novel approaches have been developed which take an actuarial approach, which have been implemented in RCT data to examine differential treatment effects in depression for Cognitive Behaviour Therapy (CBT) compared to antidepressant medication,[8] interpersonal psychotherapy,[9] and psychodynamic therapy.[10] Furthermore, research has started to use routinely collected data contained in electronic healthcare records. These have the benefit of including much large patient populations compared to those present in clinical trials. Recent research has shown that targeted prescription machine learning algorithm to assign patients to CBT vs. person-centred counselling for depression (CFD) resulted in approximately 20% greater improvements when patients were assigned to the optimal treatment as indicated by the model.[7] Further research using a patient profiling algorithm demonstrated that certain patient profiles saw greater improvement in CBT compared to counselling and vice versa.[11] Due to the relatively early stage of these novel approaches being implemented in healthcare records, less is known about replicability and generalisability of the results. Triangulation of evidence with different methodological approaches and using different samples will add to the evidence base. As such, we used a large-scale sample of healthcare records to assess the benefits of a differential treatment allocation of CBT vs. CFD, based on baseline patient characteristics, and to understand which variables contribute to potential differences in clinical outcomes between treatments. The validity of the models was tested in an external dataset.

5.4 Method

5.4.1 Settings

Improving Access to Psychological Therapies (IAPT) is a national programme that delivers psychological therapy for depression and anxiety across England. IAPT has implemented routine data collection, where detailed information is gathered about patients, their treatment, and their clinical outcomes.[12] These data are collected on a session-by-session basis to increase complete-case recording, even when patients drop out of treatment early.[12] The clinical records for the present study were obtained from a 15 IAPT services who were approached based on convenience and feasibility and agreed to participate. The services are located across the southwest of England and London, with the average Index of Multiple Deprivation of the sample population ranging from 12.91 vs 29.62 between services, the proportion of individuals from a Black, Asian, and ethnic minority background ranging 2.9% vs 58.9%, and services being located in a range of setting such as more urban inner-city areas as well as more rural settings. Data from 2012 to late 2019 was included. All data were extracted and fully anonymised by Mayden; the providers of the patient management software used in IAPT, who hold 61% of the market share for adult IAPT services.

5.4.2 Interventions

IAPT operates on a stepped care model whereby low-intensity therapy (LIT) are offered in the first instance and high-intensity therapy (HIT) is offered where response to LIT is insufficient or where there is a clinical necessity, such as a high baseline severity.[12] CBT and CFD are two of the most commonly available high-intensity therapies for depression in IAPT. CBT in IAPT is intended to be delivered in accordance with Beck's cognitive model.[13, 14] CFD in IAPT is intended to be delivered based on a person-centred, experiential therapy based on the humanistic model.[15] All therapies are delivered by accredited mental health professionals trained in accordance with the national curriculum.[14, 15]

5.4.3 Sample Selection

We identified all patients who received treatment for clinical levels of depression, based on a diagnosis of depression as well as a depression severity threshold of 10 points on the Patient Health Questionnaire-9 (PHQ-9) at baseline.[16] Patients were included in the present analysis if the majority of HIT was CBT or CFD. Majority HIT was defined as the most frequently recorded treatment label within all treatments that fall under the umbrella of HIT in IAPT.

LIT was not considered in this definition but prior LIT was accounted for in the analysis. Patients who received equal amounts of two different high-intensity therapy were excluded. In order to allow for pre-and post-treatment measures to be taken, patients had to attend at least 2 appointments. Amongst patients in this sample, the most recent referral was chosen where patients had a record of multiple prior treatments of CBT or CFD. All patients with missing outcome data at their last attended appointment were excluded. Due to the session-by-session recording of outcome measures in IAPT, this does not exclude patients who dropped out of treatment as their final outcome measure who be the one collected at their last attended appointment. All patients who completed outcome questionnaires at their last attended session either before dropping out or completing treatment were included. See supplementary figure C1 for a flowchart of the sample selection.

5.4.4 Measures

5.4.4.1 Outcome Measure

The PHQ-9 is a 9-item self-report questionnaire assessing the severity of depressive symptoms over the past 2 weeks.[16] Each item is rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). The total PHQ-9 score has a range of 0-27, with higher scores indicating greater symptom severity. Scores of 5, 10, 15, and 20 denote mild, moderate, moderately severe, and severe depression, respectively. The evidence suggests that a score of ≥10 on the PHQ-9 has an 88% sensitivity and specificity for identifying major depressive disorder.[16]

5.4.4.2 Patient Characteristics

The baseline variables consisted of data that are routinely collected at the point of referral or assessment. These include sociodemographic data: age, gender, ethnicity, employment status, sexual orientation. We additionally assessed the Index of Multiple Deprivation (IMD) as proxy for socioeconomic status.[17] A range of clinical variables are also collected: disability, and long-term health condition (LTC), diagnosis, depression symptoms (PHQ-9), anxiety symptoms (Generalised Anxiety Disorder Scale, GAD-7),[18] functional impairment (Work and Social Adjustment Scale, WSAS),[19] baseline psychotropic medication, and the referral source. From the available dataset we additionally determined who also received low-intensity therapy and the referral number measuring how many times a patient has been referred.

The variables included in the analysis are those that are routinely collected as part of the IAPT minimum dataset as they are available consistently across services. Due to a lack of clear and consistent evidence on moderators of psychotherapy outcomes (see section 3.4), the present analysis took an exploratory and inclusive approach that used all available variables which have also commonly been used in previous research exploring clinical outcomes in IAPT samples.[7, 11, 20-24] In addition to the variables available from the IAPT minimum dataset, the IMD was further included due to its association with clinical outcomes.[25] See supplementary table C1 for variable descriptions and coding.

5.4.5 Data Analysis

All analyses were performed using the R programming language.[26]

5.4.5.1.1 Test-train split

Prior to any data analysis, the dataset was randomly split into a training and testing set at a 3:1 ratio, respectively in order to have a held-out validation sample. This has the benefit of allowing the evaluation of the model in a previously unseen dataset. To assure training and testing datasets have similar characteristics and therefore assuring they are comparable, the balance of the partitioning was assessed on all variables included in the data analysis, including services and referral year, using standardized mean difference (SMD). Balance on all variables in the train and test samples were < 0.1, meeting a conservative threshold of balance (see supplementary table C2).[27]

5.4.5.1.2 Imputation

To address missing data, a non-parametric imputation for all baseline characteristics was performed using "MissForest" which uses a random-forest algorithm.[28] Random forest imputation has been shown to perform well in datasets with different data types and outperforms other methods of imputation where there are possible complex interactions and non-linear trends. [28] As missing outcome data at the last attended appointment was an exclusion criterion, these were not imputed for either train or test data. Random forest imputation was implemented to impute both categorical and continuous variables with 500 trees per forest. Service and year were also included to account for potential differences in patient populations across areas and time. Out-of-bag (OOB) imputation error estimates are reported to assess the imputation error using the normalized root mean squared error

(NRMSE) for continuous variables and the proportion of falsely classified entries (PFC) for categorical variables.[28] Imputation was performed separately for the training and testing datasets. Imputation was successful with an NRMSE of 0.40 and a PFC of 0.16.

5.4.5.1.3 Propensity Score Estimation

Due to the observational nature of the data, the allocation of treatment is not random – certain patients may be more likely to receive one type of treatment over another because of certain characteristics. Propensity scores estimate the probability of receiving one treatment over another based on observed baseline characteristics and can therefore, at least partially, account for patients' non-random treatment allocation. The propensity scores were added into all subsequent analyses as a covariate in addition to regression adjustment, resulting in a doubly robust approach. Previous research has demonstrated that doubly robust regression adjustment with propensity scores performs well in electronic healthcare records.[29]

5.4.5.1.4 Treatment Model

Arguably, differential treatment allocation is only useful when comparing two treatments that are equally effective - if one treatment is clearly superior it would generally be of greater value to simply provide the more effective treatment. Previous research in IAPT suggests that treatment outcomes in CBT and CFD are comparable.[30] While the aim of the present analysis is not to evaluate treatment efficacy, in order to assess the equivalence assumption a main effects model was fitted using linear regression, with post-treatment PHQ-9 score as the primary outcome. All baseline patient characteristics and the propensity score were added as covariates. We additionally adjusted for total number of appointments to control for treatment dose, the service, and the referral year.

5.4.5.1.5 Prediction Model

Non-specific predictors of treatment response are variables that influence how well a patient responds to therapy, irrespective of which treatment they receive, whereas moderators are variables that influence a better response to one treatment over another. [31] Within statistical models, predictors are coded as main effects and moderators are coded as interactions between a baseline characteristics and treatment. As interactions require more power, a different strategy is to examine specific predictors. These examine predictors in separate treatment arms to identify which variables are associated with outcomes in a particular

treatment, as has been implemented elsewhere.[7] However, only interactions are able to identify if variables produce a statistically significant difference in clinical outcomes between treatments. Due to the larger sample size, we opted to test for interactions.

In the present analysis, a linear regression was fitted in the train data with the patient's post-treatment PHQ-9 score as the primary outcome, covariate-adjusted for baseline PHQ-9.[16, 26] This approach was chosen over change-from-baseline calculations to avoid loss of power and the ability to account for measurement error. All baseline characteristics were added into the regression as main effects with an additional interaction term for treatment. We additionally accounted for the main effects of service, referral year, and propensity scores. Likelihood ratio tests were used to assess the significance of predictors and moderators for categorical variables with more than two levels.

To illustrate the magnitude of effect modification, predicted post-treatment PHQ-9 scores were estimated for CBT and CFD separately for each prescriptive variable while holding the remaining covariates of the model constant. Continuous variables were held constant at the mean, with categorical variables being set to the most frequent level. This effectively allows the moderating effects of baseline characteristics to be isolated. For example, if patient A was female and patient B was male, but they were otherwise identical on all other baseline characteristics, it is now possible to see how much of an impact gender has clinical outcomes between two different treatments (table 2).

5.4.5.1.6 External Validation

In order to test the generalisability of the results, the models were applied to the held-out test set. Within the test data, the post-treatment PHQ-9 score was predicted for CBT and CFD; thus, generating a prediction for the treatment response that patients actually received (a "factual" prediction), as well as for the treatment they did not receive (a "counterfactual" prediction). Following the Personalised Advantage Index (PAI) methodology,[8] the difference between the two predicted estimates was calculated to define the magnitude of benefit from one treatment over another. This difference can be seen as how much better or worse a patient would do if they received CBT vs. CFD, or vice versa. The treatment with the lowest predicted PHQ-9 score at the end of treatment is classified as the *optimal* treatment, whereas the treatment predicted to produce a higher score is the *suboptimal* treatment.

Previous research has shown that the PAI magnitude is not of relevance for all patients – many patients are likely to respond to both treatments similarly.[7,8] As a means of identifying patients who are likely to benefit from a differential treatment allocation, we identified patients with a high PAI score. We attempted to identify patients where the PAI exceeded the percent Minimal Clinically Important Difference (MCID). This is the smallest difference in scores where patients may experience a subjective improvement, estimated at an approximate reduction of 20% from baseline PHQ-9 scores.[32, 33] However, this number was very small and allowed for no meaningful comparison between patients. Previous research defined a high PAI as a score beyond one standard deviation from the mean.[7] We adopted a similar approach, defining a high PAI as a score beyond the first or third quartiles as the distribution was marginally skewed. As such, three groups were defined: patients who received their optimal treatment, patients who received their suboptimal treatment, and patients where no favourable treatment is indicated.

Subsequently, the observed post-treatment PHQ-9 scores were compared between patients receiving their *optimal* treatment and those receiving their *suboptimal* treatment. This comparison was made for adapted IAPT metrics of recovery, reliable change, and reliable recovery.[34] Recovery is defined as falling above clinical cut-offs on either depression or anxiety questionnaires pre-treatment and falling below these clinical cut-offs on depression and anxiety post-treatment.[34] The depression measure used in IAPT is the PHQ-9 and the clinical cut-off is ≥ 10 points.[34] Reliable change is measured as pre- and post-treatment questionnaire changes exceeding the measurement error on one or both depression or anxiety questionnaires (without a reliable deterioration on the other).[34] The reliable change threshold on the PHQ-9 is ≥ 6 points.[34] Reliable recovery is defined as a change in scores that exceeds the measurement error and scores falling below the clinical cut-offs.[34] In the present study these definitions were adapted to only incorporate the depression measure rather than a combination of depression and anxiety measures as the primary interest in the present study was depression. Furthermore, the comparison was also made for both a percent MCID, which is defined as a 20% reduction from baseline and an absolute MCID which has a range of values specific to baseline severity.[32, 33, 35] The difference/odds ratio of the outcomes between patients who received their *optimal* and *suboptimal* treatment were determined using simple linear and logistic regression, respectively.

5.5 Results

5.5.1 Sample Characteristics

The majority of the sample were women (67%), White (79.9%) with an average age of 40 years. Most patients had moderately severe depression (18 PHQ-9 points), moderate anxiety (14 GAD-7 points) and moderately severe functional impairment (23 WSAS points). There were differences in baseline characteristics in patients receiving CBT and CFD. Sample characteristics are described in Table 1, with a SMD threshold of <0.25 indicating adequate balance.[20] Patients who received CBT were more likely to have a diagnosis of recurrent depressive disorder, higher depressive symptom, and higher functional impairment. They were also more likely to be taking psychotropic medication, more likely to self-refer, and have received low-intensity therapy as well as having had a higher referral number. This could potentially suggest that clinical decision making with regards to treatment allocation incorporates baseline patient characteristics. However, it should be noted that these imbalances are not adjusted for other variables. As such, they could be a consequence of specific services having different populations and/or delivering a different ratio of CBT to CFD.

5.5.2 Main Treatment Effects

We found no evidence to suggest there are significant differences in treatment outcome between CBT vs CFD in this sample within a main effects model. After adjusting for baseline and treatment characteristics, the difference in post-treatment PHQ-9 score between was -0.10 (95% CI -0.39 to 0.18, p= 0.493).

5.5.3 Non-Specific Predictors & Moderators of Treatment Outcomes

Lower age, not working, higher IMD, having a disability or long-term health condition, as well as higher baseline PHQ-9, GAD-7, and WSAS scores were predictors of higher post-treatment PHQ-9 scores across both CBT and CFD (see supplementary table C3). Furthermore, taking medication, being referred from primary care or other services (vs. self-referring), and having a higher referral number were identified as predictors of worse outcome. Service and year were also predictive of clinical outcomes. After adjusting for other baseline characteristics, we found no evidence to suggest that gender, ethnicity, sexual orientation or also receiving low-intensity treatment was associated with outcomes.

We found weak evidence that employment status and psychotropic medication were moderators of clinical outcomes in CBT vs. CFD, but differences were of a very small clinical magnitude when other covariates are kept constant (Table 2). Other moderators also appear to make small differences in clinical outcomes when other covariates are kept constant, but none reached statistical significance.

5.5.4 External Cross-Validation

The discrepancy between the actual post-treatment score and the model predicted score was -0.19 (SD=6.44). The median PAI in the test sample was 0.11 (interquartile range: -0.29 to 0.53). This suggests that across all patients in the test sample, more patients may marginally benefit from CFD. However, as was found in previous research, these small differences suggest that not all patients benefit from a differential treatment allocation. As such, we identified patients who may benefit the most by selecting those with a PAI beyond the 1st and 3rd quartiles. In this 50% of patients, 1247 (51.8%) received their model indicated *optimal treatment*. Where CBT was the model indicated as the *optimal treatment*, i.e., where according to the model offering CBT would be indicated, 944 (78.3%) patients received CBT. Where CFD model indicated as the *optimal treatment,* i.e., where according to the model offering CFD would be indicated, 303 (25.2%%) patients received CFD.

Patients who received their *optimal treatment* scored -1.78 (95% CI -2.36 to-1.21, p <0.001) PHQ-9 points lower compared to those who received their suboptimal treatment (Table 3). Patients in the *optimal* group had a mean post-treatment PHQ-9 score of 9.63 (SD=6.95) whereas the suboptimal group scored 11.42 (SD=7.49). The odds of recovery for those receiving their *optimal treatment* vs. those who received their *suboptimal treatment* was 1.52 (78 (95% CI 1.29 to 1.79, p<0.001); 60.0% of patients in the *optimal* group recovered compared to 49.7% in the *suboptimal* group. The odds of achieving a reliable change for those receiving their *optimal treatment* vs. those who received their *suboptimal treatment* was 1.19 (95% CI 1.01 to 1.40, p=0.038); 63.8% of patients in the *optimal* group recovered compared to 59.7% in the *suboptimal* group. The odds of achieving a reliable recovery for those receiving their *optimal treatment* vs. those who received their *suboptimal treatment* was 1.35 (95% CI 1.15 to 1.58, p<0.001); 53.9% of patients in the *optimal* group recovered compared to 46.5% in the *suboptimal* group. The odds of achieving a percent MCID of a 20% improvement from baseline for those receiving their *optimal treatment* vs. those who received their *suboptimal treatment* was 1.37 (95% CI 1.15 to 1.64 p<0.001) with 74.2% of

patients in the *optimal* group showing changes of a clinically meaningful magnitude compared to 67.6% in the *suboptimal* group. The odds of achieving an absolute MCID for those receiving their *optimal treatment* vs. those who received their *suboptimal treatment* was 1.39 (95% CI 1.18 to 1.63 p<0.001); 63.8% of patients in the *optimal* group recovered compared to 56.0% in the *suboptimal* group.

When exploring the baseline characteristics of patients who were predicted to have better treatment responses in CBT, we found that they tended to be slightly older and have lower IMD, depression, anxiety, and functional impairment scores. This group also had a higher proportion of patient who self-referred, were employed, and were heterosexual. Furthermore, it had a higher proportion of patients who received LIT prior to HIT, had a long-term health condition, and were taking medication, relative to CFD group. Conversely, patients who were predicted to have better treatment responses in CFD tended to be slightly younger, as well as having higher IMD, depression, anxiety, and functional impairment scores. This group also had a higher proportion of patients who were referred from primary care, were not working, not heterosexual, and no previous LIT treatment. Furthermore, it had a higher proportion of patients who were not taking medication and no long-term health condition. Proportions of gender, ethnicity, disability status, diagnosis, and referral number appeared to be similar. Little difference in proportion of disability status or referral number were found.

5.6 Discussion

Electronic healthcare records were used to identify a cohort of patients receiving CBT or CFD for depressive symptoms in primary care settings. We investigated the benefit of differential treatment allocation on the basis of baseline characteristics. The results were validated in a held-out test sample. We found no evidence to suggest a main effect of treatment for CBT or CFD. However, we found some evidence to suggest that differential treatment allocation based on baseline characteristics can modestly improve outcomes. When allocated to their model-indicated optimal treatment, patients improved 1.8 points more on the PHQ-9 compared to patients who were allocated to their suboptimal treatment. This resulted in 4-10% more patients achieving favourable clinical outcomes. However, there were very few patients where the predicted difference between treatments was of a clinically meaningful magnitude at the individual level. However, benefits may nonetheless be meaningful from a public health perspective when applied at the population-level.

5.6.1 Discussion of findings

Similar to previous research, which compared CBT and counselling, we found no evidence of a main treatment effect of CBT vs. CFD in patients with depression.[30] However, previous research has shown that some patients can benefit if they are differentially allocated to CBT vs. CFD on the basis of baseline characteristics.[7] Previous research used a supervised machine learning algorithm to identify predictors separately within each treatment.[7] This approach of examining predictors separately in each treatment group is favourable, relative to testing for interactions, when sample sizes are smaller as there is insufficient power to assess moderating effects (i.e., to test for interactions). The present study tested for moderation in a larger sample, which has the benefit of additional power to assess differential effects of characteristics in different treatments. In the previous research study, 62.5% of patients experienced a reliable recovery if they were assigned to their optimal treatment whereas only 41.7% of patients achieved this if they were assigned to their suboptimal treatment (among the 30% of people who benefited from a differential treatment allocation).[7] This approximately 20% difference in improvement translated into post-treatment PHQ-9 differences in the range of approximately 1-2 points and effect sizes ranging from 0.16 to 0.33.[7] We found comparable benefits on the post-treatment PHQ-9 but much more modest improvements in reliable recovery. It was suggested that higher deprivation was associated with worse outcomes in CBT and better outcomes in CFD.[7] Ethnicity was found to only be a predictor in CBT, where ethnic minority groups had worse outcomes.[7] Higher baseline anxiety, lower outcome expectancy, longer chronicity, and not taking antidepressant medication were found to be associated with better outcomes in CFD only.[7] Our research suggests that only two variables were marginally statically significant moderators. Similar to previous the previous study we found some evidence to suggest that medication status is a moderator; however, contrary to previous research we also found that employment status was a moderator whereas this was found to be a general predictor in the previous study.[7] In our work, all other variables did not reach statistical significance when testing for effect modification. However, due to the previous study including additional variables only crude comparisons of variables can be made. Further research used a patient profiling algorithm to identify distinct groups of patients with specific profiles and examined differences in treatment response. It was found that certain patient profiles saw greater clinical improvements in CBT whereas other patient profiles appeared to benefit more from

counselling, although the point estimates for the latter groups had wider confidence intervals.[11]

We found no strong evidence to support the idea that any of the examined moderators produce meaningfully different clinical outcomes when examined individually. Perhaps surprisingly, we still observed benefits between patients who received their optimal vs. suboptimal treatment. This potentially suggest that no individual characteristic is sufficient to result in substantive effect modification, but rather a cumulative effect - small differences may add up across multiple characteristics. It should be noted that the benefits were only observed at the population level – almost no patients had a PAI score that reached the threshold of minimal clinically important difference.[32, 33, 35] This suggests that benefits may not be immediately tangible to every individual patient; rather, they appear to be relevant from a public health perspective where clinical outcomes are improved to a small degree but at scale. However, it should be noted that achieving a difference beyond the MCID, approximately 20% reduction from baseline, may be a relatively ambitious threshold given that both treatments are generally effective treatments.[32, 33]

5.6.2 Strengths & Limitations

The present study used a large, retrospective cohort of patients receiving treatment for depression in primary care, from multiple services across different geographic locations. This, in addition to the naturalistic settings, increases the external validity and generalisability of findings. Furthermore, we used pre-post treatment outcome measures which are favourable to retain power and account for measurement error. We additionally validated the model in an external test sample.

Despite the large, diverse sample, it is still possible that the heterogeneity which exists between services may nonetheless limit the generalisability to other services.[36] A further limitation of the present research is the observational nature. Unlike in randomised controlled trials, patients in routine clinical practice are not randomly allocated to treatments. We found differences in the baseline characteristics of patients between treatments which may suggest that patients with a higher clinical severity are more likely to receive CBT. We applied doubly robust propensity adjustment which has been established as performing well in electronic healthcare records.[29] However, adjustment can only be made for observed variables leaving the possibility of unmeasured confounding. Possible examples may include,

but are not limited to, mental health comorbidities,[37] childhood maltreatment,[38] cognitive biases, [39] competency in cognitive skills, [40] and shame.[41] Additionally, while all treatments in IAPT are delivered by mental health professional who are trained in accordance with the national curriculum, there are currently no measures of treatment fidelity making judgements about the adherence to treatment protocols difficult.[42] As such, the present study serves an explanatory exploration with more rigorous and causal research required prior to application in practice, such as evaluation in a randomised controlled trial.

A further consideration of the modelling approach is the split that was used to partition the data into a test and train dataset. In the present study, the data was randomly partitioned resulting in a test and train dataset that had very similar characteristics. However, there are multiple ways to select a test dataset. Examples could include specifically selecting a test dataset that has different characteristics to the training dataset or using a subset of data from service providers for training and then evaluating the model in the data from a service which was not included in the training dataset. Evaluating the model in such a way provides the benefit of further examining how the model generalises to unfamiliar contexts. This can result in larger or smaller effects, depending on the characteristics of the test dataset. However, the choice of characteristics to partition the data to make the datasets strategically different or the choice of service to evaluate in can be somewhat arbitrary. The approach used in the present study removes the somewhat arbitrary selection of the testing dataset but has the limitation of evaluating model performance in more similar contexts. The present approach may be more useful in circumstances where models are used in the settings that contributed to modelling, i.e., where services that contributed data to the models used these models as a clinical decision-making aid. Conversely, using an approach that evaluates the model in a test dataset with different characteristics may be more useful in settings where the model is intended to be used outside of its development context.

Furthermore, the present results are limited by the data quality of routinely collected data. Electronic healthcare records contain missing data and are collected by various clinicians across different services and years. We used robust methods of data imputation to address missingness but are unable to account for any systemic differences in data recording by individual therapists and/or services. We examine broad sociodemographic and clinical characteristics. Previous research has identified more detailed, psychological moderators of psychotherapy such as cognitive problems, attributional style, or interpersonal self-

sacrificing.[9] These, more in-depth psychological characteristics, may be promising moderators that produce greater differential improvements at the patient-level, but we were limited by data availability. Lastly, a pragmatic limitation is that the present research does not take organisational factors into account, such as treatment availability. Given the small benefit at the patient level, therapy that is available immediately may outweigh the benefit of waiting for the model indicated treatment. Such decisions would need to be weighed up by the treating clinician.

5.6.3 Implications

The present approach can potentially inform clinical decision-making. While the benefits of differential treatment allocation may not produce large effects and no characteristics emerged as strong moderators at the patient-level, they may still be relevant from a public health perspective when applied at scale. Currently, IAPT services in England receive approximately 1.7 million referrals a year.[4] As such, improvements in clinical outcomes ranging from approximately 4 -10% may still produce some benefit to a large number of patients. However, only randomised controlled trials can determine the true extent of the benefits that differential treatment allocation based on baseline characteristics have. A benefit of the present approach is that it comes at minimal cost and are easy to implement resulting in little burden to healthcare systems. Additionally, there is little risk concerning the implementation as patients receive one of two effective treatments.

Beyond the immediate implications, the present research touches on a debate in the current literature concerning the mechanisms by which therapy produces change. The debate focuses on whether these mechanisms are common and shared across therapeutic modalities or whether there are specific factors unique to different approaches.[43] The identification of differential outcomes in CBT vs. CFD based on baseline characteristics potentially suggests each may possess specific factors; however, our effects were modest. Furthermore, the finding that treatments appear to be equally effective and that most characteristics are stronger general predictors of response also speaks to the idea that various common factors are likely to exist. Our research suggests that common factors are likely to contribute to outcomes but that specific factors may also contribute to a small, but potentially clinically relevant, degree.[43] In order to have greater confidence in differential treatment allocation, an understanding of the mechanisms of depression as well as how treatments work is necessary. However, no clear consensus has been established in process research that attempts to

elucidate the mechanism of action that underpin psychotherapy to date.[44] This is further complicated by the fact that the evidence for mechanisms of depression remain unclear as well as the complexity of depression evidenced by the significant symptom heterogeneity.[45] This makes it difficult to reconcile depressive and therapeutic mechanism and moderator research to assess if they converge, at the very least on a theoretical basis. Future research investigating both the mechanisms of psychotherapy and well as the mechanisms of psychopathology will undoubtedly provide invaluable insights into efforts of matching patients to their optimal treatment.

5.6.4 Conclusion

The present research suggest that targeted allocation of psychotherapy based on baseline characteristics has the potential to personalise therapy; however, only to some degree. While the effects are modest at the patient-level, the impact from public health perspective may nonetheless be meaningful. The ease, minimal risk and low cost associated with the implementation of such models provides a simple way to support clinicians in clinical decision-making in the future. However, causal research is necessary to truly evaluate the magnitude of benefit. Furthermore, significant advances in personalised psychotherapy are likely dependant on advancements in the mechanistic understanding of psychopathology itself as well as how psychotherapy works in order to optimally match treatments to disease-specific processes.

5.7 Tables

Table 1. *Baseline characteristics of patients, stratified by treatment*

		Total Sample	Cognitive Behavioural Therapy	Counselling for Depression	Standardised Mean Difference
n		*19,269*	*14,544*	*4725*	
Age		40.25 (13.99)	39.59 (14.01)	42.31 (13.72)	0.196
Gender	Female	12,837 (66.6)	9479 (65.2)	3358 (71.1)	0.127
	Male	6432 (33.4)	5065 (34.8)	1367 (28.9)	
Ethnicity	White	15,401 (79.9)	11,983 (82.4)	3418 (72.3)	0.242
	BAME	3868 (20.1)	2561 (17.6)	1307 (27.7)	
Employment Status	Employed	10570 (54.9)	7766 (53.4)	2804 (59.3)	0.120
	Not working	8699 (45.1)	6778 (46.6)	1921 (40.7)	
Index of Multiple Deprivation		21.47 (11.85)	21.55 (12.12)	21.23 (10.99)	0.027
Sexual Orientation	Heterosexual	18,366 (95.3)	13,816 (95.0)	4550 (96.3)	0.064
	Not heterosexual	903 (4.7)	728 (5.0)	175 (3.7)	
Disability	No	16,449 (85.4)	12,415 (85.4)	4034 (85.4)	<0.001
	Yes	2820 (14.6)	2129 (14.6)	691 (14.6)	
Long-Term Health Condition	No	12,602 (65.4)	9448 (65.0)	3154 (66.8)	0.038

	Yes	6667 (34.6)	5096 (35.0)	1571 (33.2)	
Diagnosis	Depressive episode	14,325 (74.3)	10,288 (70.7)	4037 (85.4)	0.361
	Recurrent depressive disorder	4944 (25.7)	4256 (29.3)	688 (14.6)	
Baseline PHQ-9		18.24 (4.39)	18.52 (4.35)	17.40 (4.39)	0.255
Baseline GAD-7		14.34 (4.52)	14.52 (4.49)	13.80 (4.58)	0.159
Baseline WSAS		23.07 (8.56)	23.76 (8.38)	20.94 (8.76)	0.329
Psychotropic Medication	Yes	11,112 (57.7)	8879 (61.0)	2233 (47.3)	0.279
	No	8157 (42.3)	5665 (39.0)	2492 (52.7)	
Referral Source	Self	10,669 (55.4)	8607 (59.2)	2062 (43.6)	0.432
	Primary care	7373 (38.3)	4850 (33.3)	2523 (53.4)	
	Other	1227 (6.4)	1087 (7.5)	140 (3.0)	
Referral Number		1.79 (1.25)	1.86 (1.31)	1.56 (1.00)	0.252
Low-Intensity Therapy	No	14,007 (72.7)	10,169 (69.9)	3838 (81.2)	0.266
	Yes	5262 (27.3)	4375 (30.1)	887 (18.8)	

** PHQ-9: Patient Health Questionnaire (9-item); GAD-7: Generalised Anxiety Disorder Scale (7-item); WSAS: Work and Social Adjustment Scale. Continuous data are presented as mean (standard deviation) and categorical data are presented as n (%).*

Table 2. *Illustration of moderating effects for baseline characteristics on predicted post-treatment PHQ-9 scores in Cognitive Behavioural Therapy vs. Counselling for Depression with other covariates held constant*

		Cognitive Behaviour Therapy			**Counselling for Depression**			
		Beta	***95% Confidence Intervals***		***Beta***	***95% Confidence Intervals***		***p-value***
Age	28	*9.50*	9.06	9.94	*9.56*	8.90	10.21	0.116
	50	*8.70*	8.28	9.11	*9.10*	8.47	9.73	
Gender	Female	*9.05*	8.64	9.46	*9.30*	8.69	9.91	0.679
	Male	*9.16*	8.71	9.61	*9.52*	8.82	10.22	
Ethnicity	White	9.05	8.64	9.46	*9.30*	8.69	9.91	0.205
	BAME	9.39	8.86	9.92	*10.02*	9.27	10.77	
Employment Status	Employed	*9.05*	8.64	9.46	*9.30*	8.69	9.91	**0.050**
	Not working	*10.97*	10.52	11.42	*10.69*	10.00	11.37	
Index of Multiple Deprivation	12	*8.81*	8.39	9.23	*9.18*	8.55	9.81	0.237
	29	*9.25*	8.83	9.67	*9.39*	8.76	10.02	
Sexual Orientation	Heterosexual	*9.05*	8.64	9.46	*9.30*	8.69	9.91	0.262
	Not heterosexual	*9.56*	8.88	10.23	*9.08*	7.78	10.37	
Disability	No	*9.05*	8.64	9.46	*9.30*	8.69	9.91	0.924
	Yes	*9.73*	9.18	10.28	*9.94*	9.06	10.82	
Long-Term Condition	No	*9.05*	8.64	9.46	*9.30*	8.69	9.91	0.603
	Yes	*9.62*	9.17	10.06	*10.01*	9.32	10.70	
Diagnosis	Depressive episode	*9.05*	8.64	9.46	*9.30*	8.69	9.91	0.686
	Recurrent depressive disorder	*9.35*	8.85	9.85	*9.46*	8.63	10.29	
Baseline PHQ-9	15	*7.87*	7.45	8.29	*8.20*	7.57	8.84	0.466
	22	*10.43*	9.99	10.87	*10.57*	9.91	11.24	
Baseline GAD-7	11	*8.62*	8.19	9.04	*8.86*	8.23	9.50	0.982

	18	*9.53*	9.10	-	9.96	*9.78*	9.13	-	10.42	
Baseline WSAS	17	*8.60*	8.18	-	9.02	*8.92*	8.30	-	9.54	0.463
	30	*9.57*	9.13	-	10.00	*9.73*	9.08	-	10.38	
Psychotropic Medication	Yes	*9.05*	8.64	-	9.46	*9.30*	8.69	-	9.91	**0.049**
	No	*8.59*	8.18	-	9.00	*8.32*	7.72	-	8.92	
Referral Source	Self	*9.05*	8.64	-	9.46	*9.30*	8.69	-	9.91	0.211
	Primary Care	*9.71*	9.24	-	10.18	*9.58*	8.96	-	10.21	
	Other	*10.02*	9.42	-	10.63	*10.83*	9.48	-	12.18	
Referral Number	1	*8.81*	8.39	-	9.22	*9.05*	8.42	-	9.68	0.983
	2	*9.12*	8.71	-	9.53	*9.37*	8.76	-	9.98	
Low-Intensity Therapy	No	*9.05*	8.64	-	9.46	*9.30*	8.69	-	9.91	0.205
	Yes	*9.24*	8.71	-	9.76	*9.88*	9.07	-	10.70	

**PHQ-9: Patient Health Questionnaire (9-item); GAD-7: Generalised Anxiety Disorder Scale (7-item); WSAS: Work and Social Adjustment Scale. For each characteristic, the treatment and one moderator are varied while all other baseline characteristics are held constant at the mean for continuous variables or most common level for categorical variables to illustrate the magnitude of effect modification.*

Table 3. *Evaluation of a data-driven treatment allocation model in held-out test sample*

	Optimal Treatment	**Suboptimal Treatment**	**Beta/Odds Ratio**	**95% Confidence Intervals**	**p-value**
n	*1247*	*1162*			
Post-treatment PHQ-9	9.63 (6.95)	11.42 (7.49)	-1.78	-2.36 - -1.21	<0.001
Recovery	748 (60.0)	577 (49.7)	1.52	1.29 - 1.79	<0.001
Reliable Change	796 (63.8)	694 (59.7)	1.19	1.01 - 1.40	0.038
Reliable Recovery	672 (53.9)	540 (46.5)	1.35	1.15 - 1.58	<0.001
MCID (%)	925 (74.2)	786 (67.6)	1.37	1.15 - 1.64	<0.001
MCID (absolute)	796 (63.8)	651 (56.0)	1.39	1.18 - 1.63	<0.001

**PHQ-9: Patient Health Questionnaire (9-item); MCID: Minimal Clinically Important Difference.*

5.8 References

1. National Institute for Health and Care Excellence. *Depression in adults: recognition and management (CG90)*. National Institute for Health and Care Excellence, 2009 (https://www.nice.org.uk/guidance/cg90/resources/depression-in-adults-recognition-and-management-pdf-975742636741).

2. Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychother Res*. 2020; **30**: 279-293. Available from: https://doi.org/10.1080/10503307.2019.1649732.

3. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *J Affect Disord*. 2014; **159**: 118-126. Available from: https://doi.org/10.1016/j.jad.2014.02.026.

4. NHS Digital. *Psychological Therapies - Annual report on the use of IAPT services, England 2019-20.* NHS Ditigal, 2020 (https://files.digital.nhs.uk/B8/F973E1/psych-ther-2019-20-ann-rep.pdf).

5. Cuijpers P. Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Can Psychol*. 2017; **58**: 7-19. Available from: https://doi.org/10.1037/cap0000096.

6. Cuijpers P, Ebert DD, Acarturk C, Andersson G, Cristea IA. Personalized Psychotherapy for Adult Depression: A Meta-Analytic Review. *Behav Ther*. 2016; **47**: 966-980. Available from: https://doi.org/10.1016/j.beth.2016.04.007.

7. Delgadillo J, Gonzalez Salas Duhne P. Targeted prescription of cognitive–behavioral therapy versus person-centered counseling for depression using a machine learning approach. *J Consult Clin Psychol*. 2020; **88**: 14-24. Available from: https://doi.org/10.1037/ccp0000476.

8. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One*. 2014; **9**: e83875. Available from: https://doi.org/10.1371/journal.pone.0083875.

9. Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, et al. Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. *PLoS One*. 2015; **10**: e0140771. Available from: https://doi.org/10.1371/journal.pone.0140771.

10. Cohen ZD, Kim TT, Van HL, Dekker JJM, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychother Res*. 2020; **30**: 137-150. Available from: https://doi.org/10.1080/10503307.2018.1563312.

11. Saunders R, Buckman JEJ, Pilling S. Latent variable mixture modelling and individual treatment prediction. *Behav Res Ther*. 2020; **124**: 103505. Available from: https://doi.org/10.1016/j.brat.2019.103505.

12. Clark D. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *Int Rev Psychiatry*. 2011; **23**: 318-327. Available from: https://doi.org/10.3109/09540261.2011.606803.
13. Beck AT. The past and future of cognitive therapy. *J Psychother Pract Res*. 1997; **6**: 276-284.

14. Department of Health and Social Care. *National curriculum for Cognitive Behavioural Therapy courses*. Department of Health and Social Care, 2011 (https://www.uea.ac.uk/documents/746480/2855738/national-curriculum-for-high-intensity-cognitive-behavioural-therapy-courses.pdf).

15. Hill A. *Curriculum for Counselling for Depression: Continuing professional development for qualified therapists delivering high intensity interventions*. National IAPT Programme Team, 2011 (https://webarchive.nationalarchives.gov.uk/20160302160209/http://www.iapt.nhs.uk/silo/files/curriculum-for-counselling-for-depression.pdf).

16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; **16**: 606-613. Available from: https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

17. McLennan D, Noble S, Noble M, Plunkett E, Wright G, N G. *The English Indices of Deprivation 2019: technical report.* Ministry of Housing, Communities, and Local Government, 2019 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833951/IoD2019_Technical_Report.pdf).

18. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006; **166**: 1092-1097. Available from: https://doi.org/10.1001/archinte.166.10.1092.

19. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry*. 2002; **180**: 461-464. Available from: https://doi.org/10.1192/bjp.180.5.461.

20. Delgadillo J, Moreea O, Lutz W. Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Beh Res Ther* 2016; **79**: 15-22. Available from: https://doi.org/10.1016/j.brat.2016.02.003.

21. Green SA, Honeybourne E, Chalkley SR, Poots AJ, Woodcock T, Price G, et al. A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health programme. *BMJ Open*. 2015; **5**: e006103. Available from: https://doi.org/10.1136/bmjopen-2014-006103.

22. Amati F, Banks C, Greenfield G, Green J. Predictors of outcomes for patients with common mental health disorders receiving psychological therapies in community settings: a

systematic review. *J Public Health*. 2017; **40**: e375-e387. Available from: https://doi.org/10.1093/pubmed/fdx168.

23. Rimes KA, Ion D, Wingrove J, Carter B. Sexual orientation differences in psychological treatment outcomes for depression and anxiety: National cohort study. *J Consult Clin Psychol*. 2019; **87**: 577. Available from: https://doi.org/https://doi.org/10.1037/ccp0000416.

24. Cairns M. Patients who come back: Clinical characteristics and service outcome for patients re-referred to an IAPT service. *Couns Psychother Res*. 2014; **14**: 48-55. Available from: https://doi.org/https://doi.org/10.1080/14733145.2013.770895.

25. Delgadillo J, Asaria M, Ali S, Gilbody S. On poverty, politics and psychology: the socioeconomic gradient of mental healthcare utilisation and outcomes. *Br J Psychiatry*. 2016; **209**: 429-430. Available from: https://doi.org/10.1192/bjp.bp.115.171017.

26. R Core Team. R: A Language and Environment for Statistical Computing.Vienna, Austria: R Foundation for Statistical Computing, 2020. Available from: https://www.R-project.org/

27. Stuart EA, Lee BK, Leacy FP. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *J Clin Epidemiol*. 2013; **66**: S84-S90.e81. Available from: https://doi.org/10.1016/j.jclinepi.2013.01.013.

28. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012; **28**: 112-118. Available from: https://doi.org/10.1093/bioinformatics/btr597.

29. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *J Am Coll Cardiol*. 2017; **69**: 345-357. Available from: https://doi.org/10.1016/j.jacc.2016.10.060

30. Pybis J, Saxon D, Hill A, Barkham M. The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2nd UK National Audit of psychological therapies. *BMC Psychiatry*. 2017; **17**: 215. Available from: https://doi.org/10.1186/s12888-017-1370-7.

31. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*. 2002; **59**: 877-883. Available from: https://doi.org/10.1001/archpsyc.59.10.877.

32. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, et al. Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychol Med*. 2015; **45**: 3269-3279. Available from: https://doi.org/10.1017/s0033291715001270.

33. Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care.

Psychol Med. 2020; **52**: 1875-1882. Available from: https://doi.org/10.1017/S0033291720003700.
34. David C, Oates M. *Improving Access to Psychological Therapies - Measuring Improvement and Recovery Adult Services* NHS England, 2014 (http://www.oxfordahsn.org/wp-content/uploads/2015/11/measuring-recovery-2014.pdf).

35. Bauer-Staeb C, Kounali D-Z, Welton NJ, Griffith E, Wiles NJ, Lewis G, et al. Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials. *J Clin Epidemiol*. 2021; **137**: 200-208. Available from: https://doi.org/10.1016/j.jclinepi.2021.04.002.

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

37. Newton-Howes G, Tyrer P, Johnson T. Personality disorder and the outcome of depression: Meta-analysis of published studies. *Br J Psychiatry*. 2006; **188**: 13-20. Available from: https://doi.org/10.1192/bjp.188.1.13.

38. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry*. 2012; **169**: 141-151. Available from: https://doi.org/10.1176/appi.ajp.2011.11020335.

39. Hollon SD, DeRubeis RJ, Evans MD. Causal mediation of change in treatment for depression: discriminating between nonspecificity and noncausality. *Psychol Bull*. 1987; **102**: 139-149.

40. Strunk DR, DeRubeis RJ, Chiu AW, Alvarez J. Patients' competence in and performance of cognitive therapy skills: relation to the reduction of relapse risk following treatment for depression. *J Consult Clin Psychol*. 2007; **75**: 523-530. Available from: https://doi.org/10.1037/0022-006x.75.4.523.

41. Nikolić M, Hannigan LJ, Krebs G, Sterne A, Gregory AM, Eley TC. Aetiology of shame and its association with adolescent depression and anxiety: results from a prospective twin and sibling study. *J Child Psychol Psychiatry*. 2022; **63**: 99-108. Available from: https://doi.org/10.1111/jcpp.13465.

42. Martin C, Iqbal Z, Airey ND, Marks L. Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: How can it better meet the range of primary care mental health needs? *Br J Clin Psychol*. 2022; **61**: 157-174. Available from: https://doi.org/10.1111/bjc.12314.

43. Mulder R, Murray G, Rucklidge J. Common versus specific factors in psychotherapy: opening the black box. *Lancet Psychiatry*. 2017; **4**: 953-962. Available from: https://doi.org/10.1016/s2215-0366(17)30100-1.

44. Cuijpers P, Reijnders M, Huibers MJH. The Role of Common Factors in Psychotherapy Outcomes. *Annu Rev Clin Psychol*. 2019; **15**: 207-231. Available from: https://doi.org/10.1146/annurev-clinpsy-050718-095424.

45. Malhi GS, Mann JJ. Depression. *The Lancet*. 2018; **392**: 2299-2312. Available from: https://doi.org/10.1016/S0140-6736(18)31948-2.

5.9 Declaration of Interests

None.

5.10 CReDiT Author Statement

Clarissa Bauer-Staeb: conceptualisation, methodology, formal analysis, writing – original draft. **Emma Griffith**: writing – review & editing, supervision. **Julian J. Faraway:** conceptualisation, methodology, formal analysis, writing – review & editing, supervision. **Katherine S. Button**: conceptualisation, methodology, writing – review & editing, supervision.

5.11 Acknowledgements

The data extraction and anonymisation was kindly performed by Mayden, the developers of iaptus – a patient management software used within IAPT.

5.12 Funding

This study was funded by the University of Bath PhD Research Programme Studentship awarded to CBS.

5.13 Ethics Statement

The research received approval from the University of Bath Psychology Research Ethics Committee (19-015).

6 Trajectories of depression and generalised anxiety symptoms over the course of Cognitive Behavioural Therapy in primary care: An observational, retrospective cohort

6.1 Statement of Authorship

This declaration concerns the article entitled:			
Trajectories of depression and generalised anxiety symptoms over the course of Cognitive Behavioural Therapy in primary care: An observational, retrospective cohort			
Publication status **Draft manuscript** [ ] **Submitted** [ ] **In review** [ ] **Accepted** [ ] **Published** [X]			
Publication details	Bauer-Staeb, C., Griffith, E., Faraway, J. J., & Button, K. S. (2022). Trajectories of depression and generalised anxiety symptoms over the course of cognitive behaviour therapy in primary care: an observational, retrospective cohort. *Psychological Medicine*, 1-9. https://doi.org/10.1017/S0033291722001556		
Copyright status (tick the appropriate statement) The material has been published with a CC-BY license [ ] The publisher has granted permission to replicate the material included here [X]			
Candidate's contribution to the paper	The candidate predominantly executed the formulation of ideas (85%) and the design of methodology (80%). The candidate predominantly made arrangements for the access to the secondary data used (75%) and conducted the formal analysis (90%). The candidate predominantly executed the presentation of data in journal format (85%).		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	[redacted]	**Date**	19/08/2022

Trajectories of depression and generalised anxiety symptoms over the course of Cognitive Behavioural Therapy in primary care: An observational, retrospective cohort

Clarissa Bauer-Staeb,[1] Emma Griffith,[1,2] Julian J. Faraway[3], & Katherine S. Button.[1]

[1] Department of Psychology, University of Bath, Bath, UK

[2] Avon and Wiltshire NHS Mental Health Partnership Trust

[3] Department of Mathematical Sciences, University of Bath, Bath, UK

6.2 Abstract

Background. Cognitive behavioural therapy (CBT) has been shown to be an effective treatment for depression and anxiety. However, most research has focused on the sum scores of symptoms. Relatively little is known about how individual symptoms respond.
Methods. Longitudinal models were used to explore how depression and generalised anxiety symptoms behave over the course of CBT in a retrospective, observational cohort of patients from primary care settings (n = 5306). Logistic mixed models were used to examine the probability of being symptom-free across CBT appointments, using the 9-item Patient Health Questionnaire and the 7-item Generalised Anxiety Disorder Scale as measures.
Results. All symptoms improved across CBT treatment. The results suggest that *low mood/hopelessness* and *guilt/worthlessness* improved quickest relative to other depressive symptoms, *with sleeping problems, appetite changes,* and *psychomotor retardation/agitation* improving relatively slower. *Uncontrollable worry* and *too much worry* were the anxiety symptoms that improved fastest; *irritability* and *restlessness* improved the slowest.
Conclusions. This research suggests there is a benefit to examining symptoms rather than sum scores alone. Investigations of symptoms provide the potential for precision psychiatry and may explain some of the heterogeneity observed in clinical outcomes when only sum scores are considered.
Funding. This work was conducted as part of a PhD studentship from the University of Bath awarded to Clarissa Bauer-Staeb.

6.3 Introduction

Depression and anxiety are leading causes of disability worldwide (GBD 2019 Mental Disorders Collaborators, 2022). While effective interventions exist, the current treatment success leaves room for improvement (Cuijpers, van Straten, Bohlmeijer, Hollon & Andersson, 2010). Much of the research efforts evaluating treatments to date have focused on sum scores. When sum scores are used, individual symptoms are assumed to be equivalent in value as common indicators of an underlying disorder. Subsequently, individual symptoms have received comparably little attention (Fried & Nesse, 2015a). However, there may be important insights to be gained from better understanding individual symptoms and how they respond over the course of treatment (Fried & Nesse, 2015a).

Depression and anxiety are heterogeneous disorders. For example, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) lists nine depressive symptoms (American Psychiatric Association, 2013). In order to meet the criteria for a diagnosis of depression, patients have to meet at least five of the nine symptom criteria, including at least one of the core symptoms - depressed mood or diminished interest/pleasure (American Psychiatric Association, 2013). These criteria include compounded symptoms that are grouped together, such as worthlessness or inappropriate guilt, as well as symptoms that lie on the opposite spectrum, such as psychomotor agitation or retardation. Given the multitude of possible symptom combinations for depression alone, it is perhaps unsurprising that considerable symptom profile variability has been reported, with few exact symptom combinations being shared between patients (Fried & Nesse, 2015b). When considering individual symptoms, they do not appear to be equal, with evidence suggesting that individual symptoms may vary in the degree to which they contribute to functional impairment (Fried & Nesse, 2014), their risk factors (Fried, Nesse, Zivin, Guille, & Sen, 2014), and their heritability (Jang, Livesley, Taylor, Stein, & Moon, 2004). Furthermore, they may differ in their response to different treatments (Bekhuis et al., 2018; Boschloo et al., 2019a) and in terms of their association with clinical outcomes (O'Driscoll et al., 2021). High variability in the presence of symptoms, combined with evidence suggesting that symptoms are not interchangeable, implies that rather than being clearly defined and bounded disorders, depression and anxiety have a complex and heterogenous presentation. Sum scores, which assume equivalence of symptoms, may therefore hide important differences.

Outcome research for depression and anxiety is further complicated by the fact that psychotherapies such as cognitive behavioural therapy (CBT) are often considered to be a '*black box'* (Huibers, Lorenzo-Luaces, Cuijpers, & Kazantzis, 2021). They commonly target a broad group of symptoms (Eronen, 2020) and identifying how treatments work and for whom is arguably the main scientific challenge facing depression and anxiety outcome researchers (Paul, 1967; Carey, Griffiths, Dixon, & Hines, 2020; Huibers, Lorenzo-Luaces, Cuijpers, & Kazantzis, 2021). Despite multiple putative mechanisms being proposed, that reflect both 'active ingredients' of specific psychotherapies and non-specific effects shared across psychotherapies, the evidence remains ambiguous, with no universal consensus concerning specific mechanisms of change (Carey, Griffiths, Dixon, & Hines, 2020; Huibers, Lorenzo-Luaces, Cuijpers, & Kazantzis, 2021).

Exploring how symptoms respond to treatment, rather than sum scores alone, may lead to important insights and provide the first steps towards precision psychiatry. If specific therapeutic interventions could be mapped onto the change in specific symptoms, it may begin to explain how psychotherapies work and for whom. More pragmatically, perhaps, understanding possible symptom-specific effects of treatments could also potentially lead to a more personalised matching of symptom profiles to treatments as well as having the potential to explain heterogeneity in clinical outcomes. Some treatments may cause side effects that increase some of the very symptoms used to measure depression, such as antidepressant medication and the possible side effect of sleep disturbances (Wichniak, Wierzbicka, Walęcka & Jernajczyk, 2017). These could potentially mask or dilute specific benefits on other symptoms when sum scores are used (Fried & Nesse, 2015a).

Overall, there is a strong argument to be made around exploring symptom-specific effects. The evidence suggests that CBT is an effective treatment for depression and anxiety (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016). However, most evaluations of CBT focus on sum scores. As such, less is known about the symptom-specificity of CBT. In the present study, we aim to explore how individual symptoms of depression and generalised anxiety behave across a course of CBT in an observational, retrospective cohort by examining symptom trajectories and how they compare to one another.

6.4 Methods

6.4.1 Settings

Improving Access to Psychological Therapies (IAPT) is a national programme that delivers psychological therapies for depression and anxiety across England in primary care settings (Clark, 2011; Clark et al., 2018). IAPT offers a variety of psychological therapies, including both low-intensity therapy (LIT) and high-intensity therapy (HIT) (Clark, 2011; Clark et al., 2018). IAPT has implemented routine outcome monitoring, where detailed information is gathered about patients, their treatment, and their clinical outcomes (Clarke et al., 2011). These data are collected on a session-by-session basis to increase complete-case recording (Clark et al., 2018).

6.4.2 Sample

The clinical records for the present study were obtained from ten IAPT practices across the southwest of England and London, who consented to share their data via Mayden, the providers of a patient management software used in IAPT, for the purposes of this research. The data from participating services were extracted and fully anonymised by Mayden before sharing with us for processing and analysis. The present analysis contained data for referrals from the year 2014 to mid-2019 – earlier years were excluded due to more operational issues early in IAPT, such as missing or inappropriate diagnostic labels (Saunders et al., 2020). These lead to poorer outcomes as a result of inappropriate matching of treatment protocols to clinical needs of the patient (Saunders et al., 2020).

Referrals were included if they had received high-intensity CBT for depression or generalised anxiety. Patients receiving treatment for depression and generalised anxiety were identified by their diagnostic labels of having a depressive episode, recurrent depressive disorder, or generalised anxiety disorder as well as having their primary outcome measures as the 9-item Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer & Williams, 2001) and 7-item Generalised Anxiety Disorder (GAD-7) Scale (Spitzer, Kroenke, Williams, & Löwe, 2006), which are used to monitor treatment process and evaluate treatment response - other anxiety measures indicate treatment for different anxiety disorders. Treatment was defined as CBT if all the recorded treatment labels were labelled as CBT. Referrals receiving other therapies in addition to CBT were excluded due to difficulties isolating the effects of CBT. In order to examine the effect of CBT on specific questionnaire items, only patients who had at least one

full set of item-level questionnaire scores recorded across any appointment were included in the analyses. Due to the observational nature of the data, no fixed amount of treatment is provided. To counteract the issue of unequal treatment doses, a minimum treatment dose was defined as eight appointments – the overall average number of delivered appointments in IAPT (NHS Digital, 2021). Eight appointments may also be more likely to at least partially capture some of the later responses that are observed for some subgroups of patients in therapy, which can occur approximately after six appointments (Saunders, Buckman, Cape, Fearon, Leibowitz, & Pilling, 2019). As such, all subsequent analyses examine the first eight appointments from each referral that had at least eight appointments. To minimise the impact of previous treatment effects, each patient's first referral, that met the criteria above, was used. See supplementary table D1 for a flowchart of the sample selection.

6.4.3 Intervention

CBT was delivered by mental health professionals trained in accordance with the national curriculum (Department of Health, 2019). CBT is a time-limited, structured, and problem-focused psychological therapy (Fenn & Byrne, 2013). CBT explores the links between cognitions, emotions, and behaviours. It further supports patients in identifying and modifying less helpful cognitions and behaviours and in developing alternative, more adaptive ones (Fenn & Byrne, 2013).

6.4.4 Measures

6.4.4.1 Patient Health Questionnaire 9-Item (PHQ-9)

The PHQ-9 is a 9-item self-report questionnaire that assesses the severity of depressive symptoms over the past two weeks (Kroenke, Spitzer & Williams, 2001). Each item is rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). Patients are asked to rate the severity of symptoms relating to: 1) *anhedonia*, 2) *low mood/hopelessness*, 3) *sleeping problems* 4) *tired/low energy* 5) *appetite*, 6) *guilt/worthlessness*, 7) *concentration*, 8) *psychomotor retardation/agitation*, and 9) *suicidal thoughts*.

6.4.4.2 Generalised Anxiety Disorder 7-Item Scale (GAD-7)

The GAD-7 is a 7-item self-report questionnaire assessing the severity of generalised anxiety over the past two weeks (Spitzer, Kroenke, Williams, & Löwe, 2006). Each item is rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). Patients are asked

to rate the severity of symptoms relating to: 1) *nervous/anxious*, 2) *uncontrollable worry*, 3) *too much worry*, 4) *trouble relaxing*, 5) *restlessness*, 6) *irritability*, and 7) *fear*.

6.4.5 Statistical Analysis

All analyses were performed in the R statistical programming language (R Core Team, 2013). We used a generalised logistic mixed model to examine how items of the PHQ-9 and GAD-7 responded over the course of CBT appointments using the glmer function from the "lme4" package (Bates, Maechler, Bolker & Walker, 2014). To simplify modelling, the item scores were dichotomised to model the probability of being symptom-free, where a score of 0 denotes being symptom-free and scores of 1-3 denote having symptoms. The default call for glmer() uses Gaussian-Hermite quadrature, which we used to estimate the model parameters but is computationally expensive. To obtain the confidence intervals we used bootstrap computation with the approximation nAGQ=0 (Adaptive Gauss-Hermite Quadrature) to make the calculation feasible.

Two separate models were built for patients being treated for depression and generalised anxiety, using the PHQ-9 and GAD-7 respectively as outcome measures. All dichotomised item scores for each questionnaire were specified as the outcome, with the appointments and question items being specified as predictors with an interaction term. As such, all PHQ-9 items were analysed in one model and all GAD-7 items were analysed in a separate model. The models were adjusted for baseline characteristics: gender, age, ethnicity, employment status, Index of Multiple Deprivation (McLennan, Noble, Noble, Plunkett, Wright & Gutacker, 2019), disability status, long-term health condition status, diagnosis, sum score baseline PHQ-9 (Kroenke, Spitzer & Williams, 2001), sum score baseline GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006), sum score baseline Work and Social Adjustment Scale (Mundt, Marks, Shear, & Greist, 2002), medication status, referral number, referral source, service and year. IAPT services collect various information about patients at baseline in a core dataset, with some services collecting additional information. Baseline characteristics were chosen on the data availability such as variables that were most consistently recorded across services, data quality of the recorded variables, and informed by variables that were considered in previous literature using IAPT records (see for example Delgadillo, Moreea & Lutz, 2016; Delgadillo & Gonzalez Salas Duhne, 2020; Green, Honeybourne, Chalkley, Poots, Woodcock, Price, et al., 2015; Saunders, Buckman, & Pilling, 2020). Missing baseline covariates were singly imputed using random forest with "missForest" (Stekhoven, &

Bühlmann, 2012). The cost of computation and the complexity of the model made the preferred option of multiple imputation impractical. Single imputation is adequate here because a) the small amount of missing baseline data (maximum 7.4%), b) evidence that missForest is effective in similar scenarios (Stekhoven, & Bühlmann, 2012; Waljee, Mukherjee, Singal, Zhang, Warren, Balis, et al., 2013) and c) satisfactory out-of-bag error suggest the imputation was successful with a normalized root mean square error of 0.39 for continuous variables and a proportion of falsely classified of 0.15 for categorical variables. Random intercepts were specified in all models for referral id and appointment id to account for repeated observations.

As there is no natural reference group on questionnaires which would provide an appropriate comparator to measure all other items against, sum coding was used. Sum coding allows the trajectory of each individual question to be compared to the mean trajectory of all other questions. Thus, providing a comparison of how specific trajectories compare relative to all others. Variability estimates for all questions relative to other questions are provided, except the last question of each questionnaire, as these are estimated by the inverse of the sum of the log odds of all other questions. Longitudinal models were used to predict the probability of being symptom-free on the questionnaire items across CBT treatment appointments.

Predictions were made for a hypothetical "average" patient with continuous variables fixed at the mean and categorical covariates set to the most frequently occurring group. The first eight appointments were used as a basis for predictions of up to 20 and 15 appointments to depict how item trajectories might respond for the National Institute for Clinical Excellence-recommended durations of treatment for depression and generalised anxiety, respectively (National Institute for Clinical Excellence, 2009; National Institute for Clinical Excellence, 2011). Bootstraps with 5000 simulations were used to approximate 95% confidence intervals for predictions. The "ggplot2" package was used to visualise questionnaire item trajectories (Wickham, 2011).

6.5 Results

6.5.1 Baseline Characteristics

Of the 5306 referrals included in the main analysis, the majority of patients were female, White, with an average age of 38 years, experiencing severe depression and moderate/severe

anxiety (Table 1). See supplementary table D2 for mean item-level depression and anxiety scores.

*[*Table 1. *Baseline clinical and sociodemographic patient characteristics]*

All PHQ-9 and GAD-7 symptoms appeared to improve across CBT appointments. However, the rate at which they improved relative to the average rate of improvement across all other questionnaire items varied.

*[*Table 2. *Effects of appointment on the odds of being symptom-free for item-level scores stratified by questionnaire]*

6.5.2 Patient Health Questionnaire 9-Item

We found the strongest evidence that PHQ-9 items 2 (*low mood/hopelessness*) and 6 (*guilt/worthlessness*) improved fastest across CBT appointments in comparison to the average rate of all questionnaire items (Table 2). Patients had 7% higher odds of having no l*ow mood/hopelessness* with every appointment compared to all other depression symptoms, or 56% across eight appointments. Patients had 6% higher odds of having no symptoms of *guilt/worthlessness* with every appointment compared to all other depression symptoms, or 48% across eight appointments. We also found the strongest evidence that PHQ-9 items 3 (*sleeping problems*), 5 (*appetite*), and 8 (*psychomotor retardation/agitation*) improved at a slower rate across CBT appointments in comparison to the average response of all other questionnaire items. Patients had 6%, 7%, and 6% lower odds of having no symptoms with every additional CBT appointment on *sleeping problems*, *appetite* and *psychomotor retardation/agitation* respectively, compared to all other depression symptoms; or 48%, 56% and 48% across eight appointments. Some evidence suggested that the odds of being symptom-free were higher for PHQ-9 item 1 (*anhedonia*) and PHQ-9 item 4 (*tired/low energy*); however, these associations were weaker. PHQ-9 item 7 (*concentration*) and PHQ-9 item 9 (*suicidal thoughts*) did not appear to differ in their rate of improvement compared to the average rate. See supplementary table D3 for the odds ratios of model covariates.

Figure 1 shows that the probability of being symptom free has different starting points for the PHQ-9 items. The predicted probability of being symptom free is much higher at baseline for PHQ-9 items 8 (*psychomotor retardation/agitation*) and 9 (s*uicidal thoughts*). This is a result

of the distribution of the response scores on each item at baseline, with items 8 (*psychomotor retardation/agitation*) and 9 (s*uicidal thoughts*) having a higher frequency of 0 scores at baseline and throughout treatment.

*[*Figure 2. *Predicted probability of being symptom-free on individual Patient Health Questionnaire-9 items across Cognitive Behavioural Therapy appointments]*

6.5.3 Generalised Anxiety Disorder 7-Item Scale

We found the strongest evidence that GAD-7 items 2 (*uncontrollable worry*) and 3 (*too much worry*) improved fastest across CBT in comparison to the mean response of all questionnaire items (Table 2). For both items, patients had 9% higher odds of having no symptoms compared to the average of all other anxiety symptoms, or 72% across eight appointments. We also found the strongest evidence that GAD-7 items 5 (*restlessness*) and 6 (*irritability*) improved slower across CBT appointments in comparison to all other questionnaire items. Patients were 11% and 7% less likely to have no symptoms of *restlessness* and *irritability*, respectively, compared to all other anxiety symptoms; or 88% and 56% across eight appointments. There was some evidence to suggest that the odds of being symptom-free was higher for GAD-7 item 1 (*nervous/anxious*); however, this association was weak. While no variability estimates are computed for the last question, it appears that the GAD-7 item 7 (*fear*) improved at a slower pace compared to all other questions. GAD-7 item 4 (*trouble relaxing*) did not appear to differ in the rate of improvement compared to the average rate. See supplementary table D4 for the odds ratios of model covariates.

Figure 2 shows that the probability of being symptom-free has different starting points for the GAD-7 items. The predicted probability of being symptom-free is somewhat higher at baseline for GAD-7 items 5 (*restlessness*), 6 (*irritability*), and 7 (*fear*). This is a result of the distribution of the response scores on each item at baseline, with these items having a somewhat higher frequency of 0 scores at baseline and throughout treatment.

*[*Figure 3. *Predicted probability of being symptom-free on individual Generalised Anxiety Disorder-7 items across Cognitive Behavioural Therapy appointments]*

We performed a sensitivity analysis in a subset of patients who had all their item-level scores recorded and found effects of a similar magnitude and direction (supplementary tables D5

and D6). We further examined whether the slope of the trajectories for each item varied by baseline medication status and found no evidence of a differential improvement across items by medication status (PHQ-9: p = 0.769 and GAD-7: p=0.561). Furthermore, we assessed whether there were inherent differences in baseline patient characteristics amongst those who had item-level data recorded and those who did not but found no evidence to suggest that this was the case, with all standardised mean differences between baseline characteristics falling < 0.25 (Rubin, 2001; Panos, & Mavridis, 2020).

6.6 Discussion

We examined the trajectories of individual depression and anxiety symptoms as measured by two widely used measures, the PHQ-9 and the GAD-7, in a large, retrospective, observational cohort of patients receiving CBT. We found evidence to suggest *low mood/hopelessness* and *guilt/worthlessness* on the PHQ-9, as well as *uncontrollable worry* and *too much worry* on the GAD-7, improved at a faster rate relative to other symptoms. We found that *sleeping problems, appetite,* and *psychomotor agitation/retardation* on the PHQ-9 and *restlessness* and *irritability* on the GAD-7 improved at a slower rate compared to the average response of all other symptoms.

Worry is clinically central to generalised anxiety (National Institute for Clinical Excellence, 2011; American Psychiatric Association, 2013) and working with worry and tolerance of uncertainty forms a critical part of treatment protocols for generalised anxiety disorder (University College London, n.d; Department of Health, 2019). Low mood is considered one of the core clinical features of depression (National Institute for Clinical Excellence, 2009; American Psychiatric Association, 2013) and has been shown to be the biggest contributor of depressive symptoms to functional impairment, explaining ~ 20% of the variance, with self-blame accounting for ~ 6% (Fried & Nesse, 2014). Various cognitive and behavioural components are delivered during CBT for depression, commonly starting with behavioural work in severe depression (Department of Health, 2019). However, working with negative automatic thoughts and themes of guilt or self-blame are also emphasised (Department of Health, 2019). Working with hopelessness may also be relevant to clinicians due to its association with clinical risk (McMillan, Gilbody, Beresford, & Neilly, 2007). Our research suggests that some of the symptoms that traditionally feature strongly in the clinical conceptualisation and treatment protocols improved relatively faster than other symptoms.

However, *sleeping problems, poor appetite or overeating,* and *psychomotor agitation/retardation* appeared to improve to a lesser degree in the present research yet account for ~ 4%, 11% and 8% of impairment respectively and may still be relevant symptoms to patients that may require additional attention in treatment (Fried & Nesse, 2014).

Our findings that depressive symptoms like *low mood/hopelessness* and *guilt/worthlessness* as well as generalised anxiety symptoms such as *uncontrollable worry* and *too much worry* change comparatively the most during early treatment resonate with findings from network analyses. These have previously suggested that symptoms relating to low mood or hopelessness and failure, guilt or worthlessness are amongst the most central symptoms (Beard et al., 2016; O'Driscoll et al., 2021). Symptoms relating to worry have been reported to be the most central anxiety symptoms (Beard et al., 2016). Some arguments suggest that central symptoms are critical elements within networks as they interlink closely with other symptoms (Beard et al., 2016). They are often considered to be highly relevant in the maintenance of a network and are potentially important treatment targets (Beard et al., 2016). However, caution should be taken to avoid overinterpretation, given that centrality does not necessarily equate to clinical importance (Bringmann et al., 2019; O'Driscoll et al., 2021) or clinical utility, such as prognostic capacity (Buckman et al., 2021; O'Driscoll et al., 2021). As such, the literature appears ambiguous regarding the practical meaning and application of centrality metrics. It is nonetheless encouraging that our findings suggest several of the symptoms thought to be central in networks, as well as clinical conceptualisations, change comparatively more during treatment. However, given these noted limitations, more research is needed.

The present research has primarily focused on how the symptoms change across CBT broadly. Ideally, these changes could be mapped onto specific therapeutic techniques. This may begin to explain which elements of CBT are acting on specific symptoms and why some symptoms may increase at a faster pace. Disentangling the complex nature of psychotherapies and how these relate to specific symptoms would have important implications for precision psychiatry – a more detailed understanding of which therapeutic techniques have an effect on specific symptoms could potentially lead to better treatment matching on both empirical and theoretical grounds. However, having no detailed information regarding the order/nature of specific techniques, and the lack of fidelity

measures in IAPT, make this difficult to assess (Martin, Iqbal, Airey, & Marks, 2022). Component studies of CBT, which also look at symptoms rather than sum scores alone, may be beneficial in gaining more granular insights into CBT; however, to date, efforts to understand specific therapeutic techniques appear underpowered (Cuijpers, Cristea, Karyotaki, Reijnders, & Hollon, 2019). Including more detailed recording of therapeutic interventions as part of routine care may be a feasible way to generate large and rich datasets to begin to address some of these questions with sufficient statistical power.

6.6.1 Strengths & Limitations

The present study was based on data from a large retrospective cohort of patients receiving CBT that was obtained from clinical practice to examine individual symptoms. The naturalistic setting adds to the ecological validity of the findings. We also used an interpretable model to examine how individual symptoms changed relative to one another and used a treatment duration that is reasonably representative of the average adult being treated for depression and anxiety in primary care in England.

Despite these strengths, several limitations should be considered. The reliability and validity of using item-level questionnaire responses, rather than sum scores, is still under debate (Boschloo et al., 2019b). While examining how symptoms improve relative to one another is a strength, the lack of a control group limits causal conclusions regarding the symptom-specific effects of CBT. For example, the differential trajectories of specific symptoms might be explained by items with higher baseline scores, such as *guilt/worthlessness* and *uncontrollable worry*, having more scope to regress to the mean. However, not all symptoms that improved comparatively less had lower baseline scores, i.e., *sleeping problems* improved comparably less than other symptoms but had a similar baseline severity to symptoms that improved most. A further explanation may be that symptoms that improved the most may be more amendable to natural change/recovery. However, this seems unlikely given that symptoms that improved the most included low mood and worry, which are considered core symptoms of depression and anxiety (American Psychiatric Association, 2013). We also model the probability of being symptom-free. It may nonetheless be clinically meaningful if patients move from a score of 3 (experiencing symptoms “nearly every day”) to a score of 1 (experiencing symptoms “several days”). Due to the modelling choice, we do not capture this nuance.

While efforts were made to capture a sample of patients who received high-intensity CBT, there are limitations of examining treatment in observational data. While clinicians are trained in accordance with clinical protocols, the lack of treatment fidelity measures in IAPT (Martin et al., 2022) raises possible concerns about the consistency and fidelity to protocol of treatments that are delivered in routine care. A lack of fidelity is likely to dilute any observed effects. Relatedly, clinicians may already be tailoring interventions to individual clients with clinical work placing importance on prioritising problem areas that are specific to individuals (Department of Health, 2019). We also cannot exclude the possibility of any data entry errors or variations between clinicians and services on how data is recorded, related to treatment, but also for other recorded characteristics.

A further limitation of the present analysis is the missing outcome data. While recording of the sum scores is high in IAPT, the sets of individual item scores had up to 39% missing data for a given appointment (supplementary table D2). We conducted a complete-case sensitivity analysis with patients who had all individual item scores recorded at every appointment and found effects of a similar magnitude and direction (supplementary tables D5 and D6). This suggests that missingness is not strongly associated with the value of the outcome scores and a missing at random assumption may be valid. However, our analysis was based on a sample of patients who had at least one complete set of item-level questionnaire scores recorded at any appointment and therefore excludes patients who had no individual items recorded at any appointment. We compared the sociodemographic and clinical characteristics of patients who had at least one set of individual-level outcomes scores recorded and those who had none recorded and found no obvious differences. Possible reasons for not having individual-level item scores recorded that are less likely to be related to the values of the missing data themselves include some services having an online portal for filling out questionnaires which inherently collect individual scores while others do not, or some clinicians entering individual item scores whereas others enter only the sum scores. A more problematic situation occurs if the missing data is related to the values themselves; for example, if patients who score highly on a particular item are more likely to have missing data. However, this appears less probable as most patients nonetheless had sum scores recorded which implies that all questions were answered. However, whether the observed item-level scores are representative of the missing data remains speculative and cannot be determined as the missing data is, by definition, not present and this limitation must be considered in the interpretation of the results.

The main analysis was based on eight appointments; this has multiple implications. A proportion of patients will require less than eight appointments to improve whilst others drop out of treatment. This likely affects the probability of overall response but is less likely to influence relative symptom trajectories unless specific symptoms are predictors of retention in treatment or dropout. Secondly, other symptoms might improve at a later stage of treatment either by being targeted by specific CBT techniques, which may be scheduled later in treatment or as a consequence of cascading effects of improvements in other symptoms. Given our choice of eight appointments, which is the average number of appointments delivered in IAPT, and potentially covers some of the later responses which are observed for some patients after approximately six appointments (Saunders et al., 2019) we can be somewhat confident that these results may generalise to the average patient. However, we cannot exclude the possibility that the rate of change for symptoms may differ over the course of longer treatment. As predictions for appointments beyond eight are extrapolations based on the first eight appointments, they remain speculative and are primarily for illustrative purposes.

6.6.2 Implications

This study suggests that sum scores may obscure important differences of the effects of interventions on specific symptoms. However, we only captured symptoms recorded on the PHQ-9 and GAD-7, which are less extensive than other questionnaires. Given their suggested relevance, questionnaires that include non-DSM symptoms may be insightful (Fried, Epskamp, Nesse, Tuerlink, & Borsboom, 2016). Similarly, questionnaires that capture symptoms more granularly, such as disaggregating symptoms like psychomotor agitation and retardation, may be beneficial given their differential contribution to impairment (Fried & Nesse, 2014). A particular focus on anxiety may also be beneficial given that there is comparably less focus on individual symptoms.

While the findings of the present research require replication and more rigorous evaluation, they do provide a promising avenue for understanding the heterogeneity in clinical outcomes as well as precision psychiatry/psychotherapy. It could potentially be possible to match patients to treatment based on their baseline presentation of symptoms; particularly if the symptom-specific effects of CBT and other interventions are better understood and evaluated in the future. Furthermore, it provides the opportunity for more targeted combinations of interventions that may work on symptoms which respond comparatively less. That is, in the

present research, sleeping problems appeared to improve comparatively less. As such, augmenting CBT with an intervention that targets sleep problems, such as an internet-delivered insomnia programme, amongst patients with higher levels of insomnia may provide a relatively cost-effective way to improve treatment outcomes (Ye et al., 2016; Darden et al., 2021).

However, much of the evidence that will likely lead to making such improvements in personalised psychotherapy more tangible is yet to be robustly established and replicated. In addition to a lack of consensus on how psychological therapy works, the pathophysiology and etiological mechanisms of depression and anxiety are also poorly understood (Nemeroff, 2020). Further research that addresses both areas will undoubtedly contribute to improving clinical outcomes if therapeutic interventions can then be matched to clinical presentations based on their mechanisms.

6.6.3 Conclusion

Trajectories of symptoms appear to differ during treatment with CBT, highlighting the importance of examining individual symptoms rather than sum scores alone. It appears that CBT improves at least some of the core symptoms of depression and anxiety, or symptoms that could be considered very clinically/theoretically important, fastest. While the research is observational and warrants further evaluation, there are possible implications for clinical practice and potential avenues for future research. Examining symptoms may provide an avenue for precision psychiatry/psychotherapy, allowing for more targeted prescription of treatments and/or more evidence-based augmentation of treatments.

6.7 Tables and figures

Table 1. *Baseline clinical and sociodemographic patient characteristics*

	Mean (SD)/Count (%)
n	5306
Gender	
Female	3675 (69.3)
Male	1631 (30.7)
Age	37.56 (13.41)
Ethnicity	
White	4648 (87.6)
Black, Asian, and ethnic minorities	658 (12.4)
Employment Status	
Employed	3256 (61.4)
Not Working	2050 (38.6)
Index of Multiple Deprivation	20.46 (12.08)
Disability Status	
None	4664 (87.9)
Yes	642 (12.1)
Long-Term Health Condition Status	
None	3651 (68.8)
Yes	1655 (31.2)
Diagnosis	
Depressive episode	2437 (45.9)
Recurrent depressive disorder	977 (18.4)
Generalized anxiety disorder	1892 (35.7)
Medication Status	
Yes	2909 (54.8)
No	2397 (45.2)
Referral Number	1.97 (1.38)
Referral Source	
Self	3437 (64.8)

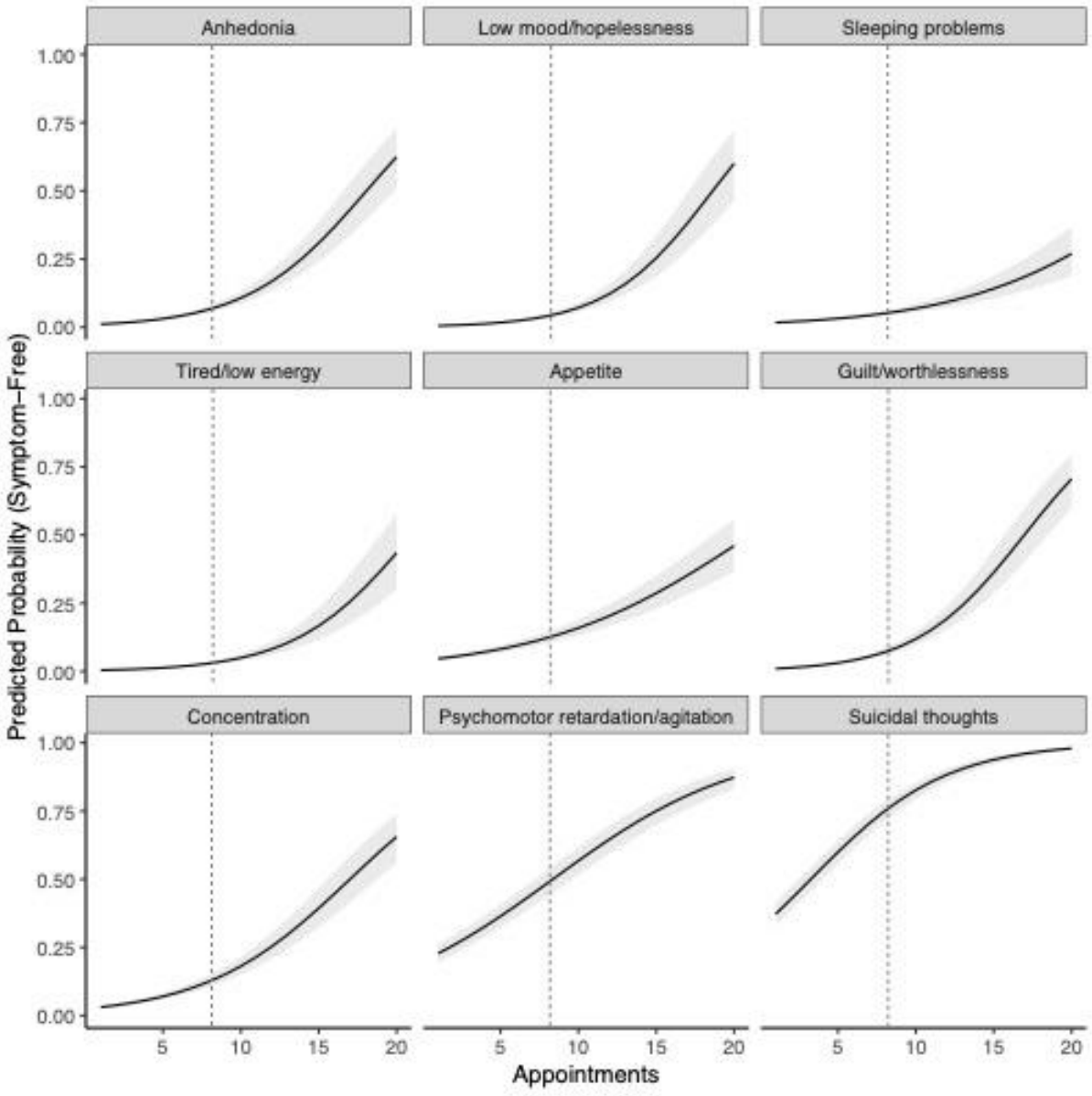
Primary Care	1572 (29.6)
Other	297 (5.6)
Patient Health Questionnaire-9	16.61 (5.61)
Generalised Anxiety Disorder Scale-7	14.70 (4.47)
Work and Social Adjustment Scale	22.19 (8.82)

Table 2. *Effects of appointment on the odds of being symptom free for item-level scores stratified by questionnaire*

	Odds Ratio	**95 % Confidence Intervals**		**p-value**
Patient Health Questionnaire-9				
Question 1 x Appointment	1.04	1.01	1.07	0.014
Question 2 x Appointment	1.07	1.04	1.11	< 0.001
Question 3 x Appointment	0.94	0.91	0.96	< 0.001
Question 4 x Appointment	1.04	1.00	1.08	0.027
Question 5 x Appointment	0.93	0.91	0.95	< 0.001
Question 6 x Appointment	1.06	1.03	1.09	< 0.001
Question 7 x Appointment	0.99	0.97	1.01	0.335
Question 8 x Appointment	0.94	0.92	0.96	< 0.001
Question 9 x Appointment	1.00	0.98	1.03	-
Generalised Anxiety Disorder-7				
Question 1 x Appointment	1.07	1.00	1.13	0.044
Question 2 x Appointment	1.09	1.04	1.14	< 0.001
Question 3 x Appointment	1.09	1.04	1.15	< 0.001
Question 4 x Appointment	1.03	0.99	1.07	0.169
Question 5 x Appointment	0.89	0.87	0.92	< 0.001
Question 6 x Appointment	0.93	0.90	0.96	< 0.001
Question 7 x Appointment	0.93	0.89	0.97	-

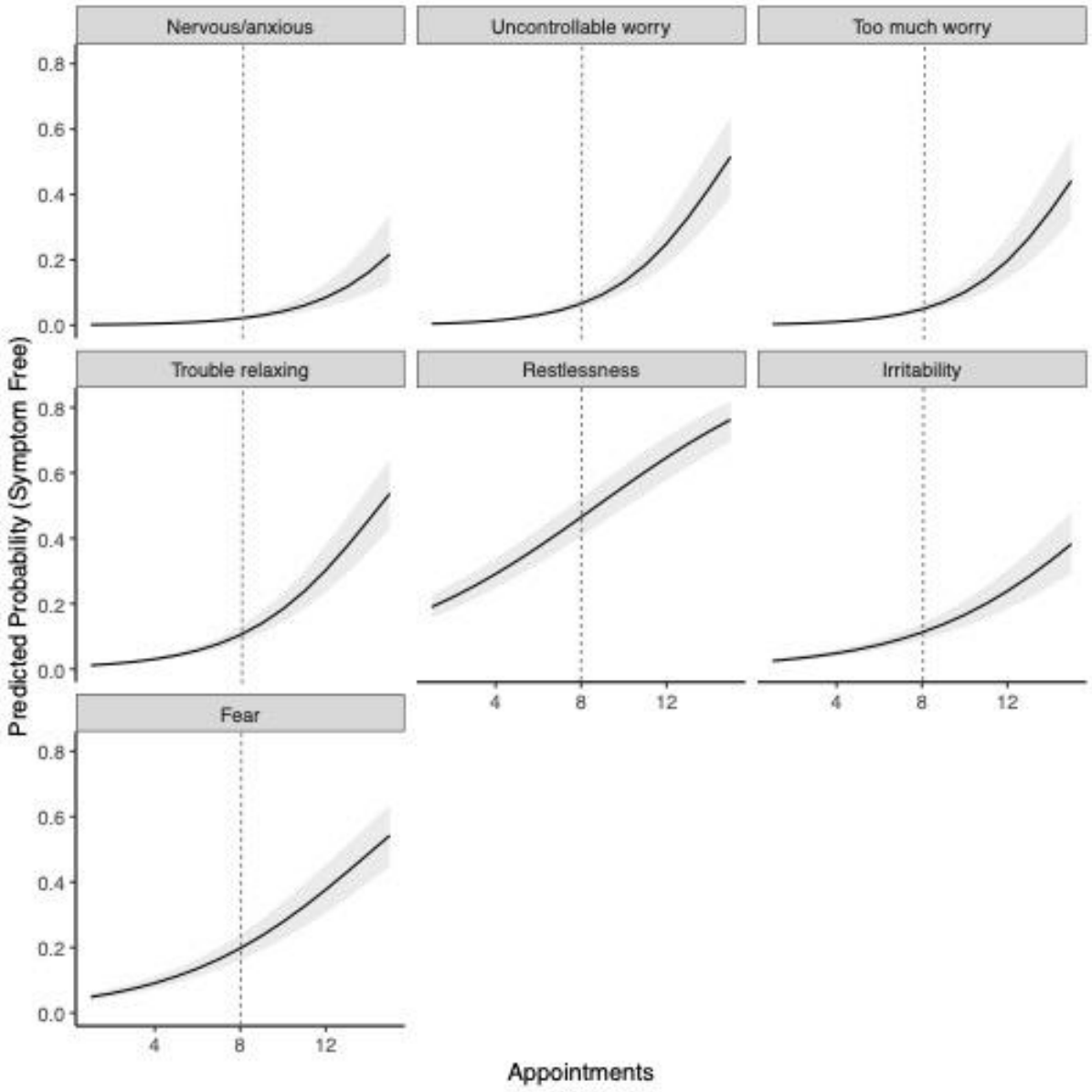
** Adjusted for main effects: age, gender, ethnicity, employment status, Index of Multiple Deprivation, disability status, long-term health conditions, diagnosis (for depression only), baseline Patient Health Questionnaire-9, baseline Generalised Anxiety Disorder Scale -7, baseline Work and Social Adjustment Scale, medication status, referral number, referral source, service, year, question, appointment as well as random effects of referral id and appointment id. P-values for the reference group of the sum coding cannot be estimated.*

Figure 4. *Predicted probability of being symptom free on individual Patient Health Questionnaire-9 items across Cognitive Behavioural Therapy appointments*



**Dotted line denotes predictions based on observed data (8 appointments) with predictions beyond being extrapolations*

Figure 5. *Predicted probability of being symptom free on individual Generalised Anxiety Disorder-7 items across Cognitive Behavioural Therapy appointments*



**Dotted line denotes predictions based on observed data (8 appointments) with predictions beyond being extrapolations*

6.8 References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425787

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7*. http://CRAN.R-project.org/package=lme4

Beard, C., Millner, A. J., Forgeard, M. J., Fried, E. I., Hsu, K. J., Treadway, M. T., ... & Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, *46*(16), 3359-3369. https://doi.org/0.1017/S0033291716002300

Bekhuis, E., Schoevers, R., de Boer, M., Peen, J., Dekker, J., Van, H., & Boschloo, L. (2018). Symptom-specific effects of psychotherapy versus combined therapy in the treatment of mild to moderate depression: A network approach. *Psychotherapy and Psychosomatics*, *87*(2), 121-123. https://doi.org/10.1159/000486793

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., ... & Cuijpers, P. (2019a). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: results from an individual patient data meta-analysis. *World Psychiatry*, *18*(2), 183-191. https://doi.org/10.1002/wps.20630

Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., & Klein, J. P. (2019b). Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: The potential of network estimation techniques. *Behaviour Research and Therapy*, *122*, 103440. https://doi.org/10.1016/j.brat.2019.103440

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892–903. https://doi.org/10.1037/abn0000446

Buckman, J., Cohen, Z., O'Driscoll, C., Fried, E., Saunders, R., Ambler, G., . . . Pilling, S. (2021). Predicting prognosis for adults with depression using individual symptom data: A comparison of modelling approaches. *Psychological Medicine*, 1-11. https://doi.org/10.1017/S0033291721001616

Carey, T. A., Griffiths, R., Dixon, J. E., & Hines, S. (2020). Identifying Functional Mechanisms in Psychotherapy: A Scoping Systematic Review. *Frontiers in Psychiatry, 11*, 291. https://doi.org/10.3389/fpsyt.2020.00291

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry, 23*(4), 318-327. https://doi.org/10.3109/09540261.2011.606803

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

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., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research, 29*(1), 15-29. https://doi.org/10.1080/10503307.2017.1395922

Cuijpers, P., van Straten A., Bohlmeijer E., Hollon SD., Andersson G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine, 40*(2), 211-223. https://doi.org/10.1002/wps.20701

Darden, M., Espie, C. A., Carl, J. R., Henry, A. L., Kanady, J. C., Krystal, A. D., & Miller, C. B. (2021). Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. *Sleep, 44*(4), zsaa223. https://doi.org/10.1093/sleep/zsaa223

Delgadillo, J., & Gonzalez Salas Duhne, P. (2020). Targeted prescription of cognitive–behavioral therapy versus person-centered counseling for depression using a machine learning approach. *Journal of Consulting and Clinical Psychology, 88*(1), 14-24. https://doi.org/10.1037/ccp0000476

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79*, 15-22. https://doi.org/10.1016/j.brat.2016.02.003

Department of Health. (2011). *National curriculum for cognitive behavioural therapy courses* (2nd ed.). London, UK: Kings College London. https://www.uea.ac.uk/documents/246046/11919343/national-curriculum-for-high-intensity-cognitive-behavioural-therapy-courses.pdf/e893a788-8df2-4f2b-a1bd-a1c9c0ae37a1

Eronen, M. I. (2020). Causal discovery and the problem of psychological interventions. *New Ideas in Psychology, 59*, 100785. https://doi.org/10.1016/j.newideapsych.2020.100785

Fenn, K., & Byrne, M. (2013). The key principles of cognitive behavioural therapy. *InnovAiT*, *6*(9), 579-585. https://doi.org/10.1177/1755738012471029

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders, 189*, 314-320. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PloS One, 9*(2), e90311. https://doi.org/10.1371/journal.pone.0090311

Fried, E. I., & Nesse, R. M. (2015a). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Medicine, 13*(1), 1-11. https://doi.org/10.1186/s12916-015-0325-4

Fried, E. I., & Nesse, R. M. (2015b). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders, 172*, 96-102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2014). Depression is more than the sum-score of its parts: individual DSM symptoms have different risk factors. *Psychological Medicine, 44*(10), 2067-2076. https://doi.org/10.1017/S0033291713002900

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry, 9*(6), 137-150. https://doi.org/10.1016/S2215-0366(21)00395-3

Green, S. A., Honeybourne, E., Chalkley, S. R., Poots, A. J., Woodcock, T., Price, G., ... & Green, J. (2015). A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health

programme. *BMJ Open*, *5*(5), e006103. http://dx.doi.org/10.1136/bmjopen-2014-006103

Huibers, M. J., Lorenzo-Luaces, L., Cuijpers, P., & Kazantzis, N. (2021). On the road to personalized psychotherapy: A research agenda based on cognitive behavior therapy for depression. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.607508

Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of Affective Disorders, 80* (2-3), 125-133. https://doi.org/10.1016/S0165-0327(03)00108-3

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

Martin, C., Iqbal, Z., Airey, N. D., & Marks, L. (2022). Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: How can it better meet the range of primary care mental health needs?. *The British Journal of Clinical Psychology*, *61*(1), 157–174. https://doi.org/10.1111/bjc.12314

McLennan D, Noble S, Noble M, Plunkett E, Wright G, Gutacker N. *The English Indices of Deprivation 2019: technical report.* Ministry of Housing, Communities and Local Government, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833951/IoD2019_Technical_Report.pdf

McMillan, D., Gilbody, S., Beresford, E., & Neilly, L. (2007). Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. *Psychological Medicine*, *37*(6), 769-778. https://doi.org/10.1017/S0033291706009664

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 Institute for Clinical Excellence (2009). *Depression in adults: recognition and management – Clinical guidelines [CG90].* https://www.nice.org.uk/guidance/cg90

National Institute for Clinical Excellence (2011). *Generalised anxiety disorder and panic disorder in adults: management– Clinical guidelines [CG113]. https://www.nice.org.uk/guidance/cg113*

Nemeroff, C. B. (2020). The state of our understanding of the pathophysiology and optimal treatment of depression: glass half full or half empty?. *American Journal of Psychiatry*, *177*(8), 671-685. https://doi.org/10.1176/appi.ajp.2020.20060845

NHS Digital. (2021, November 25). *Psychological Therapies, Annual report on the use of IAPT services, 2020-21*. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2020-21#

O'Driscoll, C., Buckman, J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., DeRubeis, R. J., Gilbody, S., Hollon, S. D., Kendrick, T., Kessler, D., Lewis, G., Watkins, E., Wiles, N., & Pilling, S. (2021). The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: analysis of data from six randomised control trials. *BMC Medicine*, *19*(1), 109. https://doi.org/10.1186/s12916-021-01971-0

Panos, A., & Mavridis, D. (2020). TableOne: an online web application and R package for summarising and visualising data. *Evidence-Based Mental Health*, *23*(3), 127-130. http://dx.doi.org/10.1136/ebmental-2020-300162

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109-188. https://doi.org/10.1037/h0024436

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169-188. https://doi.org/10.1023/A:1020363010465

Saunders R, Buckman JEJ, Cape J, Fearon P, Leibowitz J, Pilling S (2019) Trajectories of depression and anxiety symptom change during psychological therapy. *Journal of Affective Disorders, 249*, 327-335. https://doi.org/10.1016/j.jad.2019.02.043

Saunders R, Cape J, Leibowitz J, Aguirre E, Jena R, Cirkovic M, Wheatley J, Main N, Pilling S, Buckman JEJ (2020) Improvement in IAPT outcomes over time: are they driven by changes in clinical practice? *The Cognitive Behaviour Therapist, 13*, 1-15. https://doi.org/10.1017/S1754470X20000173

Saunders, R., Buckman, J. E., & Pilling, S. (2020). Latent variable mixture modelling and individual treatment prediction. *Behaviour Research and Therapy*, *124*, 103505. https://doi.org/10.1016/j.brat.2019.103505

Spitzer, R. L., Kroenke, K., & Williams, J. B. (2006). Generalized anxiety disorder 7-item (GAD-7) scale. *Archives of Internal Medicine*, *166*, 1092-7. https://doi.org/10.1001/archinte.166.10.1092

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

University College London. (n.d.). *Competency Frameworks - Cognitive and Behavioural Therapy*. https://www.ucl.ac.uk/pals/sites/pals/files/all_problem-specific_competences.pdf

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., ... & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open, 3*(8), e002847. http://dx.doi.org/10.1136/bmjopen-2013-002847

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180-185. https://doi.org/10.1002/wics.147

Wichniak, A., Wierzbicka, A., Walęcka, M., & Jernajczyk, W. (2017). Effects of antidepressants on sleep. *Current Psychiatry Reports*, *19*(9), 1-7. https://doi.org/10.1007/s11920-017-0816-4

Ye, Y. Y., Chen, N. K., Chen, J., Liu, J., Lin, L., Liu, Y. Z., ... & Jiang, X. J. (2016). Internet-based cognitive–behavioural therapy for insomnia (ICBT-i): a meta-analysis of randomised controlled trials. *BMJ Open*, *6*(11), e010707. http://dx.doi.org/10.1136/bmjopen-2015-010707

6.1 Declaration of Interests

None.

6.2 CReDiT Author Statement

Clarissa Bauer-Staeb: conceptualisation, methodology, formal analysis, writing – original draft. **Emma Griffith**: writing – review & editing, supervision. **Julian J. Faraway:** conceptualisation, methodology, formal analysis, writing – review & editing, supervision. **Katherine S. Button**: conceptualisation, methodology, writing – review & editing, supervision.

6.3 Acknowledgements

The data extraction and anonymisation was kindly performed by Mayden, the developers of iaptus – a patient management software used within IAPT.

6.4 Funding

This study was funded by the University of Bath PhD Research Programme Studentship awarded to CBS.

7 The Early Impact of COVID-19 on Primary Care Psychological Therapy Services: A Descriptive Time Series of Electronic Healthcare Records

7.1 Statement of Authorship

This declaration concerns the article entitled:					
The Early Impact of COVID-19 on Primary Care Psychological Therapy Services: A Descriptive Time Series of Electronic Healthcare Records					
Publication status					
Draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	X
Publication details	Bauer-Staeb, C., Davis, A., Smith, T., Wilsher, W., Betts, D., Eldridge, C., ... & Button, K. S. (2021). The early impact of COVID-19 on primary care psychological therapy services: A descriptive time series of electronic healthcare records. *EClinicalMedicine*, *37*, 100939. https://doi.org/10.1016/j.eclinm.2021.100939				
Copyright status					
The material has been published with a CC-BY license	X	The publisher has granted permission to replicate the material included here			
Candidate's contribution to the paper	The candidate predominantly executed the formulation of ideas (75%) and the design of methodology (80%) as well as conducting the formal analysis (90%). The candidate predominantly executed the presentation of data in journal format (85%).				
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.				
Signed	[redacted]		**Date**	22/08/2022	

The Early Impact of COVID-19 on Primary Care Psychological Therapy Services: A Descriptive Time Series of Electronic Healthcare Records

Clarissa Bauer-Staeb,[a] Alice Davis,[b,c] Theresa Smith,[b] Wendy Wilsher,[c] David Betts,[c] Chris Eldridge,[c] Emma Griffith,[a,d] Julian Faraway,[2] & Katherine S. Button.[a]

[a] Department of Psychology, University of Bath, Bath, United Kingdom

[b] Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

[c] Mayden, Bath, United Kingdom

[d] Avon and Wiltshire Mental Health Partnership NHS Trust, Bath, United Kingdom

7.2 Abstract

Background. There are growing concerns about the impact of the COVID-19 pandemic on mental health. With government-imposed restrictions as well as a general burden on healthcare systems, the pandemic has the potential to disrupt the access to, and delivery of, mental healthcare.

Methods. Electronic healthcare records from primary care psychological therapy services (Improving Access to Psychological Therapy) in England were used to examine changes in access to mental health services and service delivery during the early stages of the COVID-19 pandemic. A descriptive time series was conducted using data from five NHS trusts to examine patterns in referrals to services (1st January 2019 to 24th May 2020) and appointments (1st January 2020 to 24th May 2020) taking place.

Findings. The number of patients accessing mental health services dropped by an average of 55% in the early weeks after the March 2020 lockdown was announced, reaching a maximum reduction of 74% in the initial three weeks after the lockdown in the UK, which gradually recovered to a 28% reduction by May. We found some evidence suggesting changes in the sociodemographic and clinical characteristics of referrals. Despite a reduction in access, the impact on appointments appeared limited, with service providers shifting to remote delivery of care.

Interpretation. Services appeared to adapt to provide continuity of care in mental health services. However, patients accessing services reduced, potentially placing a future burden on services. Despite the observational nature of the data, the present study can inform the planning of service provision and policy.

Funding. AD and TS were funded by Innovate UK (KTP #11105).

7.3 Research in Context

7.3.1 Evidence before this study

Google Scholar was searched using the terms “COVID-19” and “mental health services”. Due to the novelty of the research, few peer-reviewed articles had been published (pre-prints excluded) at the time of data analysis. Since then, research surveying mental health staff reports rapid innovation in services to adapt to COVID-19, with an emphasis on remote working. Furthermore, research using electronic healthcare records has been conducted that suggests a drop in referrals across primary and secondary mental health services at both regional and national levels, with remote mediums being increasingly used for clinical contacts.

7.3.2 Added value of this study

This is one of the first studies looking at the impact of COVID-19 on access and service delivery in primary care psychological therapy services. Specifically, examining the number of referrals and the sociodemographic and clinical characteristics of these as well as the number of appointments and how appointments were delivered. We observed reductions in referrals during March 2020 lockdown, with some evidence potentially indicating changes in sociodemographic and clinical characteristics of referrals. This may suggest changes in demand amongst different groups. The impact of COVID-19 on the total number of appointments was limited, with a shift to remote care. This possibly suggests that the care of patients who sought treatment, or were already in contact with services, may have only seen small disruptions.

7.3.3 Implications of all the available evidence

Although the observational nature of the data should be noted, the research has the potential to support planning of clinical practice and policy. Despite service providers in the present study appearing to adapt to the pandemic by offering remote care, there was a reduction in access to mental health services compared to what would have been expected at that time of year. This reduction has likely left some people without adequate mental health support, particularly as a switch to remote care may not have occurred rapidly across all of England. Although speculative, this deficit may result in greater pressures to treat a possible backlog as well as dealing with the potential aftermath of the long-term consequences of the pandemic on mental health.

7.4 Introduction

In public healthcare systems such as the National Health Service (NHS) in England, primary care services are often the first port of call for patients with common mental health problems. Patients show a preference for psychological therapy over medication.[1] In England, psychological therapy in primary care settings is predominantly delivered by Improving Access to Psychological Therapy (IAPT) services.[2] IAPT services deliver a range of low- and high-intensity psychological interventions for depression and anxiety.[2] IAPT have implemented routine data collection, measuring sociodemographic and clinical patient characteristics as well as treatment data.[2] These data are nationally reported on a monthly basis.[3] In 2019-20 IAPT received approximately 1·69 million referrals.[3]

There is growing concern about the profound and long-lasting impact of COVID-19 on mental health from multiple areas, including academia, healthcare, and lived experience advocates .[4] Research suggests that clinically significant levels of mental distress rose during the pandemic in England – from approximately 19% in 2018-9 to 27% in 2020.[5] People who have previously, or are currently, suffering from mental health conditions as well as those who become mentally unwell during the pandemic, may potentially be vulnerable groups.[4] The pandemic may also disproportionately affect the mental health of other groups, including those with pre-existing mental and physical health conditions, individuals facing financial instability, ethnic minority groups, as well as young and older adults.[4–8] The provision of adequate mental health support to address the psychological impact of the pandemic and meet mental health needs is critical.

Despite the growing concerns about the COVID-19 pandemic on mental health, less focus has been placed on how individuals with mental health problems are supported.[9] Concerns have been raised about adequate service provision during the pandemic, with staff shortages and service reconfigurations as well as the pressures of implementing infection control measures posing challenges to mental health staff.[4,9] Research suggests that referrals to primary and secondary mental health services reduced after lockdown regionally and nationally, with an increase in remote mediums to conduct clinical contacts.[10-12] However, less is known about how the impact on psychological care provision varied by patient characteristics.

The use of electronic healthcare records provides a first avenue to examine the impact of COVID-19 on primary care mental health services at scale.[4] Using electronic healthcare records from routinely collected data in IAPT services, we aim to understand service use during the pandemic. Specifically, we investigate access to psychological therapy services generally as well as how this may have varied by patient characteristics early during the pandemic. Furthermore, we aim to understand the impact of COVID-19 on how clinical care was impacted and delivered.

7.5 Methods

7.5.1 Settings & Design

IAPT are primary care services in England delivering psychological interventions for depression and anxiety.[2] A minimum dataset is routinely collected for all patients, recording data relating to patient characteristics, treatment, as well as the routine measurement of clinical outcome questionnaires for depression and anxiety.[2] The present study examines data from five NHS trusts in England, which were chosen for convenience purposes. All incoming referrals between 1st January 2019 until 24th May 2020 were examined as well as all corresponding appointments occurring between 1st January 2020 until 24th May 2020, covering the first national lockdown in England. Appointments were only examined from 2020 onwards due to pragmatic reasons – appointment data from 2019 would have included appointments from referrals that took place prior to 2019. In order to have coverage of all appointments occurring in 2019, these earlier referrals would have had to been selectively added to the data extract which would have been challenging due to service reconfigurations. The lockdown in England was implemented on the 23rd March 2020.[13] All members of the public were asked to stay at home and not leave their house other than to shop for basic necessities, medical reasons, one form of exercise a day, or travelling to and from work in instances where this was absolutely necessary.[13] All shops selling non-essential goods were instructed to close, and social events and gatherings were prohibited.[11]

All data were extracted and fully anonymised by WW from Mayden, the software providers of the patient management system used by a large proportion of IAPT services. The authors CBS, KSB, and JF were provided with the dataset on 28th May 2020 and have access for ten years thereafter.

7.5.2 Ethical Approval

Due to the anonymous nature of the data, the present study was exempt from NHS Ethical Review. The project received ethical approval by the University of Bath (PREC: 19-015). Due to the anonymous nature of the data, it was not possible to retrieve individual patient consent. However, patients who had records indicating they did not want their data to be used for further processing were not included in the dataset provided by Mayden.

7.5.3 Measures

To examine the impact of COVID-19 on access to services, we examined the data of all incoming referrals to IAPT. Specifically, we consider the total number of referrals as a measure of the impact of the pandemic on patients accessing services. We further examined the characteristics of these referrals to explore possible changes in the demography or means of accessing services. Specifically, we examine characteristics including age, gender, ethnicity, the Index of Multiple Deprivation (IMD) as a proxy for socioeconomic status, population density (people per square kilometre) as a proxy for urbanicity, and referral source.[14,15] IMD and population density were determined at the Lower Super Output Area level via linkage to the Office for National Statistics data. Clinical characteristics of referrals were also explored, examining comorbid long-term health condition status, number of previous referrals, baseline depression measured via the Patient Health Questionnaire-9 (PHQ-9), and baseline anxiety measured by the Generalised Anxiety Disorder Scale-7 (GAD-7).[16,17] In order to examine the impact of COVID-19 on service delivery, we examined appointment data, specifically the number of appointments and the consultation medium for attended appointments. Variable definitions can be found in supplementary table E1.

7.5.4 Statistical Analysis

Descriptive time series are presented containing the weekly total count for categorical variables and weekly averages for continuous variables. Variables were examined individually, with the exception of ethnicity and referral source. Ethnicity was stratified by referral source as self-referrals were introduced into IAPT to increase access for minority groups such as Black, Asian and minority ethnic (BAME) groups.[18] To quantify changes in access, weekly counts of incoming referrals for the 9 weeks after lockdown were compared to the corresponding weeks in 2019.

Missing data for factor variables were defined as an additional factor level classed as *unknown*. Missing data for continuous variables were excluded. At the patient-level, the frequency of missing data for IMD and population density was approximately 0.5%. There may be various reasons as to why no baseline clinical measures are taken, including patients having never attended an appointment. For all reporting of the PHQ-9 and GAD-7, the last week of data (18th to 24th May 2020) was excluded as it contained approximately 55% missing data, exceeding the maximum weekly missing data observed throughout the year. This possibly reflects that people referred close to the data extraction date were unlikely to have had a first appointment booked within this short timeframe. After excluding the last week of data, the patient-level data contained approximately 28% missing data for the baseline PHQ-9 and GAD-7.

All analyses were performed in the R programming language.[19] All data analysis and visualisations were performed using base R and 'ggplot2'. [19,20]

7.5.5 Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

7.6 Results

7.6.1 Sample Characteristics

In the timeframe of 1st January 2019 to 24th May 2020, 171,823 referrals came into IAPT services across five different areas (Table 1). The majority of referrals were self-referrals (76%), typically female (66%), White (68%), with an average age of 38 years. However, patient characteristics demonstrate that services in different areas serve heterogeneous populations, with two serving populations from more urban areas, with slightly greater deprivation and a larger proportion of BAME individuals (supplementary table E2). From 1st January 2020 to 24th May 2020 248,628 appointments were scheduled, amongst which the majority were attended (75%) and took place remotely (59%; Table 2).

[Table 1. *Characteristics of referrals from 1st January 2019 to 24th May 2020*]

[Table 2. *Characteristics of appointments from 1st January 2020 to 24th May 2020*]

7.6.2 Total Referrals

There was a decline in referrals in March 2020 (Figure 1). The decline in referrals commenced approximately one week prior to the official government announcement of a lockdown in England, beginning on 23rd March 2020. The decline in referrals relative to those observed at the same time in 2019 was greatest in the immediate three weeks of lockdown, reaching a maximum reduction of 74%. The decline in referrals is of a similar magnitude to a decline observed at the end of December during the Christmas holidays, albeit slightly larger.

[Figure 1. *Total weekly referrals from 1st January 2019 to 24th May 2020*]

Referrals started to gradually increase again over time. However, referrals had not fully recovered by the end of May. The total number of referrals in the week commencing on 18th May were still 28% lower than the corresponding week in 2019; that is, 72% of their previous volume.

In the early weeks after England entered lockdown, there was an average 55% reduction in referrals compared to the corresponding weeks in 2019 (supplementary table E3). In the present dataset, this translated into approximately 12,000 fewer patients accessing mental health services than might be expected for that time of year.

7.6.3 Sociodemographic characteristics of referrals

There are no clear changes in referrals by gender or long-term condition status (supplementary figures E1 and E2).

[Figure 2. *Total weekly referrals by ethnicity and referral source from 1st January 2019 to 24th May 2020*]

There was a reduction in referrals across all referral sources and ethnicities after the lockdown was imposed in March 2020 (see Figure 2). Self-referrals and referrals from other sources returned to baseline fastest, while referrals from primary care were increasing at a slower rate across all ethnicities. Compared to self-referrals from a White background,

BAME self-referrals appeared to increase again at a faster rate after the initial drop observed around lockdown, being slightly higher than the corresponding time point in 2019. There were 382 self-referrals from patients with a BAME background in the week commencing the 18th May 2020 compared to 338 in the corresponding week in 2019. When examining ethnicity subgroups, there was a particular increase in referrals towards the end of May 2020 by patients with a Black ethnic background, reaching the highest number of self-referrals observed across the entire follow-up period. There were 112 Black self-referrals in the week commencing the 18th May 2020 compared to 78 in the corresponding week in 2019.

[Figure 3. *Weekly average age of referrals from 1st January 2019 to 24th May 2020*]

It appears that age shows a time trend, where the average age increases during the summer and decreases again towards the winter (Figure 3). There appeared to be a slight decrease in average age at referral in the 9 weeks after lockdown (36·8 years), when compared to the same time in 2019 (38·3 years). Although appearing to return to average levels, the average age had not returned to expected levels for the given time of year in May 2020.

[Figure 4. *Weekly average Index of Multiple Deprivation of referrals from 1st January 2019 to 24th May 2020*]

There appeared to be a brief spike in IMD early during lockdown (Figure 4). This increase in IMD appeared to return to levels observed throughout the year relatively fast, with the average IMD of referrals in the nine weeks post-lockdown being only marginally higher (21·5) compared to the same time in the previous year (21·0). However, it appears that IMD levels may be rising towards the end of May (Figure 4).

[Figure 5. *Weekly average population density of referrals from 1st January 2019 to 24th May 2020]*

While there does not appear to be an obvious change in population density immediately after the lockdown was imposed (7515 people per Km^2 in 2020 vs. 7408 people per Km^2 for the same time in 2019), it appears that the population density of referrals may be increasing towards the end of May 2020 (Figure 5).

7.6.4 Clinical Characteristics of Referrals

There appears to be a slight increase in average PHQ-9 and GAD-7 immediately after the lockdown was imposed (Figure 6), with respective scores of 14·8 and 13·4 in the weeks following lockdown compared to 14·3 and 12·6 at the same time in 2019. This increase appeared to be sustained throughout lockdown until May 2020.

[Figure 6. *Weekly average baseline depression and anxiety scores for incoming referrals from 1st January 2019 to 17th May 2020*]

[Figure 7. *Weekly average number of previous referrals from 1st January 2019 to 24th May 2020*]

There appeared to be a small increase in the average number of previous referrals in the nine weeks after the lockdown was imposed (Figure 7), with patients having an average of 1·2 previous referrals post-lockdown compared to 1·0 over the same time in the previous year.

7.6.5 Appointments

There appears to be a brief, relatively small dip in attended appointments around lockdown. There appear to be no meaningful change in Did Not Attend (DNA) appointments or attended too late to be seen after lockdown. There was a relatively large reduction in appointments cancelled by patients after lockdown, whereas there was a brief spike in appointments cancelled by providers around lockdown (see Figure 8).

[Figure 8. *Weekly total appointments by attendance from 1st January 2020 to 24th May 2020*]

Out of all attended appointments, face-to-face consultations reduced after the lockdown was imposed, with remote consultations increasing. Prior to lockdown, 64,201 (60·0%) of appointments took place *face-to-face* and 39,985 (37·4 %) were recorded as *remote*. After lockdown, the majority of appointments took place *remotely* (69,782, 89·2%) with 4,576 (5·8%) being recorded as *face-to-face* appointments. There was also an increase in appointments labelled *other* as well as a short spike in consultation mediums labels being *unknown*. Prior to lockdown, 1,845 (1·7%) appointments had a consultation medium labelled

as *other* and 882 (0·8%) were *unknown*. After lockdown, appointments with the label *other* rose to 2,650 (3·4%) and 1,229 (1·6%) had an unknown consultation medium.

7.7 Discussion

Using electronic healthcare records, we examined the short-term impact of COVID-19 on primary care psychological therapy services in England, with regards to access and service delivery. There was a clear drop in referrals to IAPT around the implementation of lockdown, resulting in approximately 55% fewer patients accessing services in the early weeks after the lockdown was imposed in the UK. There appeared to be a trend indicating faster increases in the number of referrals from BAME, especially Black patients, relative to White patients once referrals started to increase again. While the changes were relatively small, there was some evidence to suggest an increase in referrals from younger patients, patients living in higher deprivation and urban areas, those with higher baseline depression and anxiety scores as well as those who have previously sought treatment. Despite reductions in the number of people accessing services, it appears that the care of patients receiving treatment showed somewhat short-lived disruptions, with services quickly moving to provide remote consultations.

Overall, there was an average reduction of 55% in referrals in the early weeks after lockdown compared to the same timeframe in 2019. This decline in referrals began approximately one week before the lockdown and reached the maximum level within three weeks after the lockdown was announced, with a 74% reduction of referrals in 2020 compared to the same time in 2019. This appears somewhat consistent with regional reports as well as the national trend observed through the monitoring of activity in the patient management software used by a majority of IAPT services, where referrals dropped by approximately 70% early after the lockdown was announced.[10-12,21] The use of primary care psychological therapy services is similar to those observed in other health services, such as a reduced number of patients accessing Accident and Emergency Departments and General Practice.[22,23] Although not returning to baseline, the number of referrals appeared to gradually increase again over time, with referrals being at 72% of their usual volumes towards the end of the follow-up period. Albeit slightly lower, a similar pattern is observed at a national level, where referrals in July 2020 were at 60% of the volumes observed prior to COVID-19.[21] Over a more extended time period time, the referrals appeared to increase again to a greater degree with national data

showing that referrals to adult mental health services, including IAPT, were approximately 10% lower when examined over a longer period from April 2020 to August 2020 compared to the same time in 2019.

Despite referral rates increasing again as the lockdown progressed, a deficit in referred patients was observed. If the present research is extrapolated across England, with an assumed 1·69 million referrals per year, as observed in 2019-20, approximately 160,900 patients who may have normally been referred did not access mental health services in the first weeks after the lockdown was imposed.[3] However, this deficit estimate is likely conservative – figures may be higher in the longer term as referrals had not returned to baseline towards the end of May and the proposed figure does not account for a possible increase in mental health needs or the annual increase in referrals – from 2018/19 to 2019/2020 the annual increase in referrals was 5.7%.[3,5]

Self-referrals appeared to have recovered from the effects of lockdown most rapidly, which may have been facilitated through an increase in the availability and use of online referral systems.[21] Nationally, online self-referrals to IAPT were 13% higher in July 2020 than pre-lockdown.[21] This recovery of self-referral rates was most pronounced in BAME groups, which appeared to return to baseline most rapidly after the initially observed decline. It is tentatively suggested that this could reflect a greater impact of the COVID-19 pandemic on the mental health of the BAME community, which would be consistent with emerging findings highlighting inequalities of the pandemic and a higher percentage of people from a BAME background reporting worse than usual mental health – approximately 50% compared to 35% across all adults.[24,25] However, it also indicates a greater opportunity for BAME groups to access mental health services and may speak to the efficacy of opening up services through self-referrals as a means of doing so.[18] Population-based surveys suggest a higher number of young adults reporting worse mental health during the pandemic than before, as well as higher levels of depression and anxiety during lockdown amongst people from low-income households, which may be related to unstable housing and job or life transitions, amongst others.[25] These may have been exacerbated as a result of lockdown. Similarly, in the present study, we observed an increase in average IMD and a decrease in age at referral, which may indicate an increased demand for psychological therapy by these groups. Population-based surveys have further identified higher levels of depression and anxiety during lockdown amongst people living in urban areas.[24] We found some evidence to suggest

that the population density of referrals was increasing over time after the lockdown. This could potentially be a result of a higher impact of the pandemic in these areas, with more urban areas showing a higher age-standardised mortality rate of COVID-19 between 1st March and 31st July 2020 compared to more rural settings.[26]

There was a slight change in the clinical severity of referrals, with an increase in average depression and anxiety scores of incoming referrals after the lockdown had been implemented in the UK. This increase remained somewhat stable throughout the lockdown. Due to the observational nature of the data, it is difficult to discern whether slight increases in depression and anxiety resulted from a rise in symptoms amongst the general population or whether patients with more severe symptoms were accessing services to a greater degree after the lockdown was imposed. However, population-based surveys showed a decrease in both depression and anxiety from the start of lockdown until May.[24] As such, the latter is suggested as more probable. Furthermore, there was a small increase in the number of previous referrals after lockdown, suggesting that patients who had already accessed services previously were returning to a greater degree. There is some evidence to suggest that patients who re-refer present with more complex psychological problems and may be more likely to have higher baseline depression and anxiety.[27]

While there was an evident impact of the lockdown on people accessing primary care mental health services, it appears that service delivery for patients already being treated pre-lockdown or starting treatment during lockdown may have only seen a limited disruption as services appeared to adapt quickly to new practices. This is consistent with national trends showing no dramatic drop in clinician activity recorded by the IAPT patient management software.[21] Services appeared to rapidly adapt, implementing infection control measures by switching to remote consultations almost exclusively. Mental health staff accounts mirror this, reporting rapid innovation with a particular emphasis on remote working.[9] A small proportion of appointments were still recorded as *face-to-face* after lockdown. There may have been a clinical necessity to continue to see patients face-to-face. However, it is also likely a result of data error – service providers may not have had appropriate labels early during lockdown as patient management systems were being updated to reflect rapid transitions to new ways of working. An example of this might be the use of remote meeting software. This is consistent with an increased practice of labelling appointments with consultations medium classed as *other* and *unknown*. However, it should be noted that there

is a sparsity of patient reports regarding their experiences of accessing and using mental health services during the pandemic.

To our knowledge, the present study is one of the first to examine a quantifiable impact of COVID-19 on primary care mental health services at scale, with data from a wide and diverse range of service providers across multiple geographic regions in England. Nonetheless, the dataset contains only a small subset of service providers in England and may therefore not be nationally representative. There may be variation by service providers and regions that is not captured by the data used in the present study. However, the observed trends appear consistent with those detected by the monitoring of activity within IAPT service's patient management software.[21] The nationally reported IAPT data will provide further insights in the long-term and has the benefit of a larger sample that is representative of all service providers; however, the present analysis allows for more detailed insights. It should be noted that the present research uses observational data, taking a descriptive approach. As such, it is not possible to draw causal conclusions, and the capacity to estimate future impact is limited and remains speculative. We also made no adjustments to accommodate the annual increase in referrals. While this has the benefit of taking into account seasonal variations, which appear to exceed annual increases in magnitude, it potentially leads to deficit figures being underestimated. The figure may further be underestimated as mental health needs may have increased during the pandemic.[5] We also only examine appointment data from referrals occurring in 2019/2020. Furthermore, we cannot quantify the influences of broader organisational and societal events that may have influenced mental healthcare. For example, the NHS ran an 'Open for Business' campaign promoting the public to access healthcare, which may have increased confidence in people seeking treatment.[28] Similarly, organisational influences could include staffing levels and/or service restructuring that influence the capacity to deliver healthcare. Despite occurring after the timeframe of the present study, and thus having little influence on the present findings, protests in support of the 'Black Lives Matter' movement occurred during Spring/Summer of 2020 concomitant to the pandemic. This will be an important factor to consider in future research as it may disproportionately affect minority groups.

A reduction of referrals took place during the early stages of COVID-19, producing approximately a 55% deficit in patients receiving mental healthcare. A concern may be that a backlog of patients has accumulated, which may cause future pressures on service providers

to treat these patients, in addition to a possible excess of patients who may seek mental health support as the long-term consequences of COVID-19 become more apparent. Given the faster increase in self-referrals after the initial drop from BAME groups compared to White groups, as well as a potential trend showing increases in deprivation and urbanity of referrals, service providers catering to these populations may experience a particular surge in demand. Services may also see changes in the demography of referrals, such as a higher proportion of younger patients, patients who have previously sought treatment, or patients with a higher clinical severity. This may serve as a reminder for the need of cultural competency in psychological therapy to meet the needs of all patients accessing services.[29] Periodical horizon scanning of the demography of patients accessing services may provide an avenue to assure that developments in cultural competency adequately reflect demographic changes.

Despite access to mental health services being impacted by COVID-19, the data suggests that service providers in the present study were able to adapt to the pandemic with the adoption of remote consultations. This shift likely provided essential continuity of care to patients receiving psychological interventions. Previous research suggests that remote cognitive behavioural therapy is effective and may increase treatment adherence.[30-32] However, approximately 40% of community and psychological therapy staff have reported difficulties such as having to learn about new technologies too quickly or without enough training, as well as experiencing technical difficulties with remote consultation.[9] Furthermore, remote therapy may come at the cost of poorer maintenance after treatment.[32] Early evidence from the nationally reported data shows a brief dip in clinical outcomes in March/April 2020 but an increase in clinical outcomes in June/July 2020; reaching higher levels than prior to COVID-19.[33] The long-term effects on clinical outcomes remain to be determined, with a pressing need for future research.

The present findings provide insight into the short-term impact of the COVID-19 pandemic on psychological therapy services. Due to the observational nature of the data, results should be interpreted with caution. However, they have the potential to support the planning of clinical practice and public health policy, particularly as restrictions are likely to remain in place until national vaccination programmes have been completed and managing the aftermath of a pandemic. The current findings also highlight a need for further research, presenting avenues for future directions. The long-term impact of COVID-19 on mental

health services and mental health more generally remains to be determined as the delayed consequences, such as economic hardship, become more apparent.

7.8 Tables and Figures

Table 1. *Characteristics of referrals from 1st January 2019 to 24th May 2020*

n	171,823
Age	37·94 (15·12)
Gender -n (%)	
Female	113,006 (65·8)
Male	58,643 (34·1)
Unknown	174 (0·1)
Ethnicity -n (%)	
White	116,964 (68·1)
Black, Asian, and ethnic minority	32,600 (19·0)
Asian	16,976 (9·9)
Black	6,857 (4·0)
Mixed	4,913 (2·9)
Other	3,854 (2·2)
Unknown	22,259 (13·0)
Index of Multiple Deprivation	21·02 (11·81)
People per Square Kilometre	7,311·40 (7306·41)
Long-Term Condition Status -n (%)	
Long-Term Condition	49,562 (28·8)
No Long-Term Condition	97,985 (57·0)
Unknown	24,276 (14·1)
Number of previous referrals	1·01 (1·68)
Baseline PHQ-9*	14·39 (6·43)
Baseline GAD-7*	12·79 (5·42)
Referral Source -n (%)	
Self	130,089 (75·7)
Primary Care	32,672 (19·0)
Other	9,062 (5·3)

**Data is presented as mean (standard deviation) unless otherwise specified. PHQ-9: Patient Health Questionnaire -9; GAD-7: Generalised Anxiety Disorder Scale -7. *Data present until the 17th of May 2020.*

Table 2. *Characteristics of appointments from 1st January 2020 to 24th May 2020*

n	248,628
Attendance	
Attended	185,150 (74·5)
Cancelled by patient	21,333 (8·6)
Cancelled by provider	12,518 (5·0)
Did Not Attend or Late	29,627 (11·9)
Consultation Medium of Attended Appointments	
Face-to-Face	68,777 (37·1)
Remote	109,767 (59·3)
Other	4,495 (2·4)
Unknown	2,111 (1·1)

*Data are presented as n (%).

Figure 1. *Total weekly referrals from 1st January 2019 to 24th May 2020*

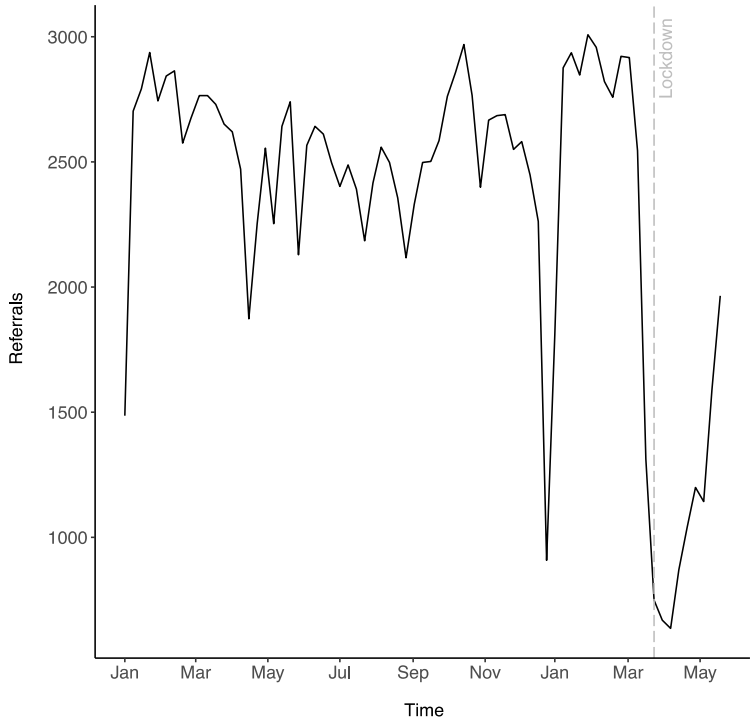


Figure 2. *Total weekly referrals by ethnicity and referral source from 1st January 2019 to 24th May 2020*

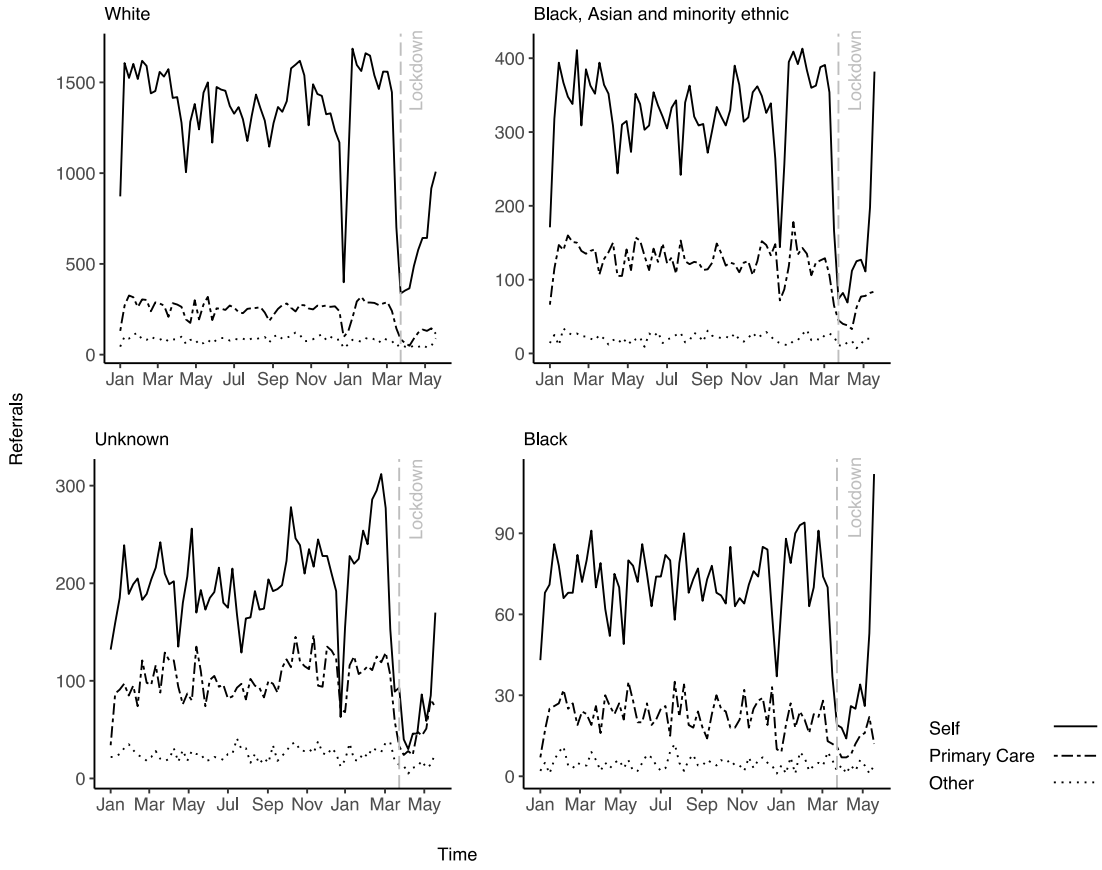


Figure 3. Weekly average age of referrals from 1st January 2019 to 24th May 2020

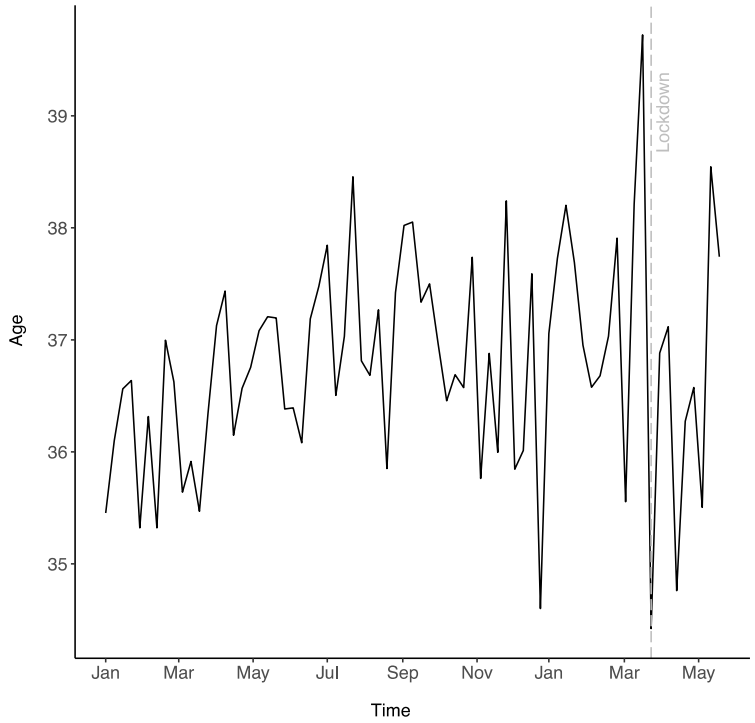


Figure 4. *Weekly average Index of Multiple Deprivation of referrals from 1st January 2019 to 24th May 2020*

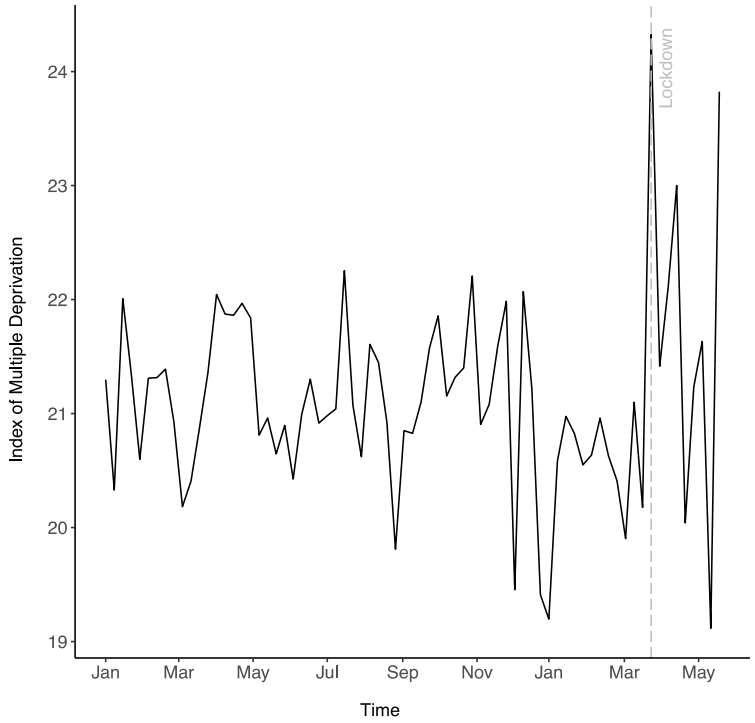


Figure 5. *Weekly average population density of referrals from 1st January 2019 to 24th May 2020*

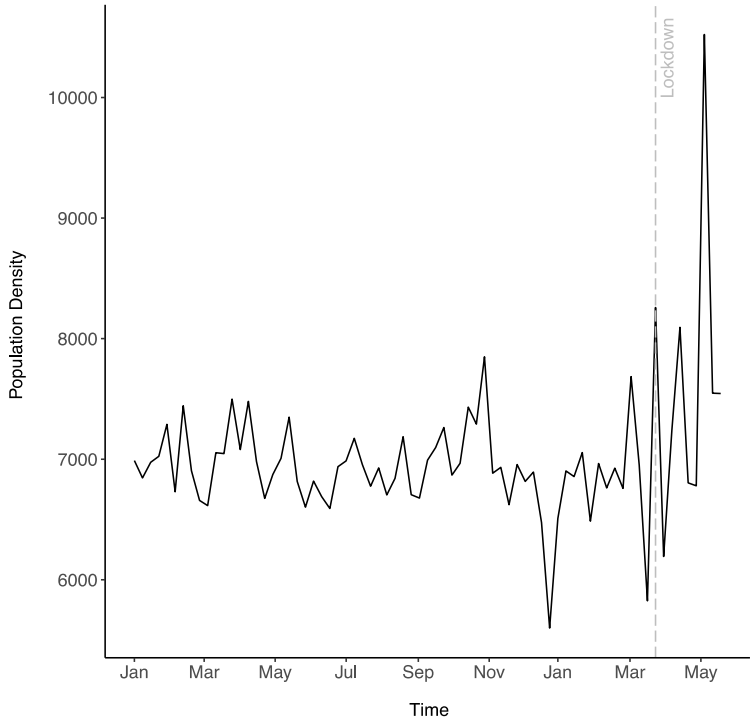


Figure 6. *Weekly average baseline depression and anxiety scores for incoming referrals from 1st January 2019 to 17th May 2020*

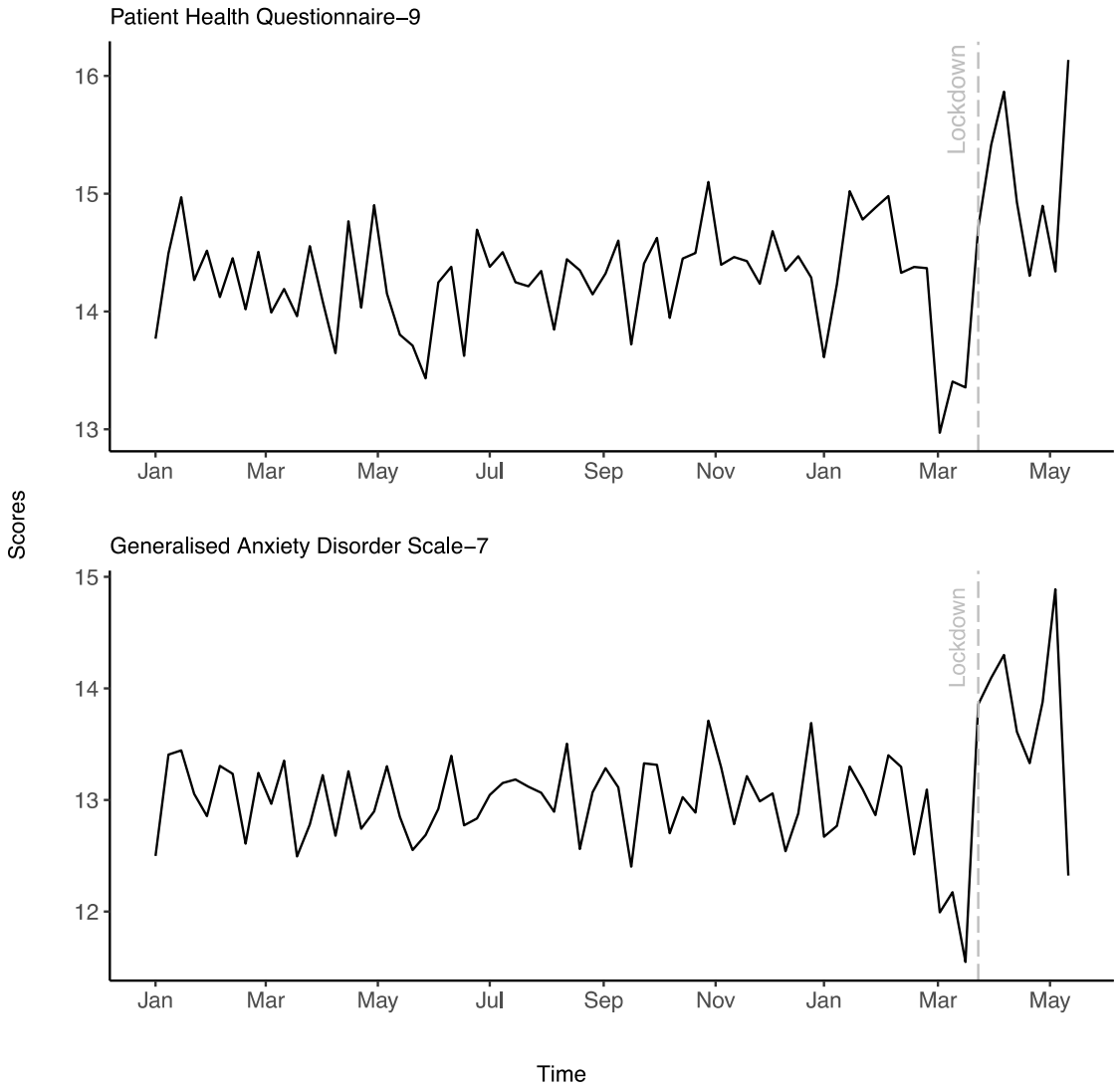


Figure 7. *Weekly average number of previous referrals from 1st January 2019 to 24th May 2020*

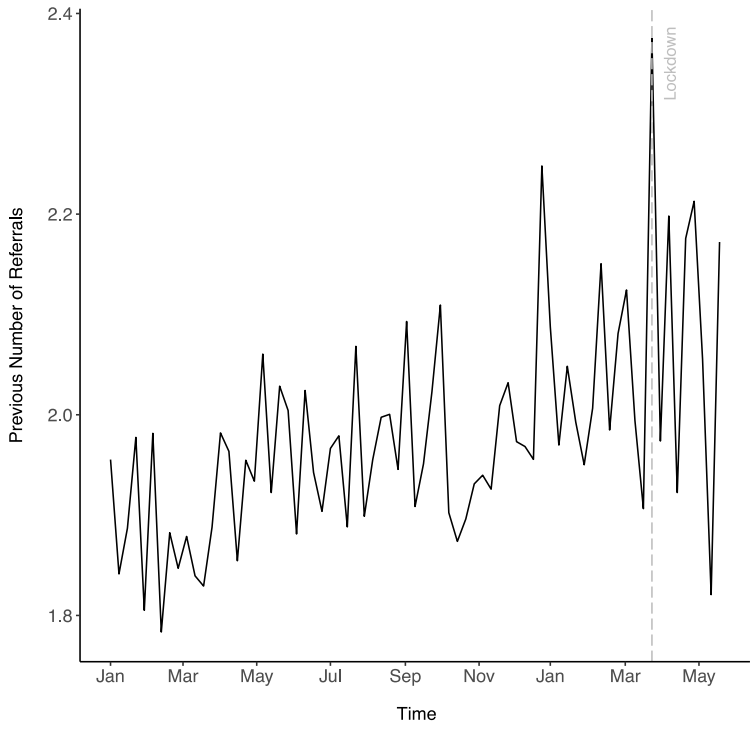
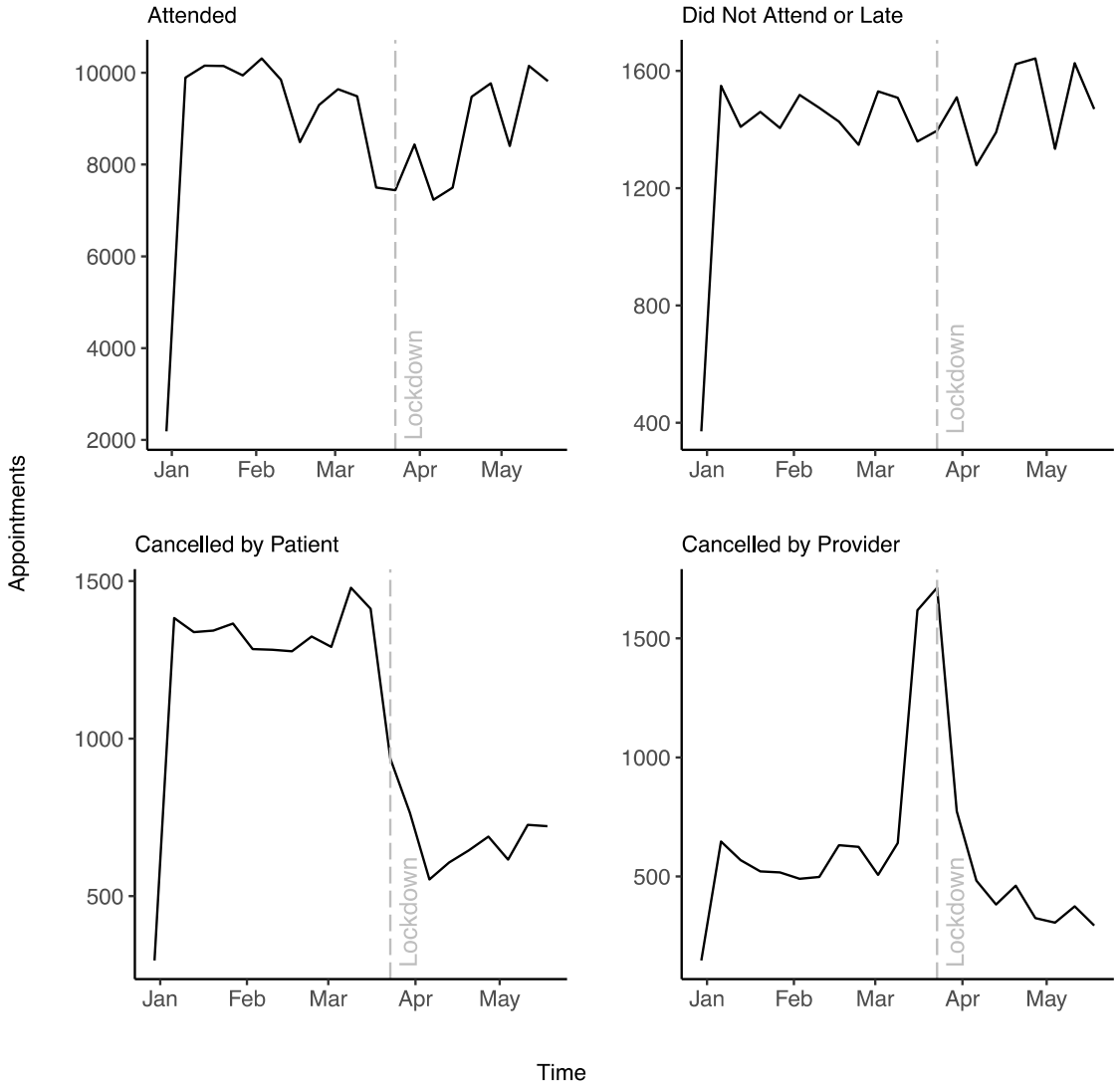


Figure 8. *Weekly total appointments by attendance from 1st January 2020 to 24th May 2020*



7.9 References

1. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient Preference for Psychological vs Pharmacologic Treatment of Psychiatric Disorders. *J Clin Psychiatry* 2013;**74**:595–602. doi:10.4088/jcp.12r07757
2. Clark DM, Canvin L, Green J, Layard R, Pilling S, Janecka M. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet* 2018; **391**:679–686. doi:10.1016/s0140-6736(17)32133-5
3. NHS Digital. Psychological Therapies Annual report on the use of IAPT services 2019-20. [Accessed 2020 August 5]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2019-20
4. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* 2020;**7**:547–560. doi:10.1016/s2215-0366(20)30168-1
5. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, Kontopantelis E, Webb R, Wessely S, McManus S, Abel KM. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *Lancet Psychiatry* 2020;**7**:883-892. doi:10.1016/S2215-0366(20)30308-4
6. Bu F, Steptoe A, Fancourt D. Loneliness during a strict lockdown: Trajectories and predictors during the COVID-19 pandemic in 38,217 United Kingdom adults. *Soc Sci Med* 2020;**265**:113521. doi:10.1016/j.socscimed.2020.113521
7. Iob E, Frank P, Steptoe A, Fancourt D. Levels of severity of depressive symptoms among at-risk groups in the UK during the COVID-19 pandemic. *JAMA Netw Open* 2020;**3**:e2026064. doi:10.1001/jamanetworkopen.2020.26064
8. Moorthy A, Sankar TK. Emerging public health challenge in UK: perception and belief on increased COVID19 death among BAME healthcare workers. *J Public Health (Oxf)* 2020;**42**:486–492. doi:10.1093/pubmed/fdaa096
9. Johnson S, Dalton-Locke C, San Juan NV, Foye U, Oram S, Papamichail A, Landau S, Olive RR, Jeynes T, Shah P, Rains LS. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. *Soc Psychiatry Psychiatr Epidemiol* 2021;**56**:25-37. doi:10.1007/s00127-020-01927-4

10. Chen S, Jones PB, Underwood BR, Moore A, Bullmore ET, Banerjee S, Osimo EF, Deakin JB, Hatfield CF, Thompson FJ, Artingstall JD. The early impact of COVID-19 on mental health and community physical health services and their patients' mortality in Cambridgeshire and Peterborough, UK. *J Psychiatr Res* 2020;**131:**244-254. doi:10.1016/j.jpsychires.2020.09.020
11. Tromans S, Chester V, Harrison H, Pankhania P, Booth H, Chakraborty N. Patterns of use of secondary mental health services before and during COVID-19 lockdown: observational study. *BJPsych Open* 2020;**6**:E117. doi:10.1192/bjo.2020.104
12. Armitage R. Antidepressants, primary care, and adult mental health services in England during COVID-19. *Lancet Psychiatry* 2021;**8**:e3. doi:10.1016/S2215-0366(20)30530-7
13. Johnson, B. (2020, March 23). Prime Minister's statement on coronavirus (COVID-19): 23 March 2020. [Accessed 2020 October 12]. Available from: https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020
14. Noble S, McLennan D, Noble M, Plunkett E, Gutacker N, Silk M, Wright G. The English indices of deprivation 2019. Ministry of Housing, Communities and Local Government, 2019. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833947/IoD2019_Research_Report.pdf
15. Park N. Population estimates by output areas, electoral, health and other geographies, England and Wales: mid-2018. Office for National Statistics, 2019. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualsmallareapopulationestimates/mid2019
16. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a Brief Depression Severity Measure. *J Gen Intern Med* 2001;**16**:606–613. doi:10.1046/j.1525-1497.2001.016009606.x
17. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder. *Arch Intern Med* 2006;**166**:1092–1097. doi:10.1001/archinte.166.10.1092
18. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behav Res Ther* 2009;**47**:910-920. doi:10.1016/j.brat.2009.07.010

19. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org/.
20. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016. Available from: https://ggplot2.tidyverse.org.
21. Eldridge C. Trends in IAPT data. Presented at Digital Innovation in Mental Health Conference; 2020 August 5-6: London, United Kingdom. Available from: https://www.youtube.com/watch?v=AtvO5ryicF8&feature=youtu.be
22. Charlesworth A. Shock to the system: COVID-19's long-term impact on the NHS. [Accessed 2020 May 28]. Available from https://www.health.org.uk/news-and-comment/blogs/shock-to-the-system-covid-19s-long-term-impact-on-the-nhs
23. The Health Foundation. Visits to A&E departments in England in April 2020 fell by 57% compared to last year. [Accessed 2020 May 15]. Available from https://www.health.org.uk/news-and-comment/charts-and-infographics/visits-to-a-e-departments-in-england-in-april-2020-fell-by-57
24. Elwell-Sutton T, Deeney S, Stafford M. Emerging findings on the impact of COVID-19 on black and minority ethnic people. [Accessed 2020 May 20]. Available from https://www.health.org.uk/news-and-comment/charts-and-infographics/emerging-findings-on-the-impact-of-covid-19-on-black-and-min
25. Fancourt D, Bu F, Mak H, Steptoe A. Covid-19 Social Study Results Release 15. [Accessed 2020 May 15]. Available from https://b6bdcb03-332c-4ff9-8b9d-28f9c957493a.filesusr.com/ugd/3d9db5_17cc74c304664db8ac9ea56e1dd301ae.pdf
26. Caul S. Deaths Involving Covid-19 by Local Area and Socioeconomic Deprivation: Deaths Occurring Between 1 March and 31 July 2020. Office for National Statistics. Available from: https://backup.ons.gov.uk/wp-content/uploads/sites/3/2020/08/Deaths-involving-COVID-19-by-local-area-and-socioeconomic-deprivation-deaths-occurring-between-1-March-and-31-.pdf
27. Cairns M. Patients who come back: Clinical characteristics and service outcome for patients re-referred to an IAPT service. *Couns Psychother Res* 2014 Jan 2;**14**:48-55. doi:10.1080/14733145.2013.770895
28. NHS England. (2020, April 25). Help us help you: NHS urges public to get care when they need it. [Accessed 2020 October 21] Available from: https://www.england.nhs.uk/2020/04/help-us-help-you-nhs-urges-public-to-get-care-when-they-need-it/

29. Naz S, Gregory R, Bahu M. Addressing issues of race, ethnicity and culture in CBT to support therapists and service managers to deliver culturally competent therapy and reduce inequalities in mental health provision for BAME service users. *Cogn Behav Therap* 2019;**12**:e22. doi:10.1017/S1754470X19000060
30. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and Acceptability of Cognitive Behavior Therapy Delivery Formats in Adults With Depression. *JAMA Psychiatry* 2019;**76**(7): 700–707. doi:10.1001/jamapsychiatry.2019.0268
31. Linde K, Rücker G, Sigterman K, Jamil S, Meissner K, Schneider A, et al. Comparative effectiveness of psychological treatments for depressive disorders in primary care: network meta-analysis. *BMC Fam Pract* 2015;**16**:103. doi:10.1186/s12875-015-0314-x
32. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of Telephone-Administered vs Face-to-face Cognitive Behavioral Therapy on Adherence to Therapy and Depression Outcomes Among Primary Care Patients. *JAMA* 2012;**307**: 2278–2285. doi:10.1001/jama.2012.5588
33. NHS Digital. Psychological Therapies: report on the use of IAPT services, England July 2020 Final including reports on the IAPT pilots. [Accessed 2020 November 4]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-report-on-the-use-of-iapt-services/july-2020-final-including-reports-on-the-iapt-pilots/outcomes

7.10 Declaration of Interests

CBS, JF, KB, TS, AD, DB, WW, CE report no conflicts of interest. EG reports grants from the British Psychological Society outside the submitted work and is a member of the British Psychological Society, Division of Clinical Psychology, Digital Healthcare.

7.11 CReDiT Author Statement

Clarissa Bauer-Staeb – Conceptualisation, Methodology, Writing – Original Draft, Formal Analysis, Visualisation.

Alice Davis - Conceptualisation, Methodology, Writing – Review & Editing.

Theresa Smith - Conceptualisation, Writing – Review & Editing.

Emma Griffith - Writing – Review & Editing, Supervision.

Chris Eldridge - Resources, Writing – Review & Editing.

Wendy Wilsher - Data Curation.

David Betts – Data Curation, Resources, Writing – Review & Editing.

Julian Faraway - Conceptualisation, Methodology, Formal Analysis, Writing – Review & Editing, Supervision.

Katherine Button - Conceptualisation, Methodology, Writing – Review & Editing, Supervision.

7.12 Acknowledgements

We would like to thank all service providers and patients whose anonymous data was made available for the present research. We would further like to thank Mayur Bhatt, Joshua Buckman, and Rob Saunders for helpful feedback on a draft of the manuscript.

7.13 Funding

AD and TS were funded by Innovate UK (KTP #11105). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

8 General Discussion

8.1 Summary and conclusions of main findings

Several evidence-based psychotherapies exist for treating depression and anxiety, with research suggesting they are all broadly speaking effective (Cuijpers, 2017). Despite this, only roughly half of patients respond to treatment (Cuijpers et al., 2021). There is also evidence to suggest that the effects of psychotherapies have been overestimated, so there is clear room for improvement (Cuijpers, 2017). One approach to potentially reducing the burden of disease is personalised psychotherapy. In my PhD, I primarily used data from Improving Access to Psychological Therapy (IAPT) services to conduct exploratory analyses in observational data to better understand variability in treatment response, taking different approaches that contribute to personalised psychotherapy.

8.1.1 Study one: the minimal clinically important difference (MCID), personalised to baseline severity

To investigate whether an intervention has worked, a success metric is required. In clinical research and practice, self-report outcome questionnaires are commonly used to evaluate the outcome of treatment. These can be used as continuous measures or categorised. However, such approaches often rely on statistical change alone or statistically defined cut-offs that ignore the patient perspective. Furthermore, due to the complexity of measuring mental states, the question arises of how much change in questionnaires is good enough (Button et al., 2015; Kounali et al., 2020). The MCID is a concept defined as the smallest difference in symptoms where patients experience a subjective improvement (Button et al., 2015; Kounali et al., 2020). There are multiple approaches to estimating the MCID, but triangulation of methodological approaches is necessary to examine whether there is a convergence between methods. Research suggests that the MCID is baseline-dependent; how much change on questionnaires translates into patients feeling better depends on the baseline severity, with severe patients requiring greater changes for these to be noticeable (Button et al., 2015; Kounali et al., 2020). In study one, I adopted the concept of the effective dose to conceptualise the MCID. The effective dose 50 (ED50) is a commonly used methodology in pharmacology research and is often used to identify the minimum clinically effective dose (Kenny & McPhee, 2022). I used data from two randomised controlled trials for depression, PANDA and CoBalT, to develop a generalised additive model that mapped the change in depression and anxiety questionnaire scores onto the patient's subjective experience of

whether they felt better or not (Lewis et al., 2019; Wiles et al., 2013). Subsequently, the ED50 was estimated from the model, indicating the symptom change associated with a 50% probability of feeling better. Given the importance of the baseline, the MCID was personalised to each level of baseline severity. The ED25 and ED75 were also estimated. As with previous research, there was a strong baseline-dependency (Button et al., 2015; Kounali et al., 2020). These metrics can be used as an outcome measure in research or clinical practice or for the purposes of power calculations in trials. For example, this study allowed me to use the MCID as an outcome metric that incorporates the patients' perspectives in study 2. The different ED levels also provide flexibility for clinicians and researchers in deciding how much probability of feeling better is "sufficient" and understanding how previous outcome metrics translate into a probability of feeling better.

8.1.2 Study two: a data-driven model for differential treatment allocation to cognitive behavioural therapy vs. counselling for depression

Much of the previous literature focusing on the personalisation of psychotherapy has investigated individual characteristics as moderating effects in isolation. Furthermore, previous approaches were mainly implemented in randomised controlled trials, which have commonly been underpowered to perform such analyses (Cuijpers, Ebert, Acarturk, Andersson, & Cristea, 2016). In study two, I used a large, retrospective cohort of IAPT patients who had received psychological therapy and focused on understanding whether a differential treatment allocation using an actuarial, data-driven approach could help improve clinical outcomes in the treatment of depression. Specifically, I examined whether baseline characteristics have a moderating effect on post-treatment depression scores in cognitive behavioural therapy vs. counselling for depression. A regression model was developed and subsequently used to predict clinical outcomes in an observed scenario (clinical response to the treatment a patient received) and a counterfactual scenario (clinical response had the patient received the alternate treatment). In a held-out validation sample, the differences in actual post-treatment scores were compared in patients who received their model-indicated optimal treatment vs. their model-indicated suboptimal treatment. There was no strong evidence to suggest that any specific baseline characteristic was a particularly important moderator. However, there was a small benefit in post-treatment scores amongst those who received their model-indicated optimal vs. suboptimal treatment. This suggests that no individual characteristic can be used before treatment to make robust predictions about who should get what treatment; instead, there appears to be a cumulative effect of many minor

differences across multiple characteristics. Using such a data-driven approach to support clinical decision-making in treatment allocation has the potential to personalise treatment, as the risk is low when equally effective therapies are delivered, and the cost of implementing such models is relatively low. However, the effects I found in study two were modest, suggesting any benefit may not be immediately tangible to patients but may be more relevant from a public health perspective.

8.1.3 Study three: individual symptom trajectories of depression and anxiety in cognitive behavioural therapy

There is significant heterogeneity in the presentation of common mental health problems such as depression and anxiety, with symptom profiles varying significantly between individuals (Fried & Nesse, 2015). However, sum scores are frequently used to measure clinical improvements, which may mask any insights into the heterogeneity in clinical outcomes, such as a given treatment being particularly beneficial for specific symptoms. In study three, the focus shifted towards exploring individual symptoms rather than sum scores alone, which assumes symptoms can be used interchangeably to define underlying psychopathology. Here I examined the trajectories of individual depression and anxiety symptoms over the course of cognitive behavioural therapy for depression and generalised anxiety in IAPT, respectively. Furthermore, I compared how rapidly individual symptoms changed relative to one another. While all symptoms reduced across treatment, I found that specific symptoms appear to improve faster relative to other symptoms. Specifically for depression, low mood/hopelessness and guilt/worthlessness improved fastest while sleeping problems, appetite changes, and psychomotor retardation/agitation improved at a relatively slower pace. For generalised anxiety, uncontrollable worry, and too much worry improved more rapidly, with irritability and restlessness improving at a somewhat slower pace. This suggests differences between how fast different symptoms respond during therapy and speculatively indicates that cognitive behavioural therapy is "working on" specific symptoms faster. However, symptoms were compared relative to one another, and there was no control group to evaluate the extent of any natural fluctuations in symptoms over time. As such, the ability to draw conclusions about the efficacy of cognitive behavioural therapy on specific symptoms is very limited and remains speculative. Nonetheless, the symptoms that appeared to change most are considered core symptoms of each disorder, so faster changes appear promising. Understanding these differences in symptom change has the potential to personalise treatment to some degree, such as matching patients to treatments based on their

symptom profile. For example, cognitive behavioural therapy may be a promising treatment option for those who experience more of the symptoms that change faster during treatment. In contrast, adjunct or alternative treatment may be required for those who predominantly experience symptoms that appear to improve more slowly.

8.1.4 Study four: the impact of the COVID-19 pandemic on referrals and appointments in IAPT

Due to the COVID-19 pandemic, I had a unique opportunity to explore the impact of the pandemic on mental health service provision. While this topic is less focused on the personalisation of treatment, it provides insights into the organisational response of adapting care to the challenges of a global pandemic. I conducted a brief, descriptive analysis examining the referrals and their sociodemographic characteristics before and during the COVID-19 pandemic, as well as exploring what happened to appointments. The result indicated a dip in overall referrals to the included IAPT services and a reduction in appointments around the first national lockdown. However, services appeared to adapt rapidly, with referrals and appointments rising steadily after the initial dip. This seemed to be facilitated by a rapid switch to remote mediums of consultation. There were differences in the sociodemographic characteristics of referrals to the participating IAPT services during the early phase of the pandemic, which may indicate that certain groups of people were more impacted by the pandemic in terms of access to mental healthcare. However, due to the descriptive analyses and observational nature of the data, the reasons for this change remain speculative.

8.1.5 Conclusions

In this PhD, I used different methodological approaches aimed at informing personalised psychotherapy. These studies included an outcome metric that is personalised to baseline severity (i.e., the MCID), personalised treatment allocation based on baseline patient characteristics, and a symptom-based approach to examine heterogeneity in clinical outcomes as a potential avenue to the personalisation of psychotherapy. A key question arises: is personalised psychotherapy for depression and anxiety possible? Many patients with similar characteristics appear to have similar clinical outcomes in different treatments. There is a less clear-cut segregation in treatment outcomes compared to biological sciences, where personalised medicine has been more successful. For example, the presence or absence of gene variants producing a more clear-cut disambiguation of treatment outcomes, such as in

oncology. As such, it seems improbable that crude baseline patient characteristics examined in this PhD alone are sufficient to evoke any large differences in clinical outcomes between treatments. It appears that different symptoms may respond differently across cognitive behavioural therapy; however, given the observational nature of the data, this requires replication. Furthermore, whether differential treatment allocation based on specific symptoms is warranted depends on other treatments having different symptom-specific effects. It would only be beneficial to potentially allocate someone to an alternative treatment, such as counselling, based on their symptom profile if the symptoms that improved slower in cognitive behavioural therapy improved faster in counselling. Furthermore, evaluations are required to examine the efficacy and associated effect sizes of such personalised treatment allocation.

Overall, my research, in conjunction with broader research in this area, suggests that potentially small benefits can be made in clinical outcomes with personalised approaches given the currently available data and methodologies used. Still, these are unlikely to translate into large clinical benefits across the board like those observed in, for example, oncology, where treatments can be stratified based on a clearer understanding of the causal pathway. However, it should be noted that such extreme examples of personalisation, such as the BRCA genes and phenylketonuria, are extremely rare. While the risk is low when equally effective treatments are delivered, and the cost of approaches discussed is generally low, there may be a trade-off between effort and size of benefits observed – directing efforts in other research areas may lead to more substantial benefits in reducing the overall burden of disease. It may very well be possible that we have hit the ceiling in terms of the clinical outcomes with current approaches.

8.2 Strengths & limitations/challenges

8.2.1 Strengths

Study one estimated the MCID for two of the most frequently used measures of depression and anxiety – the 9-item Patient Health Questionnaire and the 7-item Generalised Anxiety Disorder Scale (Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, Williams, & Lowe, 2006). The research was based on secondary data from two high-quality randomised controlled trials with a wide range of patients with different baseline severities, which was critical given the baseline dependency of the MCID (Button et al., 2015; Kounali et al., 2020;

Lewis et al., 2019; Wiles et al., 2013). Studies two, three, and four were based on secondary data collected as part of routine care from IAPT. The samples were obtained from a wide range of services in different geographies, with varying patient populations adding to the findings' generalisability. Furthermore, while clinical records come with difficulties, they benefit from larger samples than can be feasibly enrolled in any randomised controlled trial and inherently have a higher ecological validity by being representative of real-world clinical practice. This is an aspect randomised controlled trials commonly fall short of due to the highly controlled settings (Kim, Lee, & Kim, 2018).

8.2.2 Limitations

8.2.2.1 Missing data

Like most observational data, IAPT data suffers from missingness. Due to the session-by-session recording of outcome measures, missing data for clinical scores is less problematic, as even when patients drop out of treatment early, they are likely to have a score at their last attended appointment (Clark, 2018) – 98% of referrals nationally have pre- and post-treatment outcome measures (Clark et al., 2018). An outcome-based payment structure has been implemented in IAPT, where reimbursement for each care episode contains a basic service component as well as an outcome-based element where performance is measured against agreed quality and outcome metrics (NHS England & NHS Improvement, 2018). In addition to outcome measures being a mandatory component of the core IAPT dataset, services may be incentivised to keep the measurement of outcomes high due to the payment structure. Similarly, baseline patient characteristics are often present as they are typically recorded at the point of referral or a routine initial assessment. For missing baseline characteristics, robust data imputation was performed using a random-forest algorithm under the missing at random assumption, which has been shown to perform well with mixed data types (Stekhoven & Bühlmann, 2012). Compared to multiple imputation, single imputation ignores the uncertainty around the imputed values and can frequently underestimate the variance (Jakobsen, Gluud, Wetterslev, & Winkel, 2017). However, multiple imputation is often more computationally expensive. Given the relatively small amount of missing data, single imputation was deemed appropriate when also considering time demands and computing resource availability for the current studies.

One of the most significant limitations of the IAPT data I found during my PhD project is missingness related to the diagnosis, also referred to as the problem descriptor in IAPT. Both patients with depression and anxiety are treated in IAPT with similar approaches such as cognitive behavioural therapy. Treatment protocols for depression and anxiety differ despite being based on the same principles. However, the treatment labels for each session are recorded under the broad category of "cognitive behavioural therapy" without an indication if cognitive behavioural therapy was delivered for depression or anxiety. As such, the diagnosis is the only reliable way to identify who was primarily treated for depression and who was treated for anxiety in the dataset (Clark et al., 2018). The missingness of the problem descriptor is frequently high, at approximately 25% nationally – but there is much variation between services (Clark et al., 2018). The presence of a problem descriptor has been associated with better clinical outcomes after treatment (Clark et al., 2018), potentially suggesting a better quality of care may be delivered to this subgroup of patients. Patient records also contained fewer missing data where a diagnosis was present. The analyses in my PhD project relied heavily on identifying, as accurately as possible, whether treatment was delivered for depression or anxiety. A subgroup of patients with the relevant problem descriptor was identified, excluding any patients with a missing problem descriptor. Many of the excluded patients were likely nonetheless treated for depression or generalised anxiety. Having a subgroup of patients who are likely to have better clinical outcomes may overestimate the findings of the current project. For example, when examining the trajectories of individual symptoms across a course of cognitive behavioural therapy (study three), having a sample with better clinical outcomes will likely inflate the predicted probability of response. However, the aim of the research was not to estimate the magnitude of overall response, which was noted to be interpreted with caution, but rather to draw comparisons between trajectories of different symptoms. More problematic is the possibility that better clinical outcomes are perhaps linked to treatment-related factors such as closer adherence to treatment protocols. If we assume there are specific factors of cognitive behavioural therapy, then the effects showing which symptoms might respond faster may be diluted or not generalise to patients without a problem descriptor where a treatment protocol may not have been as strictly adhered to. With the same assumptions, any benefit of differential treatment allocation may also be diluted amongst the subgroup who do not have a diagnosis label (study two). However, there is a lack of research as to why missingness may occur in the diagnosis label, and there are no treatment fidelity measures which would assess this; as such, the assumptions remain speculative (Martin, Iqbal, Airey, & Marks, 2022).

8.2.2.2 Causal inference and confounding

With most observational data making inferences about causality is not possible. Unlike with randomised controlled trials, there is no control group to which comparisons can be made. For example, in study three, we compare the trajectories of individual depression and anxiety questionnaire items throughout a course of cognitive behavioural therapy. Whilst providing a form of internal control by comparing one trajectory to the average of all other trajectories, we cannot be sure that cognitive behavioural therapy caused these changes – fluctuations in symptoms may occur naturally over time, and certain symptoms may be more amenable to natural recovery. A control group would allow for more robust causal inference. As such, we may over- or underestimate the effect cognitive behavioural therapy has on specific symptoms depending on the natural trajectory of symptoms without treatment. Furthermore, the data presented in study two, where the benefit of using baseline characteristics for differential treatment allocation was used, suggests there are differences in baseline characteristics between patients in different treatment groups. This could simply be the result of different services having different patient populations and varying availability of different treatments, which could incorrectly appear to suggest systemic differences in the allocation of treatment. However, it is also likely that clinicians are already personalising care for their patient's needs to some degree, and therefore, allocation may not be random. Various ways to adjust for confounding exist, including regression adjustment and propensity score methods (Elze MC et al., 2017). In the present PhD project, propensity scores denote the probability of receiving one treatment over another based on known covariates. These can then be used to perform propensity score matching, inverse probability weighting, stratification on the propensity score, or regression adjustment using the propensity score (as completed in the present PhD in study two) (Austin, 2011). However, only known characteristics can be adjusted, leaving the possibility of unmeasured confounding. The clinical decision-making, psychotherapy process, and the response of individuals are likely dependent on many more complex idiosyncrasies that cannot be accounted for from crude characteristics alone, leaving a strong probability of unmeasured confounding.

8.2.2.3 Misclassification bias

The analyses for this project also relied heavily on identifying which treatment was delivered. Treatment is recorded on a session-by-session basis in IAPT. Whilst the treatment label does not change for many referrals, some treatment episodes have missing data for some

appointments, and others have different treatment labels within the same care episode. The latter can result from patients being switched to a further treatment where the labels change systematically and consistently with time. More difficult to interpret are scenarios where the majority of labels are for one treatment, and there are one, or less commonly, a few different labels dispersed across appointments. Speculatively, this may result from a patient having to see a different therapist on a given day due to absences or changes in appointment days. The studies in the present project used relatively strict criteria for defining treatment whilst still trying to be as exhaustive as possible. Misclassification is therefore unlikely to be present for many patients, but undoubtedly some episodes have been misclassified in terms of the defined treatment. If we assume treatments have specific effects, misclassification is likely to dilute the treatment groups. It is, therefore, more likely to result in an underestimation, potentially resulting in a slightly more conservative estimate than expected if no misclassification occurred.

8.2.3 Challenges

Some of the key methodological challenges and limitations of using observational clinical records are outlined above. However, one somewhat neglected challenge is the difficulty and time of obtaining clinical records in the first place and converting these into useable datasets. Obtaining clinical records for my PhD took over two years, which was challenging given the average PhD duration is three years. Outlined below are some of the processes and associated challenges of obtaining routine clinical records.

8.2.3.1 Approvals

Prior to seeking approvals, it was necessary to find service providers that would consent to their data being used in my PhD research. This was challenging as I did not have a clinical background or contacts in the NHS. Services were approached based on convenience through a combination of academic, clinical, and industry partners making introductions, which was a lengthy process in and of itself. Furthermore, any research involving NHS data requires approval by the Health Research Authority (HRA). Alongside the HRA approval, data sharing agreements had to be developed with support from the University of Bath's and Mayden's legal advisors, agreed upon, and signed between Mayden (the data holders) and the NHS (the data owners). Furthermore, data sharing agreements were signed between Mayden and the University of Bath. Due to the anonymous nature of the data, the research was exempt from NHS Research Ethics Committee approval and only required local ethics from

the University of Bath's Psychology Research Ethics Committee. After these approvals were in place, local approval by the Research and Development (R&D) team at individual NHS trusts were required. Data sharing agreements were amended as necessary by the responsible person in the local NHS trust (for example, the local information governance team) to fit the individual trust's preferences and requirements. Assuring all these approvals are in place was an extensive and time-intensive process, which was complicated by the General Data Protection Regulations (GDPR) being implemented during my PhD.

8.2.3.2 Data Extraction

Once approvals were in place, data extraction commenced. For the present PhD, Mayden offered support-in-kind for the data extraction, making my PhD research feasible without the necessity of extra funding, for which I was incredibly grateful. To comply with the approvals, Mayden also anonymised all data removing any identifiable information from the clinical records. Obtaining records from the developers of the IAPT patient management system has the benefit of obtaining more detailed, system-level records that are not part of the core IAPT dataset, such as individual item outcome measure scores. However, data extraction was also a lengthy process that took several months as this was one of the first projects Mayden supported with this kind of research. As such, extraction procedures had to be developed and optimised.

8.2.3.3 Data Processing and Cleaning

As stated, one of the benefits of obtaining data via Mayden is the availability of more detailed, system-level records and support-in-kind. However, these data extractions are not curated for research purposes, making data reconciliation challenging. For someone without a background in data science, it was an invaluable opportunity to learn about complex data management, the processing of data, and the curation of new datasets. The data extracts were provided in multiple CSV files relating to different aspects of the data, such as patient information, referral-specific data, or records relating to appointments, which then had to be compiled into one dataset. Specific challenges include managing the wealth of data across hundreds of files as well as not working internally at Mayden and therefore lacking in-depth knowledge of the system backend and a more practical understanding of how the data was managed and stored once inputted by clinicians. As such, much exploration was required to understand how the different data tables relate to one another and how best to collate them into a coherent dataset. In addition to standard data cleaning practices such as variables

selection, generation of variables, and the exploration of the data, decision making and rules about data cleaning were complicated by the clinical records not being curated for research purposes. For example, it appears that all data points that were ever entered in the systems are stored. As a result, a scenario might be that multiple outcomes scores are recorded for one appointment. This seems improbable from a clinical perspective as only one questionnaire is collected per appointment. Two plausible explanations might include a system error of attaching the wrong appointment ID to the outcome data or a clinician updating an entry due to an incorrect entry in the first instance. It is impossible to tell from the available data, but a decision had to be made to shape the data into a usable format. Exploring the data and creating rules for specific data processing steps is also complicated when multiple services are used, as operational discrepancies across services are also likely. As such, the dataset will likely be less accurate than data collected purely for research purposes.

8.2.3.4 Recommendations

Overall, using clinical healthcare records opens many possibilities for secondary data analysis and therefore has clear benefits. However, it is also important not to underestimate the groundwork required to operationalise the generation of a new dataset from records that are not curated for research purposes. Whilst the current project is a time-limited project, given the groundwork required, I would suggest it is important to maximise the benefit by establishing it as collaborative research used for multiple projects. This also has the benefit of quality assurance, where a robust data processing pipeline can be established based on a team member's expertise. Receiving new records annually would result in up-to-date data being available for both exploratory research and evaluations of models in previously unseen data. An alternative to an individual or collaborative project may be a national effort to create a register, such as the Clinical Practice Research Datalink (CPRD), which contains primary care data (Herrett et al., 2015). This would likely increase accessibility to individual-level psychological therapy records and simplify the application and approval through a standardised process.

8.3 Future directions for personalised psychotherapy

Given the relatively small benefits observed across the studies in my PhD and the wider literature, the question arises: is there a future for personalised psychotherapy?

8.3.1 Do we have the right type of data?

My PhD research was limited by data routinely collected as part of clinical practice. Other research has found potential characteristics that have moderating effects on treatment outcomes between therapies. These have typically included more fine-grained clinical, cognitive, and psychological factors, such as "somatic complaints, cognitive problems, paranoid symptoms, interpersonal self-sacrificing, attributional style and number of life events" (Huibers et al., 2015, p.1). However, there has not been any clear consensus and replication across research. Furthermore, important factors to better understand variation in treatment response are not limited to the patient per se but also treatment-related factors. The only systematic recording of the treatment in IAPT that is comparable between services is the overarching labels for treatment, such as "cognitive behavioural therapy". Given the lack of fidelity measures in IAPT, inferences about the specific details of how treatment protocols are delivered are therefore limited. As such, incorporating further data in routine care may provide researchers with data needed for more substantial advances in personalised psychotherapy. However, a word of caution in doing so is ensuring any additional measures to be completed as part of routine care needs to be balanced with pragmatism in terms of what is feasible for service users and staff. As researchers, we often have a mantra that *more data is better,* and if we just had sufficient information, a data mining exercise may lead to new developments. However, the time clinicians spend "collecting data" detracts from time treating patients. It may be necessary to rely more on smaller experimental or exploratory studies with promising results, or theory-driven hypotheses, to inform whether further variables should be included as part of routine care.

A potential avenue may be a new perspective discussed in the field of nutrition, which distinguishes between personalised medicine and precision medicine (Betts & Gonzalez, 2016). In most previous literature, and for the purposes of this PhD, broad terms such as personalised, precision, stratified, or targeted medicine/psychotherapy are used interchangeably to denote the same concept of improving clinical outcomes by providing treatments to patients based on who they are. In this thesis, who you are was conceptualised holistically, including more static characteristics as well as more malleable ones such as life events or your lifestyle. However, Betts and Gonzalez suggest drawing a distinction between targeting based on who we are (personalised medicine) and targeting based on daily behaviour (precision medicine) (Betts & Gonzalez, 2016). Thus, separating the concepts of

who we are (more static traits) and daily life (more malleable states), with the authors suggesting that daily life matters more than who we are (Betts & Gonzalez, 2016).
Such an approach appears to align conceptually with trauma-informed care, where there is an emphasis on managing patient care in the context of *what happened to you* rather than who you are (Harris & Fallot, 2001). Trauma-informed care takes the approach of framing and understanding patients' behaviour/experiences in the context of coping strategies that were developed to manage and cope with overwhelming circumstances or adversity (Huang et al., 2014). These circumstances or adverse experiences could include, for example, adverse childhood experiences (Oral et al., 2016). Such an approach has also been emphasised in the NHS's mental health implementation plan with the aim of increasing access to services that are underpinned by a trauma-informed approach (NHS England, 2019). As such, it may be worthwhile exploring how trauma and life events, such as adverse childhood experiences, influence variability in treatment outcomes with the possibility of using trauma data to support personalised psychotherapy. However, this data was unavailable in routine IAPT data during my PhD.

8.3.2 Mechanisms of depression, anxiety, and psychotherapy

Despite decades of research and multiple putative propositions, the underlying causes and mechanisms of depression and anxiety are still relatively poorly understood (Craske et al., 2017; Malhi & Mann, 2018). Furthermore, the validity of mental health diagnoses for understanding mechanisms has been questioned by some – most mental disorders do not fit into neatly defined categories with natural boundaries (Kendell & Jablensky, 2003). The lack of a true understanding of depression and anxiety limits the ability to develop or deliver treatments that target problematic mechanisms precisely.

Perhaps shifting the research focus to more transdiagnostic approaches, as conceptualised in the Research Domain Criteria (RDOC), may be more fruitful (National Institute of Mental Health). The research framework RDOC shifts away from the approach that diagnostic manuals take where symptoms are clustered to create discrete definitions of psychopathology (National Institute of Mental Health). Here, the focus is on identifying biological-based conceptualisations of traits/states (National Institute of Mental Health). The framework defines six main functional domains (arousal/regulatory, positive valence, negative valence, sensorimotor, cognitive, and social processes) to be understood in relation to mental health. The research agenda also aims to understand the development/lifespan aspects of these

processes, studied across various units of analysis (such as physiological or behavioural) as well as exploring the environmental impacts (such as social determinants or physical environment) (National Institute of Mental Health). Such transdiagnostic approaches may be more beneficial when attempting to pinpoint mechanisms where there is much heterogeneity and patients may share very few symptom profiles within diagnostic classifications, such as in depression and anxiety (Fried & Nesse, 2015; National Institute of Mental Health). Furthermore, given the comorbidities that frequently exist between mental disorders, particularly depression and anxiety, questions have arisen whether viewing disorders as discrete and separate problems allows commonalities between disorders to be appropriately understood (Johansson, Carlbring, Heedman, Paxling, & Andersson, 2013; National Institute of Mental Health). Clinical classifications also often rely on thresholds, which limits an understanding of the full spectrum of mental health disorders (National Institute of Mental Health). This perspective is also present in clinical practice where some have advocated for more transdiagnostic, formulation-based approaches to conceptualise patients' problems rather than these conceptualisations being predominantly diagnostically led (Johnstone, 2017). It is possible that adopting a more transdiagnostic perspectives may lead to a better understanding of the mechanisms of depression and anxiety.

Similar to a lack of insight into the exact mechanisms of depression and anxiety, it is not well-established what psychotherapy is specifically doing and the mechanisms of how psychotherapy produces change (Carey, Griffiths, Dixon, & Hines, 2020; Cuijpers, Cristea, Karyotaki, Reijnders, & Hollon, 2019). Therapeutic effects are commonly classified as common factors shared between therapies (such as therapeutic alliance) and specific factors, which are elements that are unique to particular interventions (such as emotion regulation) (Cuijpers, Cristea, et al., 2019). An approach in process research to better understand potential mechanisms is examining mediators of change. However, establishing causality has been problematic using such methods, and they have been criticised for not being empirically valid enough to constitute a definitive mechanism (Cuijpers, Reijnders, & Huibers, 2019). Component studies have been used to disentangle the therapy process. Here, studies compare a therapy protocol to one where a specific component is removed or added (Cuijpers, Cristea, et al., 2019). However, meta-analytic evidence suggests problems with low statistical power and quality of the available evidence (Cuijpers, Cristea, et al., 2019). More focus on the quality of evidence in process research or the development of methodologies to understand mechanisms that underpin psychotherapies is required.

Overall, a better understanding of the mechanisms of both psychopathology and psychotherapy are undoubtedly going to contribute to any efforts being made in the field of personalised psychotherapy by allowing therapies to be matched to clinical presentations based on mechanisms of both psychopathology and the intervention. Alternatively, they may also provide insights into where a gap between the two exist; thereby, informing any development or modification of psychotherapies or other treatments that may be required to target specific mechanisms.

8.3.3 Measurement of depression and anxiety

A further issue in the field lies in the measurement of depression and anxiety. For example, while over 280 measures of depression exist, there is a rather baffling lack of consensus on what actually *should* be measured (Fried, Flake, & Robinaugh, 2022). In a comparison of 7 frequently used questionnaires, the content overlap showed 52 differing symptoms, of which 40% were only present on one scale (Fried, 2017; Fried et al., 2022). It is therefore perhaps unsurprising that low-to-moderate correlations often exist between measures (Fried et al., 2022; Fried, van Borkulo, et al., 2016; Furukawa et al., 2020). The patient reports also don't always match the clinician's perspective, with moderate associations between self-reported symptoms and observer ratings (Fried et al., 2022; Hershenberg et al., 2020; Kaiser, Herzog, Voderholzer, & Brakemeier, 2022). Furthermore, subjective patient experiences don't always align with self-report questionnaires themselves - there is a 55% mismatch between patients' perceptions of improvement and symptom improvement as measured by questionnaires (Hobbs et al., 2020). Further, research focusing on the dimensionality of scales suggests there is a lack of unidimensionality; that is, the assumption that scales effectively measure one underlying construct (Fried et al., 2022; Fried, van Borkulo, et al., 2016). Thus, suggesting depression measures may be tapping into distinct constructs. While findings are mixed, some evidence also suggests that depression measures do not always meet the assumption of measurement invariance (measuring the same constructs consistently) – that psychometric properties may shift across time and people (Fried et al., 2022; Fried, van Borkulo, et al., 2016; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004). How we then use these questionnaires to define if patients have improved also varies greatly. A review of trials showed that within 27 trials, there were 47 different definitions of remission (Courtney et al., 2021; Fried et al., 2022). Overall, the evidence suggests there is a lack of clear consensus on how to reliably measure constructs like depression and anxiety.

Without a thorough understanding of what we are measuring and what the measurement means, interpretations of any findings will be caveated. Eiko Fried and colleagues have proposed a series of steps to develop and iterate the methodological basis of measurements (Fried et al., 2022). One key suggestion is that clear theories are needed that can then be embedded in measures - "without a clear theory, it is unclear what we ought to measure, and how to evaluate whether we have succeeded in doing so" (Fried et al., 2022, p.364). Explicit theories should ideally be formally defined, with core assumptions clearly outlined so that they can subsequently be used to evaluate the measures against (Fried et al., 2022). A second suggestion is the need for epistemic iteration, whereby theories and measures are repeatedly developed, tested, and updated based on the novel insights of each iteration. A clear example of the benefit of such iterations is given in relation to the common cause and network theories of mental health problems. The common cause hypothesis posits that there is an underlying cause for psychopathology and that symptoms are manifestations of this cause; effectively, "depression" and "anxiety" produce symptoms (Fried et al., 2022). Because the common cause hypothesis assumes that a universal construct is being measured and symptoms are just interchangeable indicators of the severity of this construct, questionnaire sum scores are often used (Fried et al., 2022). The criticisms or problems inherent within the common cause hypothesis has given rise to network theory. This theory suggests that symptoms are not merely manifestations of a latent cause; rather, depression and anxiety may be a complex, causal system arising from interactions between symptoms (Borsboom, 2017; Fried et al., 2022). This theory is now being explored extensively with a greater focus on the symptoms per se rather than sum scores alone (Beard et al., 2016; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; Mullarkey, Marchetti, & Beevers, 2019). It is also noted that a broad range of people and cultures must be kept in mind rather than developing measures solely on the basis of WEIRD (western, educated, industrialised, rich and democratic) people (Fried et al., 2022). Lastly, they recommend that any new or altered measure requires response research to examine how different people interpret and respond to questions in the measures to assure consistency in what responses mean (Fried et al., 2022).

8.3.4 Alternatives to personalised psychotherapy

The research required to better understand the mechanisms of depression and anxiety, the mechanisms of psychotherapy, as well as advances in how we measure depression and anxiety are likely to be decades away, requiring great cumulative research efforts. In the

meantime, more novel, experimental treatments may be a more effective alternative or a means of augmenting current psychotherapies. Treatments considered relatively novel a few years ago have now become part of mainstream practice, such as mindfulness, which is now delivered in IAPT. Ketamine, psilocybin, and MDMA-assisted psychotherapy are among some of the treatments that are currently being investigated for their therapeutic effects (Conley, Norwood, Hatvany, Griffith, & Barber, 2021; Goldberg, Pace, Nicholas, Raison, & Hutson, 2020; Tedesco et al., 2021). However, their long-term efficacy, cost-effectiveness, and safety are yet to be established. Therefore, it is yet to be determined if they provide a sufficient additional benefit to replace or augment existing therapies. A further alternative may be to shift the focus from treating depression and anxiety to the prevention of common mental health problems. Instead of focusing on treating disorders, it may be helpful to try and reduce the incidence of depression and anxiety in the first place. To date, prevention initiatives, like school-based or e-Health programs, have yet to show robust evidence for long-term benefits (Deady et al., 2017; Feiss et al., 2019). However, comparatively fewer research efforts have been dedicated to this field.

8.4 Conclusion

IAPT has been an immense effort to increase the availability of psychological therapies in England. The IAPT dataset is one of its kind and provides researchers with an invaluable resource that is difficult to find elsewhere at such a scale. Overall, the studies in my PhD showed that treatment personalisation had modest clinical benefits. Substantial research efforts are likely required in the field to greatly improve the modest outcomes in psychotherapy for depression and anxiety, including but not limited to understanding what depression and anxiety are, how we effectively measure them, and the mechanisms that underpin psychotherapy. In the meantime, it may be most pragmatic to increase the availability of high-quality treatment further and focus on community engagement to reduce the overall burden of disease.

8.5 References

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behaviour Research, 46*(3), 399-424. https://doi.org/10.1080/00273171.2011.568786

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., . . . Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine, 46*(16), 3359-3369. https://doi.org/10.1017/S0033291716002300

Betts, J. A., & Gonzalez, J. T. (2016). Personalised nutrition: What makes you so special? *Nutrition Bulletin, 41*(4), 353-359. https://doi.org/10.1111/nbu.12238

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5-13. https://doi.org/https://doi.org/10.1002/wps.20375

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., . . . Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychological Medicine, 45*(15), 3269-3279. https://doi.org/10.1017/s0033291715001270

Carey, T. A., Griffiths, R., Dixon, J. E., & Hines, S. (2020). Identifying Functional Mechanisms in Psychotherapy: A Scoping Systematic Review. *Frontiers in Psychiatry, 11*, 291. https://doi.org/10.3389/fpsyt.2020.00291

Clark, D. (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., 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

Conley, A. A., Norwood, A. E. Q., Hatvany, T. C., Griffith, J. D., & Barber, K. E. (2021). Efficacy of ketamine for major depressive episodes at 2, 4, and 6-weeks post-treatment: A meta-analysis. *Psychopharmacology, 238*(7), 1737-1752. https://doi.org/10.1007/s00213-021-05825-8

Courtney, D. B., Watson, P., Chan, B. W., Bennett, K., Krause, K. R., Offringa, M., . . . Zentner, T. (2021). Forks in the road: Definitions of response, remission, recovery, and other dichotomized outcomes in randomized controlled trials for adolescent

depression. A scoping review. *Depression and Anxiety, 38*(11), 1152-1168. https://doi.org/10.1002/da.23200

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H.-U. (2017). Anxiety disorders. *Nature Reviews Disease Primers, 3*(1), 17024. https://doi.org/10.1038/nrdp.2017.24

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Canadian Psychology/Psychologie Canadienne, 58*(1), 7-19. https://doi.org/10.1037/cap0000096

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research, 29*(1), 15-29. https://doi.org/10.1080/10503307.2017.1395922

Cuijpers, P., Ebert, D. D., Acarturk, C., Andersson, G., & Cristea, I. A. (2016). Personalized Psychotherapy for Adult Depression: A Meta-Analytic Review. *Behavior Therapy, 47*(6), 966-980. https://doi.org/10.1016/j.beth.2016.04.007

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*(3), 288-299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Reijnders, M., & Huibers, M. J. H. (2019). The Role of Common Factors in Psychotherapy Outcomes. *Annual Review of Clinical Psychology, 15*, 207-231. https://doi.org/10.1146/annurev-clinpsy-050718-095424

Deady, M., Choi, I., Calvo, R. A., Glozier, N., Christensen, H., & Harvey, S. B. (2017). eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. *BMC Psychiatry, 17*(1), 310. https://doi.org/10.1186/s12888-017-1473-1

Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, . . . Pocock, S. J. (2017). Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *Journal of the American College of Cardiology, 69*(3), 345-357. https://doi.org/10.1016/j.jacc.2016.10.060

Feiss, R., Dolinger, S. B., Merritt, M., Reiche, E., Martin, K., Yanes, J. A., . . . Pangelinan, M. (2019). A Systematic Review and Meta-Analysis of School-Based Stress, Anxiety,

and Depression Prevention Programs for Adolescents. *Journal of Youth and Adolescence, 48*(9), 1668-1685. https://doi.org/10.1007/s10964-019-01085-0

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders, 208,* 191-197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders, 189,* 314-320. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., Flake, J. K., & Robinaugh, D. J. (2022). Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology, 1*(6), 358-368. https://doi.org/10.1038/s44159-022-00050-2

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 172,* 96-102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28*(11), 1354–1367. https://doi.org/10.1037/pas0000275

Furukawa, T. A., Reijnders, M., Kishimoto, S., Sakata, M., DeRubeis, R. J., Dimidjian, S., . . . Cuijpers, P. (2020). Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking. *Epidemiology and Psychiatric Sciences, 29*, e24. https://doi.org/10.1017/S2045796019000088

Goldberg, S. B., Pace, B. T., Nicholas, C. R., Raison, C. L., & Hutson, P. R. (2020). The experimental effects of psilocybin on symptoms of anxiety and depression: A meta-analysis. *Psychiatry Research, 284*, 112749. https://doi.org/10.1016/j.psychres.2020.112749

Harris, M., & Fallot, R. (2001). New directions for mental health services: Using trauma theory to design service systems. *San Francisco, CA: Josey-Bass*.

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology, 44*(3), 827-836. https://doi.org/10.1093/ije/dyv098

Hershenberg, R., McDonald, W. M., Crowell, A., Riva-Posse, P., Craighead, W. E., Mayberg, H. S., & Dunlop, B. W. (2020). Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. *Journal of Affective Disorders, 266*, 22-29. https://doi.org/10.1016/j.jad.2020.01.108

Hobbs, C., Lewis, G., Dowrick, C., Kounali, D., Peters, T. J., & Lewis, G. (2020). Comparison between self-administered depression questionnaires and patients' own views of changes in their mood: a prospective cohort study in primary care. *Psychological Medicine, 51*, 853 - 860. https://doi.org/10.1017/S0033291719003878

Huang, L. N., Flatow, R., Biggs, T., Afayee, S., Smith, K., Clark, T., & Blake, M. (2014, 2014-07). *SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach*. http://hdl.handle.net/10713/18559

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. *PloS One, 10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology, 17*(1), 162. https://doi.org/10.1186/s12874-017-0442-1

Johansson, R., Carlbring, P., Heedman, Å., Paxling, B., & Andersson, G. (2013). Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. *PeerJ, 1*, e98. https://doi.org/10.7717/peerj.98

Johnstone, L. (2017). Psychological Formulation as an Alternative to Psychiatric Diagnosis. *Journal of Humanistic Psychology, 58*(1), 30-46. https://doi.org/10.1177/0022167817722230

Kaiser, T., Herzog, P., Voderholzer, U., & Brakemeier, E.-L. (2022). Out of sight, out of mind? High discrepancy between observer- and patient-reported outcome after routine inpatient treatment for depression. *Journal of Affective Disorders, 300*, 322-325. https://doi.org/10.1016/j.jad.2022.01.019

Kendell, R., & Jablensky, A. (2003). Distinguishing Between the Validity and Utility of Psychiatric Diagnoses. *American Journal of Psychiatry, 160*(1), 4-12. https://doi.org/10.1176/appi.ajp.160.1.4

Kenny, B. J., & McPhee, A. S. (2022). ED50. In *StatPearls*. Treasure Island (FL): StatPearls Publishing

Kim, H. S., Lee, S., & Kim, J. H. (2018). Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. *Journal of Korean Medical Science, 33*(34), e213. https://doi.org/10.3346/jkms.2018.33.e213

Kounali, D., Button, K. S., Lewis, G., Gilbody, S., Kessler, D., Araya, R., . . . Lewis, G. (2020). How much change is enough? Evidence from a longitudinal study on depression in UK primary care. *Psychological Medicine, 52*(10), 1875-1882. https://doi.org/10.1017/S0033291720003700

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

Lewis, G., Duffy, L., Ades, A., Amos, R., Araya, R., Brabyn, S., . . . Lewis, G. (2019). The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *The Lancet Psychiatry, 6*(11), 903-914. https://doi.org/10.1016/S2215-0366(19)30366-9

Malhi, G. S., & Mann, J. J. (2018). Depression. *The Lancet, 392*(10161), 2299-2312. https://doi.org/10.1016/S0140-6736(18)31948-2

Martin, C., Iqbal, Z., Airey, N. D., & Marks, L. (2022). Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: How can it better meet the range of primary care mental health needs? *British Journal of Clinical Psychology, 61*(1), 157-174. https://doi.org/10.1111/bjc.12314

Mullarkey, M. C., Marchetti, I., & Beevers, C. G. (2019). Using network analysis to identify central symptoms of adolescent depression. *Journal of Clinical Child and Adolescent Psychology, 48*(4), 656-668. https://doi.org/10.1080/15374416.2018.1437735

National Institute of Mental Health. *Research Domain Criteria (RDoC)*. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc

Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African Americans compared

with a nationally representative sample. *Psychiatry Research, 126*(2), 177-187. https://doi.org/10.1016/j.psychres.2004.02.004

NHS England. (2019). *NHS Mental Health Implementation Plan 2019/20–2023/24*. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/07/nhs-mental-health-implementation-plan-2019-20-2023-24.pdf

NHS England, & NHS Improvement. (2018). *Paying for quality and outcomes: IAPT case studies*. https://www.england.nhs.uk/wp-content/uploads/2018/04/iapt-payment-casestudy.pdf

Oral, R., Ramirez, M., Coohey, C., Nakada, S., Walz, A., Kuntz, A., . . . Peek-Asa, C. (2016). Adverse childhood experiences and trauma informed care: the future of health care. *Pediatric Research, 79*(1), 227-233. https://doi.org/10.1038/pr.2015.197

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

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

Tedesco, S., Gajaram, G., Chida, S., Ahmad, A., Pentak, M., Kelada, M., . . . Jolayemi, A. (2021). The Efficacy of MDMA (3,4-Methylenedioxymethamphetamine) for Post-traumatic Stress Disorder in Humans: A Systematic Review and Meta-Analysis. *Cureus, 13*(5), e15070. https://doi.org/10.7759/cureus.15070

Wiles, N., Thomas, L., Abel, A., Ridgway, N., Turner, N., Campbell, J., . . . Lewis, G. (2013). Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *The Lancet, 381*(9864), 375-384. https://doi.org/10.1016/S0140-6736(12)61552-9

9 Supplementary Material

9.1 Supplementary Material A: Chapter 3

General Introduction

The overview of previous literature presented in the general introduction of this thesis exploring moderators of treatment response in depression and generalised anxiety was a smaller subset of a wider systematic review. This broader review explored predictors and moderators of treatment in a wider clinical population incorporating both systematic reviews and meta-analyses. The wider review was completed by Joanne Oberholzer and Buse Durdurak and supervised by Clarissa Bauer-Staeb and Katherine Button. The present overview focused specifically on meta-analyses examining moderators of treatment outcome in psychotherapy for depression and generalised anxiety.

9.1.1 Search strategy

Three databases were searched – PsychINFO, EMBASE, and Cochrane. The search strategy was developed by JO and BD with support from CBS and KB. The databases were searched on the 31st May 2019 with no limits or filters such as language or date restrictions. Hand searches were performed to identify additional articles including publications until 30th December 2020. The search strategy can be found below:

PsycINFO

(ti=(((Precision OR personli?ed OR individuali?ed) AND (medicine OR psychiatry OR treatment OR psychotherapy)) OR predict* OR moderat* OR it=(Prediction OR Precision medicine OR Client Treatment Matching) OR ab=(((Precision OR personali?ed OR individuali?ed) AND (medicine OR psychiatry OR treatment OR psychotherapy) OR predict* OR moderat*))) AND (ti=(depress* OR dysthymi* OR "mood disorder" OR "affective disorder" OR anxiety OR agoraphobi* OR panic OR obsessive compulsive OR phobia OR post traumatic stress OR trauma OR hypochondri* OR MDD OR GAD OR OCD OR PTSD OR MADD OR BDD) OR ab=(depress* OR dysthymi* OR "mood disorder" OR "affective disorder" OR anxiety OR agoraphobi* OR panic OR obsessive compulsive OR phobia OR post traumatic stress OR trauma OR hypochondri* OR MDD OR GAD OR OCD OR PTSD OR MADD OR BDD) OR MeSH=Depressive Disorder) AND (ti=(response OR outcome OR success OR recovery OR remission) OR it=(Psychotherapeutic Outcomes OR

Treatment Outcomes) OR MeSH=Treatment Outcome OR ab=(response OR outcome OR success OR recovery OR remission)) AND (ti=(systematic review OR review OR (systematic AND review) OR meta analysis) OR ab=(systematic review OR review OR (systematic AND review) OR meta analysis))

EMBASE

((precision:ab,ti OR personali?ed:ab,ti OR individuali?ed:ab,ti) AND (medicine:ab,ti OR psychiatry:ab,ti OR treatment:ab,ti OR psychotherapy:ab,ti) OR 'personalized medicine'/exp OR predict*:ab,ti OR moderat*:ab,ti OR 'predictor variable'/exp OR 'evidence based medicine'/exp OR 'prediction'/exp OR 'prediction and forecasting'/exp) AND (depress*:ab,ti OR 'depression'/exp OR dysthymi*:ab,ti OR 'dysthymia'/exp OR 'mood disorder':ab,ti OR 'affective disorder':ab,ti OR anxiety:ab,ti OR 'anxiety disorder'/mj OR agoraphobi*:ab,ti OR panic:ab,ti OR 'obsessive compulsive':ab,ti OR phobia:ab,ti OR 'post traumatic stress':ab,ti OR trauma:ab,ti OR hypochondri*:ab,ti OR 'hypochondriasis'/mj OR MDD:ab,ti OR GAD:ab,ti OR OCD:ab,ti OR PTSD:ab,ti OR MADD:ab,ti OR BDD:ab,ti) AND (response:ab,ti OR outcome:ab,ti OR success:ab,ti OR recovery:ab,ti OR remission:ab,ti OR 'clinical outcome'/mj OR 'treatment response'/mj) AND ('systematic review':ab,ti OR 'meta analysis':ab,ti)

COCHRANE

((precision:ab,ti OR personali?ed:ab,ti OR individuali?ed:ab,ti) AND (medicine:ab,ti OR psychiatry:ab,ti OR treatment:ab,ti OR psychotherapy:ab,ti) OR (personalized NEXT medicine) OR predict*:ab,ti OR moderat*:ab,ti OR (predictor NEXT variable) OR (evidence NEXT based NEXT medicine) OR [mh prediction] AND [mh forecasting])AND (depress*:ab,ti OR [mh depression] OR dysthymi*:ab,ti OR [mh dysthymia] OR mood disorder:ab,ti OR affective disorder:ab,ti OR [mh "anxiety disorder"] OR agoraphobi*:ab,ti OR panic:ab,ti OR obsessive compulsive:ab,ti OR phobia:ab,ti OR "post traumatic stress":ab,ti OR trauma:ab,ti OR "body dysmorphi*":ab,ti OR [mh "body dysmorphic disorder"] OR hypochondri*:ab,ti OR [mh hypochondriasis] OR MDD:ab,ti OR GAD:ab,ti OR OCD:ab,ti OR PTSD:ab,ti OR MADD:ab,ti OR BDD:ab,ti) AND (response:ab:ti OR outcome:ab,ti OR success:ab,ti OR recovery:ab,ti OR remission:ab,ti OR "clinical next outcome" OR [mh treatment response])

9.1.2 Screening

JO and BD completed the abstract screening in duplicate. JO and BD completed the full-text screening in duplicate at the 1st stage for the broader systematic review. Any discrepancies were resolved by discussion with CBS and KB. CBS completed the screening at the 2nd stage for the purposes of this PhD thesis.

9.1.3 Inclusion/Exclusion Criteria

Articles were included in the review at the 2nd stage if they met the following criteria:

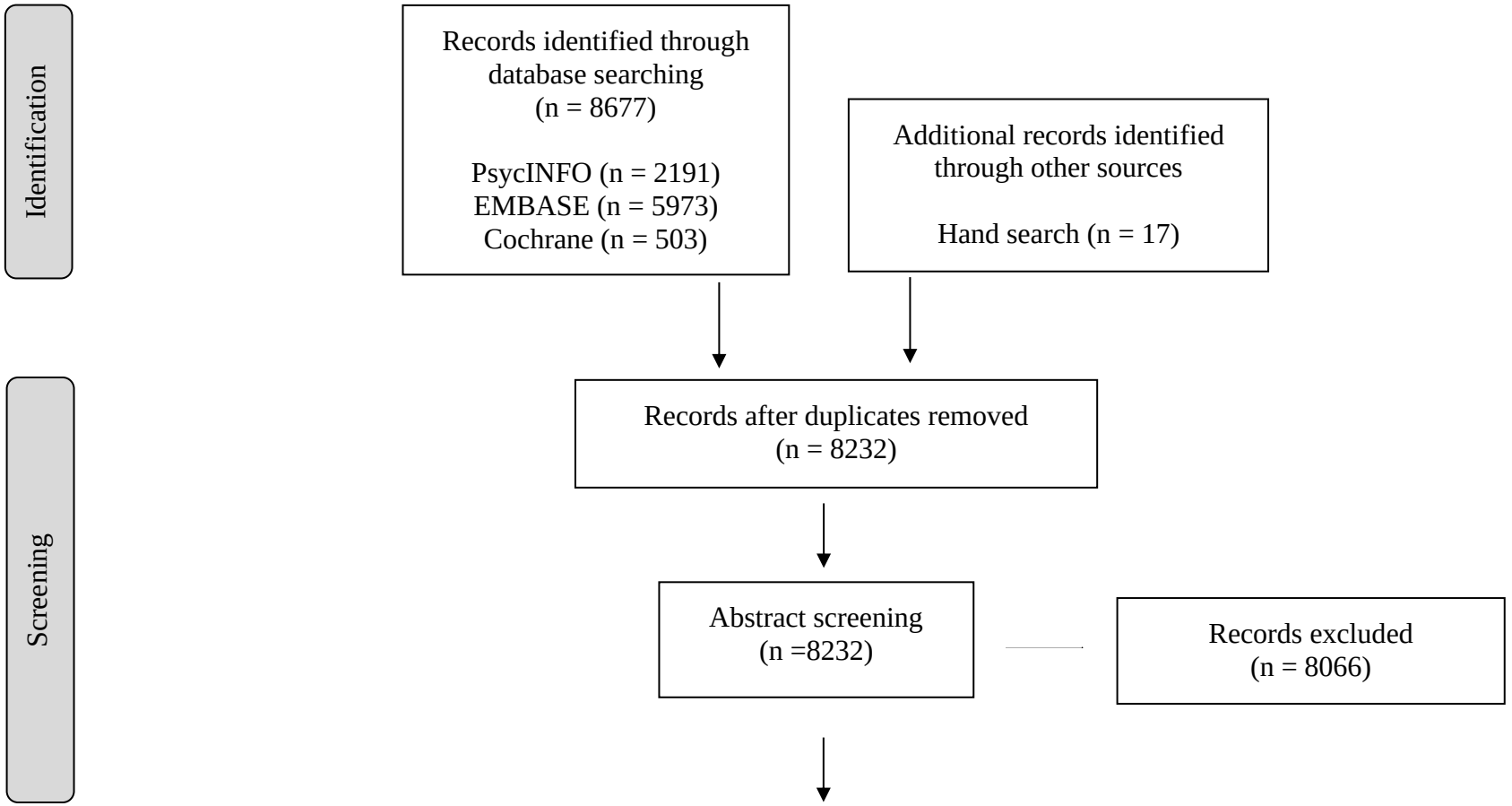
- Peer-reviewed journal articles published in English that demonstrated clear characteristics of a systematic review and meta-analysis (the development of a search strategy, a systematic search of the literature, title/abstract and full-text screening, data extraction, and quantitative synthesis of data).
- Meta-analyses that reported moderators of treatment outcomes either as a primary or secondary analysis if the examined characteristics were of a nature that can feasibly be assessed in routine clinical practice, including demographic, clinical, and/or social characteristics that were known/measured at baseline.
- Meta-analyses that included only adult populations (> 17 years of age).
- Meta-analyses where the population of interest was depression and/or generalised anxiety disorder. Meta-analyses that included other clinical populations were included only if moderator analyses were stratified by clinical presentation/diagnosis.
- The primary outcome measures were questionnaires assessing the overall severity of depression and/or anxiety.
- Meta-analyses that included peer reviewed RCTs. Meta-analyses that included observational studies were included only if the moderator analysis was stratified by study design.
- Meta-analyses that included an intervention arm focusing on psychotherapy, with at least one RCT that contained a psychotherapy delivered in IAPT. Adjunct pharmacotherapy was permitted.
- Meta-analyses that included RCTs where the comparator was an active or inactive control, pharmacotherapy, and/or psychotherapy.

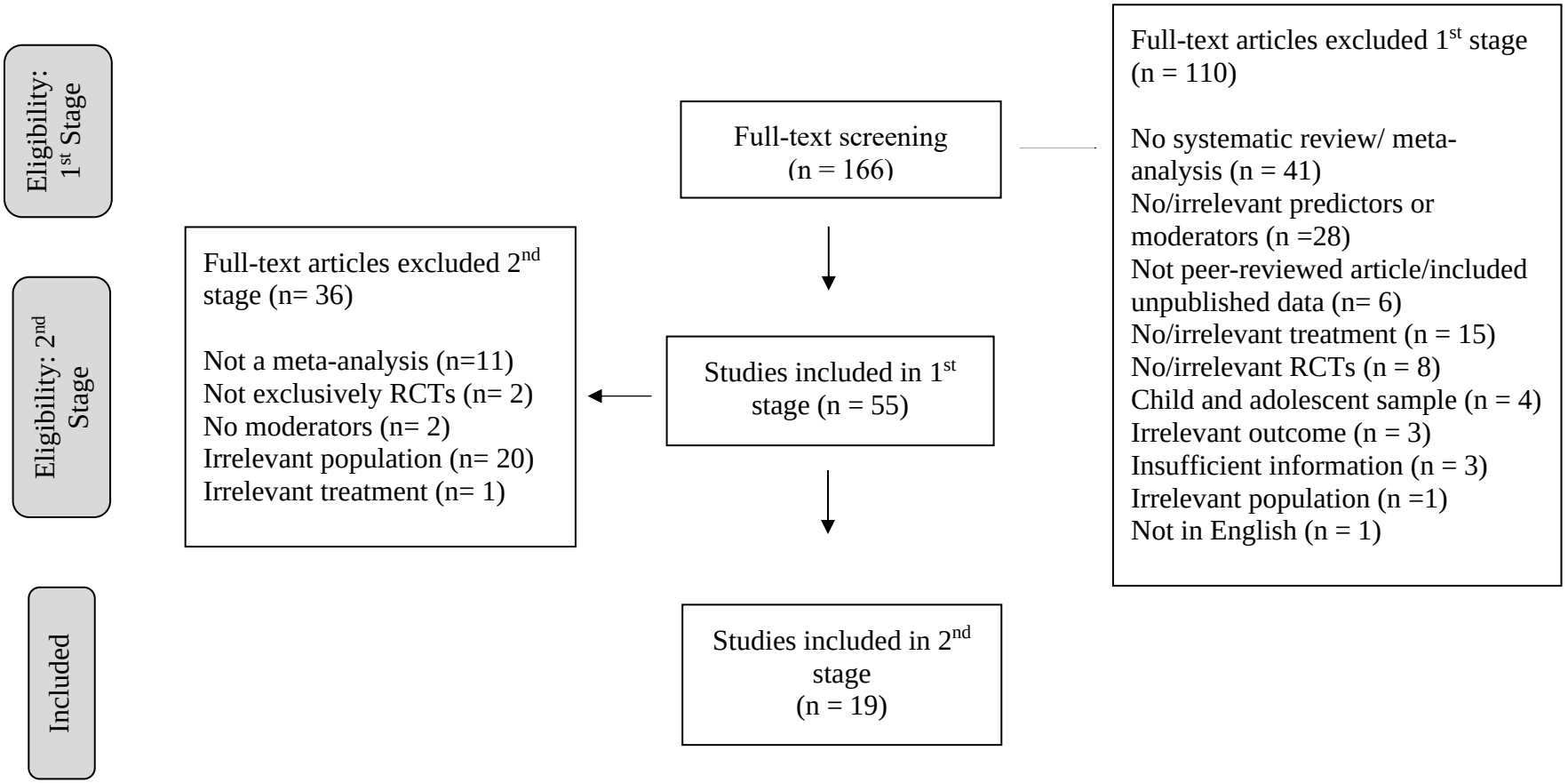
Articles were excluded from the narrative overview if they met the following criteria:

- The primary outcome measure assessed the severity of other symptoms than depression and anxiety, such as functional impairment.
- The population included children or adolescents
- The predictor/moderator analyses included only methodological or treatment-related characteristics, such as quality criteria or number of clinical appointments.

9.1.4 Prisma diagram

Supplementary figure A1. *PRIMSA flow diagram selection process for the narrative overview of meta-analyses*




Eligibility: 1st Stage
Full-text screening (n = 166)
Full-text articles excluded 1st stage (n = 110)
No systematic review/ meta-analysis (n = 41)
No/irrelevant predictors or moderators (n =28)
Not peer-reviewed article/included unpublished data (n= 6)
No/irrelevant treatment (n = 15)
No/irrelevant RCTs (n = 8)
Child and adolescent sample (n = 4)
Irrelevant outcome (n = 3)
Insufficient information (n = 3)
Irrelevant population (n =1)
Not in English (n = 1)
Eligibility: 2nd Stage
Studies included in 1st stage (n = 55)
Full-text articles excluded 2nd stage (n= 36)
Not a meta-analysis (n=11)
Not exclusively RCTs (n= 2)
No moderators (n= 2)
Irrelevant population (n= 20)
Irrelevant treatment (n= 1)
Included
Studies included in 2nd stage (n = 19)

9.1.5 Data extraction

JO and BD completed the data extraction for all included meta-analyses at the 1st stage. CBS reviewed and amended the data extraction for all included meta-analyses at the 2nd stage. The extracted data included:

- Authors
- Year
- Intervention
- Control
- Significance and direction of moderating effect
- Sample size/number of studies
- Commentary on publication bias
- Commentary on study quality

The significance and direction of moderating effects were extracted on the basis of the following criteria:

- Where multiple analyses were completed to address the same question, we present the results from the analysis deemed most appropriate/relevant by the authors to draw conclusions or, if unclear: multivariate regression analysis > meta-regression/subgroup analysis.
- Where multiple analyses were completed in different subgroups of the sample: intention-to-treat/full sample analysis > adjusted analysis > complete-case analysis.
- Where moderator analyses were stratified by intervention: psychological intervention > psychological + pharmacological.
- Where multiple outcome or patient characteristics were presented addressing the same question: continuous > categorical.
- Where multiple outcome measures are presented to address the same question: combined estimates > self-report measures > clinician report measures.
- Where multiple sample sizes are reported for moderator analyses: number of patients > number of trial arm comparisons > number of studies.
- Where power calculations were presented for predictors/moderators, only the results with sufficient power were reported.

9.2 Supplementary Material B: Chapter 4

Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials

9.2.1 Dealing with baseline dependency

Baseline dependency is a well-established problem – the change needed to feel better varies according to baseline severity, with patients who have more symptoms commonly requiring greater changes to experience a subjective improvement.[1-3] Various methods exist to address this problem. The most prominent methods include effect sizes, statistical control, proportionate change, and MCID categories.[1] Broadly speaking, mean change methods encompass approaches that examine the mean change at different levels of the GRC. That is, two of the originally developed methods include examining the mean change amongst those who *feel slightly better* or calculating the difference in mean change of those responding feeling *slightly better* and *feeling about the same*.[4,5] This method does not take baseline dependency into account. To account for baseline dependency, effect sizes can be estimated.[6,7] While effect sizes are useful for comparative purposes, they have been criticised for providing little clinical information and being difficult to interpret.[8] Statistical control can be implemented in models to reduce the effect of baseline severity.[1] However, Copay et al. reports that because extreme scores are assumed to be a result of error/chance, the true variation is masked despite the fact that medical patients might be expected to have higher symptom scores.[1] Proportionate change is the percentage of how much someone's symptoms change relative to their baseline score. While this approach is beneficial as it allows for comparisons between measures, it may increase the association with baseline scores when patients with high symptoms have small change.[1] A further approach is to provide multiple MCID estimates for categories of patients, which are grouped based on the certain levels of the measure.[1] These can sometimes create somewhat arbitrary groups and reduce the benefit of one MCID figure.[1] Previous analyses have shown that using proportionate scores in depression and anxiety provide a good rule of thumb for those with moderately-severe symptoms, but do not fully account for baseline dependency at all ranges and may additionally require categorization of the MCID based on baseline severity categories (mild, moderate, severe).[2,3] Our previous and present analyses additionally show that adjustment for baseline severity, either as an interaction term in linear

models or by using percent change, is insufficient to fully account for baseline severity across the scale.[2] Our argument is that if it is not possible to fully capture baseline dependency and produce one universal MCID with the methods above, it may be worthwhile having a more detailed and precise approach that, by definition, fully captures baseline severity to be used alongside the existing rule-of-thumb.

9.2.2 Modelling Approach

The MCID is a concept rather than being mathematically defined. As such, multiple methods have been proposed to estimate the MCID. Some of the most common methods include mean change, linear regression, and Receiver Operator Characteristics (ROC) curves.[1,7,9] A comprehensive review of MCID approaches can be found elsewhere.[1,7,9] As above, mean change methods don't account for baseline dependency and standardised effects have been criticised for providing little clinical information and being difficult to interpret.[4-9]
Using linear regression allows the mean changes to be adjusted for baseline severity to account for baseline dependency.[1,7,9,10] However,
previous research and the current analyses show that baseline adjustment is insufficient to fully account for baseline dependency in depression and anxiety across the spectrum.[2,3] ROC curves identify the point of optimal sensitivity and specificity between two groups of GRC responses (i.e., those who *feel better* vs. those who do *not feel better*) to denote the MCID.[1,2,9,10] However, they have the limitation that this estimate can be unstable and subject to changes in the ROC curve (and thus the relative balance of sensitivity/specificity) following the addition/deletion of a few data points.

We propose adopting the ED50 as an additional, novel means of estimating the MCID. The ED50 is frequently used in drug safety and pharmacotherapy research to
identify the minimum effective therapeutic dose.[11] An example of an application is the prescription of drugs, where the ED50 is used as a guideline for clinicians to identify the smallest effective dose of medication.[11] For the purposes of this analysis, the MCID is defined as the changes in scores associated with a 50% probability of *feeling better.* The ED50 has clear face validity as a MCID metric as it marks the threshold where patients are slightly more likely to feel *better* than *not*. The ED50 is based on a model derived from all data and therefore not as susceptible to the limitations of the ROC analysis and mean change methods. Using this approach, we further address the problem of baseline dependency above and beyond covariate adjustment in GLM. By incorporating baseline severity in

the GAMM model additively we provide an MCID for each level of baseline severity. Thus, addressing the limitation that covariate adjustment alone does not appear to fully account for baseline dependency in depression and anxiety. Using this method also allows for the identification of any important difference (i.e., ED25 and ED75) to examine the probabilities associated with different changes, which provides a granularity that other approaches do not. This provides flexibility to the end user to identify and select the amount of certainty in treatment response.

9.2.3 Data and statistical modelling decisions

9.2.3.1 Inclusion of all GRC responses

A common approach to estimating the MCID is to examine the mean change amongst those *who feel slightly* better or to examine the difference in mean change of those responding feeling *slightly better* and *feeling about the same* to find the minimal clinically important difference.[4,5] CoBalT patients were asked how they felt in comparison to the last assessment to which they could respond: "I feel better", "I feel about the same", and "I feel worse".[12] In PANDA, patients were asked how they felt compared to when they were last seen at all time points, with fixed responses entailing: "I feel a lot better", "I feel slightly better", "I feel about the same", "I feel slightly worse", and "I feel a lot worse".[13] As such, only one of the RCTs contained a more fine-grained breakdown and we were limited by the data available. This approach to estimating mean change is useful, but it has the limitation of throwing away a lot of data where only one or two of several fine-grained GRC categories are of interest, or when the categories are limited (as with CoBalT) the estimates can be inflated by inclusion of those people who felt very much better.

When estimating from statistical models (GLM or GAMM) it is preferrable to include all observations to reduce bias and increase the precision of the model. Examining only subgroups of patients can lead to erroneous results.[14] As such, we included all GRC responses in the present analysis. The present analysis nonetheless examines the minimal point as it is modelling the probability of feeling better by baseline severity and change in symptoms, rather than looking at the mean differences stratified by the GRC. From this model, we estimate the MCID as a threshold (a lower bound, if you will) of 50% chance or greater of feeling better. Unlike the categorical mean change approach this threshold is relatively robust to the inclusion of those with a wide range of GRC ratings.

9.2.3.2 Exclusion of time as a model effect

We found a statistically significant effect of time on the proportion of people feeling better at follow-up 1. However, we are interested in MCID estimates, and when we examined the effect of time on the ED50 (MCID estimates) we found marginal differences as a result of time, that were of little practical importance to the MCID estimates. We therefore excluded time from the final model for pragmatic purposes; future users will want to select a MCID without having to decide which follow-up period is closest to theirs. Unless there is strong evidence that time makes a practically significant difference to the MCID there is little benefit of adding time for the end-user (see Supplementary Table B3 and B4).

9.2.3.3 Pooling of studies and exclusion of study as a model term

We pooled data from two RCTs. This has the benefit of higher precision as there are more observations per level of baseline severity. We found a statistically significant effect of study on the probability of feeling better driven by differing baseline severities at follow-up 1, with PANDA having fewer observations at the very high end of scores, and CoBalT fewer lower scores.[10,11] As such, MCID estimates for the low end of scores from CoBalT at follow-up 1 alone will be unreliable. The opposite is true for PANDA. The effect of study disappeared after follow-up 1, further suggesting that the initial difference is a result of the different baseline characteristics of the RCTs and will therefore have little practical importance to the MCID estimates. Pooled data is preferable as it provides a greater coverage of baseline scores at time point 1 and the model produces a weighted average where most of the weight comes from one study. Similar to the covariate of time, the effects of study on the MCID estimates were of little practical importance (see Supplementary Table B3 and B4). We therefore exclude study from the final model. This is advantageous because future users will want to select an MCID for their study without having to decide whether it is more like one of the two studies – the results are more generalisable in this way.

9.2.3.4 Inclusion of treatment and control groups

We include both treatment and control groups in the present analyses for the purposes of generalisability. Our primary aim is to identify a change in scores that is noticeable to patients, but we are agnostic to how this change is produced. We assume a stability in the relationship between the changes in symptoms and the GRC that does not vary by treatment. We have no reason to believe that different treatments require a different

MCID, i.e., if a patient changes a given amount on the PHQ-9 they should be as likely to notice this difference regardless of whether it was brought about by SSRIs, CBT, or placebo/natural recovery. From our perspective (for the purposes of this analysis), the treatment is simply a means of inducing change.

9.2.3.5 No further covariate adjustment

While the adjustment of various other covariates is technically possible, we are unable to implement them in the current analyses. Firstly, there is a sample size consideration as we are stratifying by each level of baseline severity. This analysis would require much larger sample sizes for the adjustment of additional covariates, which would only be feasible through the analysis of electronic healthcare records or pooling of a very large number of clinical trials. Unfortunately, we are limited by the data available to us. However, there are also pragmatic issues with further covariate adjustment. In order to produce generic MCID estimates, the covariates would have to be fixed at certain points to estimate the ED50, introducing an array of assumptions that may not hold across all patients. Alternatively, an MCID can be estimated for each patient individually. However, this would firstly require that data to be available, creating the burden of additional data collection on patients and clinicians/researchers. This may be difficult in clinical practice due to time constraints but also in clinical research where these measures may otherwise not be of interest. Secondly, the calculation would be too extensive to print in any format and would require an online resource.

9.2.4 References

[1] Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *J Spine.* 2007;7(5):541-546. https://doi.org/10.1016/j.spinee.2007.01.008

[2] Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, Ades AE, Lewis G. Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychol Med.* 2015;45(15):3269-3279. http://doi.org/10.1017/S0033291715001270

[3] Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, Duffy L, Lanham P, Peters TJ, Wiles N, Lewis G. How Much Change is Enough? Evidence from a longitudinal study on depression in UK Primary Care. *Psychol Med*. 2020:52(10):1-8. https://doi.org/10.1017/S0033291720003700

[4] Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415. https://doi.org/10.1016/0197-2456(89)90005-6

[5] Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch Intern Med*. 1993;153(11):1337-1342.

[6] Borenstein M. Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC (Eds.). The Handbook of research synthesis and meta-analysis (2nd Ed.). New York: Russell Sage Foundation; 2009. p. 222-236.

[7] Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol.* 2017;82:128-136. http://doi.org/10.1016/j.jclinepi.2016.11.016

[8] Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *J Affect Disord*. 2014;159:118-126. https://doi.org/10.1016/j.jad.2014.02.026

[9] Mills KA, Naylor JM, Eyles JP, Roos EM, Hunter DJ. Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the knee injury and osteoarthritis outcome score. J Rheumatol. 2016;43(2):395-404. https://doi.org/10.3899/jrheum.150398

[10] Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum clinically important difference: current trends in the orthopaedic literature, part II: lower extremity: a systematic review. *JBJS Reviews*. 2018;6(9):e2. https://doi.org/10.2106/JBJS.RVW.17.00160

[11] Dimmitt S, Stampfer H, Martin JH. When less is more–efficacy with less toxicity at the ED50. *Br J Clin Pharmacol.* 2017;83: 1365– 1368. https://doi.org/10.1111/bcp.13281

[12] Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, Garland A, Hollinghurst S, Jerrom B, Kessler D, Kuyken W. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *Lancet*. 2013;381(9864):375-384. https://doi.org/10.1016/S0140-6736(12)61552-9

[13] Lewis G, Duffy L, Ades A, Amos R, Araya R, Brabyn S, Button KS, Churchill R, Derrick C, Dowrick C, Gilbody S. The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *Lancet Psychiatry.* 2019;6(11):903-914. https://doi.org/10.1016/S2215-0366(19)30366-9

[14] McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA*. 2014;312(13):1342-1343. https://doi.org/10.1001/jama.2014.13128

Supplementary Table B1. *Spearman rank correlation coefficients between change in symptoms and the categorical Global Rating of Change, stratified by study and follow-up*

		Baseline to Follow-up 1	**Follow-up 1 to Follow-up 2**	**Follow-up 2 to Follow-up 3**	**Follow-up 3 to Follow-up 4**
Patient Health	*PANDA*	-0.41	-0.50	-0.43	-
Questionnaire -9	*CoBalT*	-0.46	-0.43	-0.38	-0.32
Generalised Anxiety	*PANDA*	-0.37	-0.43	-0.36	-
Disorder Scale-7	*CoBalT**	-	-0.52	-	-0.41

**Generalised Anxiety Disorder Scale-7 data was not collected at follow-up one and three. Change scores are derived from previous follow-up.*

Data reported for patients with complete Global Rating of Change and each respective outcome questionnaires.

The Global Rating of Change for Panda contains 5 categories and 3 categories in CoBalT.

Supplementary Table B2. *Summary of the Generalised Additive Mixed Models*

Model	Observations	Adjusted r2	Deviance explained	UBRE	AIC	Variable(s)	P-value
Patient Health Questionnaire -9 (1)	3205	0.468	46.80	0.050	3366	Change x Baseline severity	<0.001
						ID	<0.001
Patient Health Questionnaire -9 (2)		0.468	46.60	0.045	3350	Change x Baseline severity	<0.001
						ID	<0.001
						Period 2	0.002
						Period 3	0.111
						Period 4	0.406
						Study: CoBalT	0.001
Generalised Anxiety Disorder Scale-7 (1)	2415	0.390	38.80	0.108	2676	Change x Baseline severity	<0.001
						ID	<0.001
Generalised Anxiety Disorder Scale-7 (2)		0.404	40.00	0.097	2649	Change x Baseline severity	<0.001
						ID	<0.001
						Period 2	<0.001
						Period 3	<0.001
						Study: CoBalT	0.042

**UBRE- Un-Biased Risk Estimator; AIC - Akaike Information Criterion; PHQ-9 - Patient Health Questionnaire - 9-item; GAD-7 - Generalised Anxiety Disorder 7-item*

Supplementary Table B3. *Minimal Clinically Important Differences stratified by study and follow-up period for the Patient Health Questionnaire 9-Item*

Panda	**Baseline Severity**																										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
T1	Inf	2	2	3	3	3	4	4	4	5	5	5	6	6	6	6	7	7	7	8	8	8	9	9	9	10	10
T2	0	0	0	0	0	0	1	2	3	4	4	5	5	5	5	5	5	4	4	5	5	5	6	7	7	7	8
T3	0	0	0	0	0	0	1	1	2	2	2	3	3	4	4	4	5	5	6	6	7	7	8	8	8	9	9
All	0	0	0	0	0	1	2	2	3	4	5	5	5	5	6	6	6	6	7	8	9	10	10	11	11	11	11
Excluding T1	0	0	0	0	0	0	1	2	2	3	4	4	4	5	5	5	5	5	5	5	6	6	7	8	8	9	9
CoBalT																											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
T1	0	0	0	0	0	0	0	0	0	0	1	2	2	3	4	4	5	6	6	7	8	8	9	10	10	11	12
T2	0	0	0	0	0	0	0	0	0	1	1	2	3	4	5	5	6	7	8	8	9	10	11	11	12	13	14
T3	0	0	0	0	0	0	0	0	1	2	2	3	4	4	5	6	6	7	8	8	9	10	11	11	12	13	13
T4	0	0	0	0	0	0	0	0	1	1	2	3	4	4	5	6	7	7	8	9	10	10	11	12	13	13	14
All	0	0	0	0	0	0	0	0	0	1	1	2	3	3	4	5	6	6	7	8	9	10	10	11	12	13	14
Excluding T1	0	0	0	0	0	0	0	0	0	1	2	3	3	4	5	6	6	7	8	9	9	10	11	12	13	13	14
Combined																											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
T1	Inf	2	2	2	2	2	2	3	3	3	4	4	4	5	5	6	6	7	7	8	8	9	9	10	10	10	11
T2	0	0	0	0	0	0	0	1	2	3	3	4	4	4	5	5	5	6	6	7	8	9	10	11	11	12	13
T3	0	0	0	0	0	0	0	1	1	2	2	3	3	4	4	5	6	6	7	8	8	9	10	11	12	12	13
T4*	0	0	0	0	0	0	0	0	1	1	2	3	4	4	5	6	7	7	8	9	10	10	11	12	13	13	14
All	0	0	0	0	0	0	1	1	2	3	3	4	4	5	5	5	6	7	7	8	9	10	11	11	12	13	14
Excluding T1	0	0	0	0	0	0	0	1	2	2	3	3	4	4	5	5	6	6	7	8	9	10	11	12	13	13	14

**Estimates from CoBalT only*

Supplementary Table B4. *Minimal Clinically Important Differences stratified by study and follow-up period for the Generalised Anxiety Disorder Scale 7-Items*

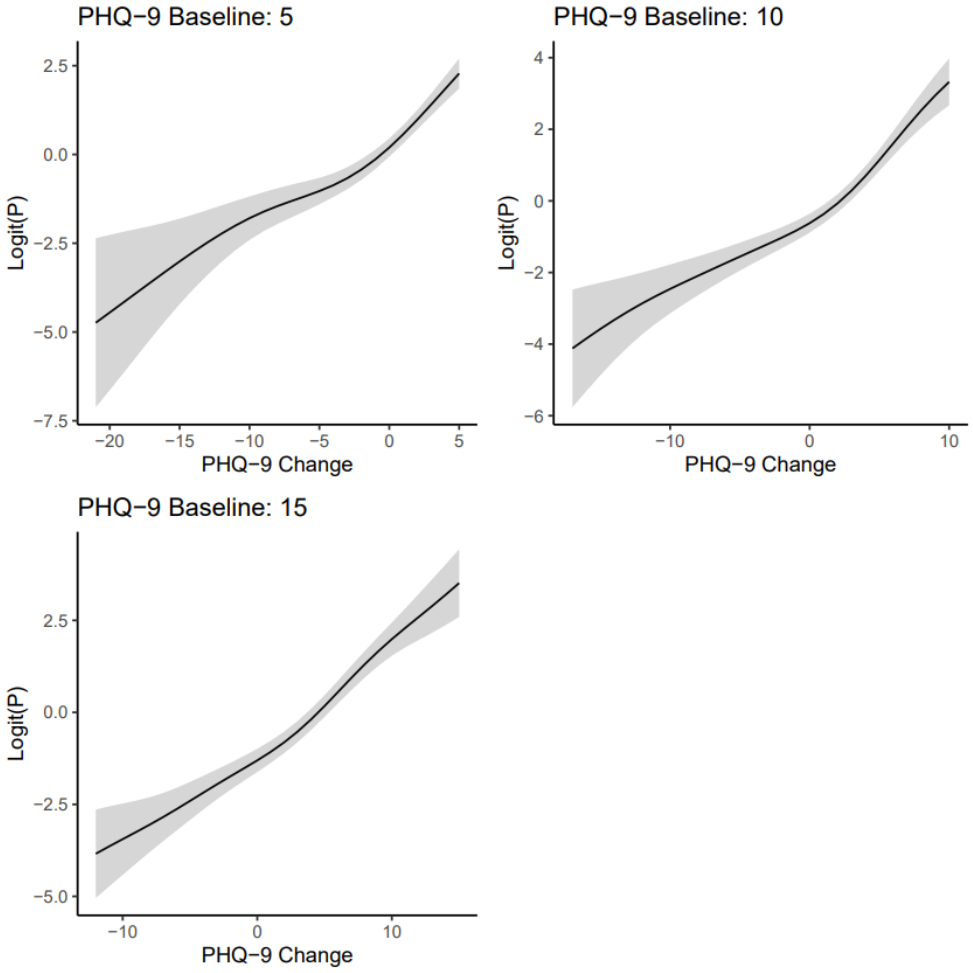
Panda	**Baseline Severity**																				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Time 1	Inf	Inf	3	3	3	3	4	4	5	5	6	6	7	7	7	8	9	10	11	12	12
Time 2	0	0	0	0	0	1	2	3	4	4	5	5	5	5	5	5	6	6	6	7	7
Time 3	0	0	0	0	1	1	1	2	2	3	3	3	4	4	5	5	5	6	6	7	7
All times	0	0	0	1	1	2	2	3	4	5	5	6	6	6	7	7	8	9	10	11	12
Excluding Time 1	0	0	0	0	0	1	2	2	3	4	4	5	5	5	6	6	6	7	7	7	8
CoBalT																					
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Time 1	0	0	0	0	1	2	2	3	3	3	4	5	7	8	8	9	9	9	9	9	9
Time 2	0	0	0	0	0	1	2	2	3	3	4	5	5	6	6	7	7	8	9	9	10
Time 3	0	0	0	0	1	1	2	3	3	4	4	5	6	7	7	8	8	8	9	9	9
All times	0	0	0	0	0	1	2	2	3	3	4	5	5	6	6	7	7	8	9	9	10
Combined																					
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Time 1	Inf	2	2	2	2	3	3	3	4	5	5	6	7	7	8	8	9	9	9	10	10
Time 2	0	0	0	0	0	1	2	3	3	4	4	5	5	5	6	6	6	7	7	8	9
Time 3*	0	0	0	0	1	1	1	2	2	3	3	3	4	4	5	5	5	6	6	7	7
All times	0	0	0	0	1	2	2	3	4	4	5	5	6	6	7	7	8	8	9	10	10
Excluding T1	0	0	0	0	0	1	2	2	3	4	4	5	5	5	6	6	7	7	8	8	9

**Estimates from Panda only*

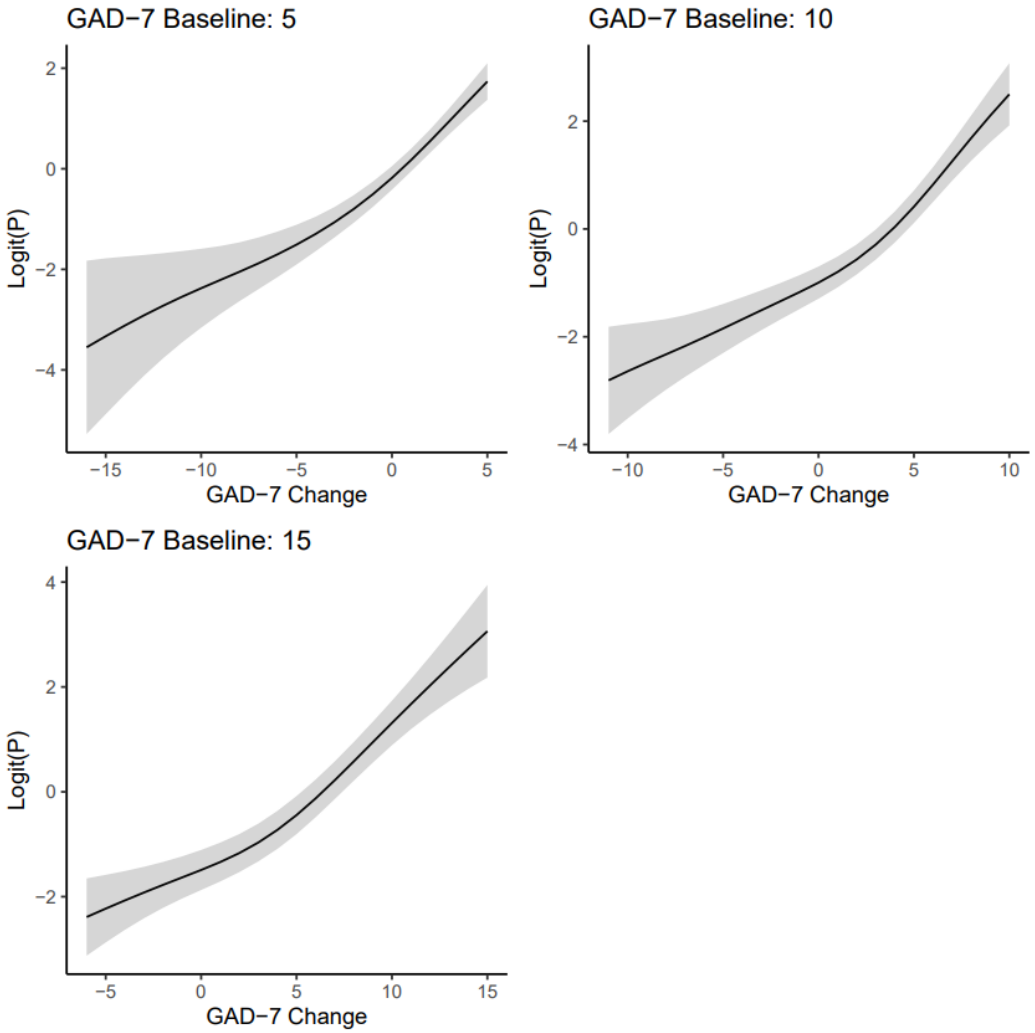
The following graphs present slices through smooth surface plots at the mild, moderate, and severe thresholds for each outcome questionnaire. The predicted values are presented on the logit scale to assess variability at each level of change, with corresponding 95% confidence intervals. Limits were set to the maximum obtainable change for each level of baseline severity.

The confidence intervals widen slightly towards the extreme ends of change, particularly when baseline severity is low and extreme reduction (deteriorations) take place as it was rare for patients with such mild symptoms to deteriorate drastically. This is unlikely to have an impact on the ED50 estimates, as they are focused on positive changes above 0, where the confidence intervals are visibly narrower.

Supplementary Figure B1. *Slices of the smoothed surface plots with 95% confidence intervals for the Patient Health Questionnaire -9*



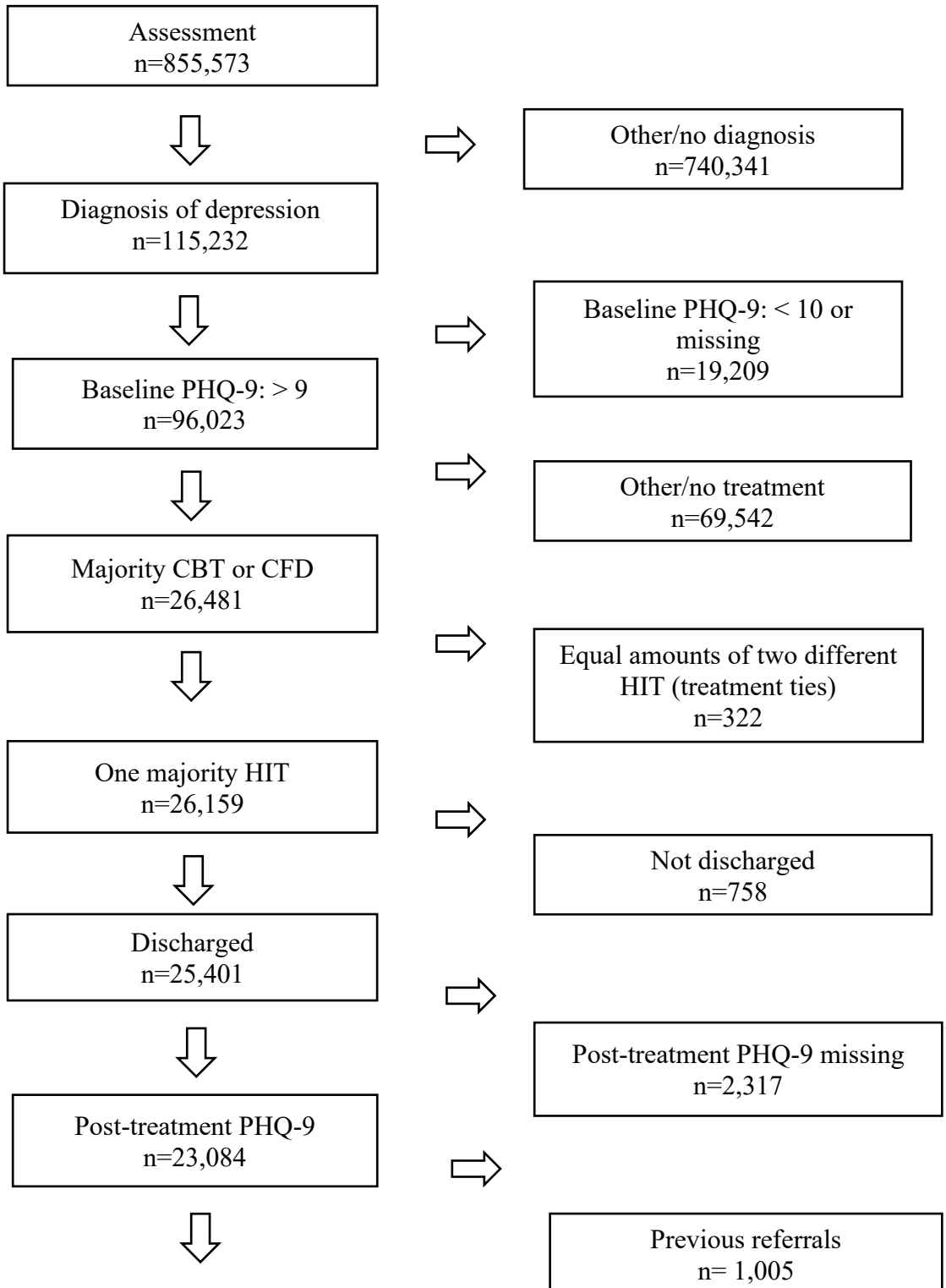
Supplementary Figure B2. *Slices of the smoothed surface plots with 95% confidence intervals for the Generalised Anxiety Disorder Scale-7*

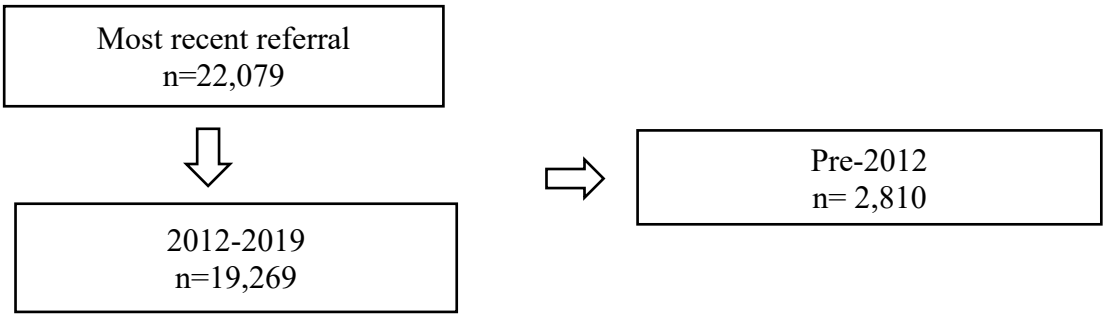


9.3 Supplementary Material C: Chapter 5

Personalised Psychotherapy in Primary Care: An Evaluation of Data-Driven Treatment Allocation to Cognitive Behavioural Therapy vs. Counselling for Depression

Supplementary Figure C1. *Flowchart of referral selection*





**PHQ-9: Patient Health Questionnaire (9-items); HIT: High-Intensity Therapy*

Supplementary Table C1. *Summary of data composition for baseline and outcome variables*

Variable	Data Type	Levels
Age	Continuous	-
Gender	Categorical	*Female*
		Male
Ethnicity	Categorical	*White*
		White
		BAME
		Mixed
		Asian or Asian British
		Black or Black British
		Other Ethnic Group
Employment Status	Categorical	*Employed*
		Not working
		Unemployed and Seeking Work
		Students who are undertaking full (at least 16 hours per week) or part-time (less than 16 hours per week) education or training and who are not working or actively seeking work

		Long-term sick or disabled, those who are receiving Incapacity Benefit, Income Support or both; or Employment and Support Allowance
		Homemaker looking after the family or home and who are not working or actively seeking work
		Not receiving benefits and who are not working or actively seeking work
		Unpaid voluntary work who are not working or actively seeking work
		Retired
Index of Multiple Deprivation	Continuous	-
Sexual Orientation	Categorical	*Heterosexual*
		Not heterosexual
		Gay/Lesbian
		Bi-sexual
Disability	Categorical	*No*
		Yes
Long-Term Health Condition	Categorical	*No*
		Yes
Diagnosis	Categorical	*Depressive episode*
		Recurrent depressive disorder

Baseline PHQ-9	Continuous	-
Baseline GAD-7	Continuous	-
Baseline WSAS	Continuous	-
Psychotropic Medication	Categorical	*Yes*
		Prescribed and taking
		No
		Prescribed but not taking
		Not Prescribed
Referral Source	Categorical	*Self*
		Self
		Carer
		Primary Care
		General Medical Practitioner
		Health Visitor
		Other Primary Health Care
		Other
		Local Authority Services
		Employer
		Justice System
		Child Health
		Independent/Voluntary Sector

		Acute Secondary Care
		Other Mental Health NHS Trust
		Internal referrals from Community Mental Health Team (within own NHS Trust)
		Internal referrals from Inpatient Service (within own NHS Trust)
		Transfer by graduation (within own NHS Trust)
		Other
		IAPT
Referral Number	Continuous	-
Low-Intensity Therapy	Categorical	*No*
		0 low-intensity appointments
		Yes
		> 0 low-intensity appointments
Post-Treatment PHQ-9	Continuous	-

**PHQ-9: Patient Health Questionnaire (9-item); GAD-7: Generalised Anxiety Disorder Scale (7-item); WSAS: Work and Social Adjustment Scale.*

Supplementary Table C2. *Baseline characteristics stratified by treatment and data split*

		Training Data				*Testing Data*				*Training & Testing Data*
		Cognitive Behavioural Therapy	**Counselling for Depression**	**Standardised Mean Difference**	**% Missing**	**Cognitive Behavioural Therapy**	**Counselling for Depression**	**Standardised Mean Difference**	**% Missing**	**Standardised Mean Difference**
n		*10,908*	*3544*			*3636*	*1181*			
Age		39.52 (14.02)	42.17 (13.68)	0.192	0.0	39.79 (14.00)	42.71 (13.84)	0.210	0.0	0.024
Gender	Female	7118 (65.3)	2502 (70.6)	0.115	0.1	2361 (64.9)	856 (72.5)	0.163	0.1	0.005
	Male	3790 (34.7)	1042 (29.4)			1275 (35.1)	325 (27.5)			
Ethnicity	White	8977 (82.3)	2549 (71.9)	0.249	4.3	3006 (82.7)	869 (73.6)	0.221	4.7	0.017
	BAME	1931 (17.7)	995 (28.1)			630 (17.3)	312 (26.4)			
Employment Status	Employed	5827 (53.4)	2105 (59.4)	0.121	0.8	1939 (53.3)	699 (59.2)	0.118	0.9	0.002
	Not working	5081 (46.6)	1439 (40.6)			1697 (46.7)	482 (40.8)			
Index of Multiple Deprivation		21.60 (12.10)	21.29 (11.01)	0.027	0.4	21.40 (12.16)	21.07 (10.94)	0.028	0.5	0.017
Sexual Orientation	Heterosexual	10,350 (94.9)	3422 (96.6)	0.083	16.5	3466 (95.3)	1128 (95.5)	0.009	15.4	0.004
	Not heterosexual	558 (5.1)	122 (3.4)			170 (4.7)	53 (4.5)			
Disability	No	9328 (85.5)	3027 (85.4)	0.003	8.3	3087 (84.9)	1007 (85.3)	0.010	8.6	0.014
	Yes	1580 (14.5)	517 (14.6)			549 (15.1)	174 (14.7)			
Long-Term Health Condition	No	7080 (64.9)	2395 (67.6)	0.057	8.3	2368 (65.1)	759 (64.3)	0.018	8.2	0.014
	Yes	3828 (35.1)	1149 (32.4)			1268 (34.9)	422 (35.7)			
Diagnosis	Depressive episode	7708 (70.7)	3035 (85.6)	0.368	0.0	2580 (71.0)	1002 (84.8)	0.339	0.0	0.001

	Recurrent depressive disorder	3200 (29.3)	509 (14.4)			1056 (29.0)	179 (15.2)			
Baseline PHQ-9		18.53 (4.34)	17.39 (4.39)	0.262	0.0	18.47 (4.40)	17.45 (4.38)	0.232	0.0	0.006
Baseline GAD-7		14.51 (4.48)	13.84 (4.53)	0.150	0.1	14.53 (4.50)	13.68 (4.71)	0.183	0.1	0.006
Baseline WSAS		23.78 (8.38)	20.79 (8.77)	0.349	0.1	23.73 (8.38)	21.42 (8.74)	0.270	1.7	0.014
Psychotropic Medication	Yes	6655 (61.0)	1704 (48.1)	0.262	2.9	2224 (61.2)	529 (44.8)	0.333	2.9	0.014
	No	4253 (39.0)	1840 (51.9)			1412 (38.8)	652 (55.2)			
Referral Source	Self	6427 (58.9)	1545 (43.6)	0.432	0.0	2180 (60.0)	517 (43.8)	0.433	0.0	0.018
	Primary care	3654 (33.5)	1895 (53.5)			1196 (32.9)	628 (53.2)			
	Other	827 (7.6)	104 (2.9)			260 (7.2)	36 (3.0)			
Referral Number		1.86 (1.31)	1.57 (1.00)	0.250	0.0	1.85 (1.29)	1.55 (0.99)	0.259	0.0	0.014
Low-Intensity Therapy	No	7604 (69.7)	2869 (81.0)	0.263	0.0	2565 (70.5)	969 (82.0)	0.273	0.0	0.020
	Yes	3304 (30.3)	675 (19.0)			1071 (29.5)	212 (18.0)			
Service	A	2803 (25.7)	292 (8.2)	1.321	0.0	938 (25.8)	126 (10.7)	1.210	0.0	0.059
	B	1522 (14.0)	7 (0.2)			515 (14.2)	4 (0.3)			
	C	904 (8.3)	805 (22.7)			295 (8.1)	252 (21.3)			
	D	1336 (12.2)	56 (1.6)			446 (12.3)	26 (2.2)			
	E	666 (6.1)	117 (3.3)			208 (5.7)	27 (2.3)			
	F	613 (5.6)	806 (22.7)			193 (5.3)	248 (21.0)			
	G	491 (4.5)	58 (1.6)			170 (4.7)	28 (2.4)			
	H	602 (5.5)	97 (2.7)			182 (5.0)	30 (2.5)			
	I	349 (3.2)	427 (12.0)			135 (3.7)	113 (9.6)			
	J	74 (0.7)	35 (1.0)			23 (0.6)	18 (1.5)			
	K	678 (6.2)	280 (7.9)			225 (6.2)	120 (10.2)			
	L	118 (1.1)	27 (0.8)			47 (1.3)	8 (0.7)			
	M	115 (1.1)	10 (0.3)			48 (1.3)	4 (0.3)			

	N	108 (1.0)	51 (1.4)			32 (0.9)	18 (1.5)			
	O	529 (4.8)	476 (13.4)			179 (4.9)	159 (13.5)			
Year	2012	989 (9.1)	169 (4.8)	0.280	0.0	335 (9.2)	47 (4.0)	0.304	0.0	0.031
	2013	1415 (13.0)	342 (9.7)			486 (13.4)	117 (9.9)			
	2014	1738 (15.9)	781 (22.0)			541 (14.9)	270 (22.9)			
	2015	2264 (20.8)	872 (24.6)			800 (22.0)	272 (23.0)			
	2016	1552 (14.2)	514 (14.5)			503 (13.8)	170 (14.4)			
	2017	1602 (14.7)	428 (12.1)			524 (14.4)	156 (13.2)			
	2018	1126 (10.3)	317 (8.9)			378 (10.4)	117 (9.9)			
	2019	222 (2.0)	121 (3.4)			69 (1.9)	32 (2.7)			

**PHQ-9: Patient Health Questionnaire (9-item); GAD-7: Generalised Anxiety Disorder Scale (7-item); WSAS: Work and Social Adjustment Scale. Continuous data are presented as mean (standard deviation) and categorical data are presented as n (%).*

Supplementary Table C3. Predictors of post-treatment PHQ-9 score in Cognitive Behavioural Therapy and Counselling for Depression

	Beta	**95% Confidence Interval**		**p-value**
Age	-0.04	-0.05	-0.03	<0.001
Gender (Male)	0.11	-0.15	0.36	0.400
Ethnicity (BAME)	0.34	-0.02	0.69	0.064
Employment Status (Not working)	1.91	1.66	2.17	<0.001
Index of Multiple Deprivation	0.03	0.01	0.04	<0.001
Sexual Orientation (Not heterosexual)	0.51	-0.04	1.05	0.071
Disability (Yes)	0.68	0.31	1.05	<0.001
Long-Term Health Condition (Yes)	0.56	0.28	0.84	<0.001
Diagnosis (Recurrent depressive disorder)	0.30	0.00	0.60	0.051
Baseline PHQ-9	0.37	0.33	0.40	<0.001
Baseline GAD-7	0.13	0.10	0.16	<0.001
Baseline WSAS	0.07	0.06	0.09	<0.001
Psychotropic Medication (No)	-0.46	-0.73	-0.20	<0.001
Referral Source (ref: Self)				<0.001
Primary Care	0.66	0.35	0.96	0.000
Other	0.97	0.50	1.44	0.000
Referral Number	0.31	0.21	0.41	<0.001

Low-Intensity Therapy (No)	0.18	-0.13	0.50	0.258
Service (ref: A)				<0.001
B	0.50	0.03	0.98	
C	1.33	0.56	2.10	
D	-0.38	-0.80	0.04	
E	-0.80	-1.34	-0.26	
F	0.79	-0.15	1.73	
G	0.50	-0.13	1.14	
H	-0.71	-1.29	-0.14	
I	0.65	-0.39	1.70	
J	-0.39	-1.71	0.93	
K	0.45	-0.16	1.05	
L	-1.06	-2.14	0.02	
M	1.06	-0.08	2.20	
N	0.47	-0.71	1.64	
O	-0.08	-0.91	0.74	
Year (ref: 2015)				0.016
2012	-0.10	-0.59	0.39	
2013	-0.48	-0.90	-0.05	
2014	0.13	-0.21	0.46	
2016	-0.28	-0.64	0.08	

2017	-0.55	-0.93	-0.18	
2018	-0.14	-0.56	0.27	
2019	0.01	-0.71	0.72	
Propensity Score	0.45	-1.65	2.54	0.676

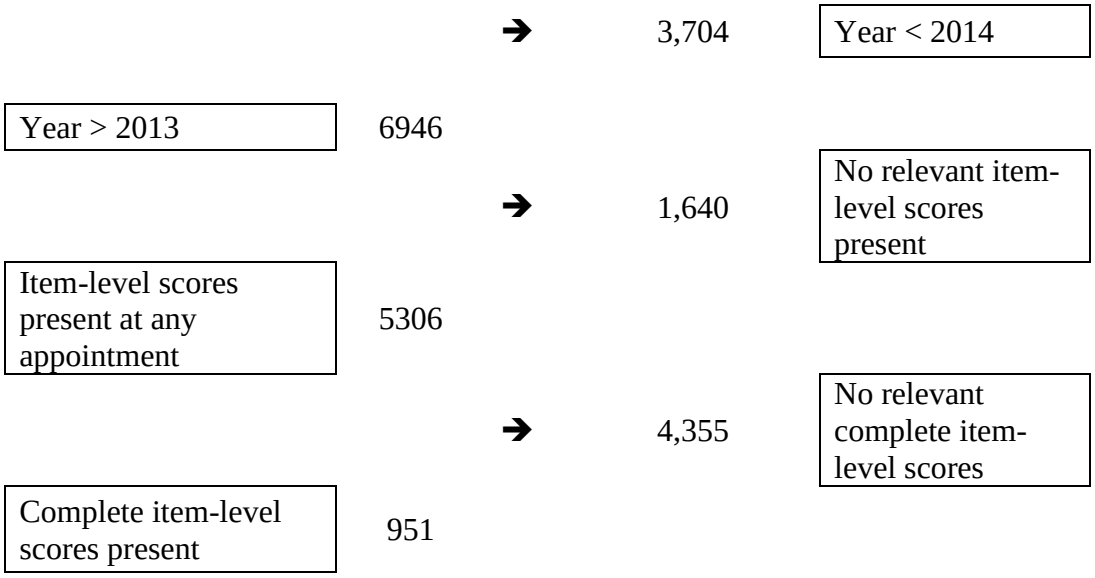
**PHQ-9: Patient Health Questionnaire (9-item); GAD-7: Generalised Anxiety Disorder Scale (7-item); WSAS: Work and Social Adjustment Scale.*

9.4 Supplementary Material D: Chapter 6

Trajectories of depression and generalised anxiety symptoms over the course of Cognitive Behavioural Therapy in primary care: An observational, retrospective cohort

Supplementary table D1. *Flowchart of sample selection*

	Included		**Excluded**	
	n		***n***	
Referrals with one appointment scheduled	666,632			
		➔	31,698	Not discharged
Discharged referrals	634,934			
		➔	71,595	No attended appointments
Referrals with at least one attendance	563,339			
		➔	478,588	Other/combination treatments
Cognitive behavioural therapy only	84,751			
		➔	41,630	< 8 appointments
Appointments: > 7	43,121			
		➔	32,045	No diagnosis present
Diagnosis present	11,076			
		➔	1	Other Anxiety Disorder Specific Measure
Anxiety Disorder Specific Measure: Generalised Anxiety Disorder-7 Scale	11,075			
		➔	425	Later Referrals
First referral	10,650			


➔
3,704
Year < 2014
Year > 2013
6946
➔
1,640
No relevant item-level scores present
Item-level scores present at any appointment
5306
➔
4,355
No relevant complete item-level scores
Complete item-level scores present
951

Supplementary Material D2. *Mean item-level questionnaire scores stratified by appointment and questionnaire*

Patient Health Questionnaire-9

		Question 1		***Question 2***		***Question 3***		***Question 4***		***Question 5***		***Question 6***		***Question 7***		***Question 8***		***Question 9***	
Appointment	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1	2980	2.15	0.88	2.32	0.79	2.32	0.91	2.44	0.77	1.95	1.06	2.38	0.84	2.01	0.98	1.28	1.08	1.01	1.03
2	1323	1.94	0.92	2.08	0.88	2.24	0.94	2.35	0.82	1.94	1.03	2.18	0.91	1.86	0.99	1.19	1.08	0.90	1.01
3	1364	1.81	0.91	1.92	0.91	2.12	0.97	2.25	0.85	1.80	1.06	2.02	0.95	1.74	1.00	1.08	1.03	0.75	0.95
4	1398	1.72	0.91	1.88	0.91	2.03	0.98	2.17	0.89	1.73	1.05	1.89	0.97	1.66	0.98	1.02	1.01	0.71	0.96
5	1436	1.61	0.89	1.76	0.90	1.96	0.97	2.03	0.91	1.66	1.05	1.80	0.95	1.58	0.98	0.97	1.00	0.67	0.92
6	1448	1.55	0.90	1.70	0.91	1.91	0.99	1.98	0.92	1.58	1.04	1.72	0.98	1.49	1.01	0.91	0.99	0.61	0.89
7	1439	1.45	0.91	1.63	0.93	1.84	1.03	1.90	0.96	1.49	1.06	1.59	0.97	1.42	1.00	0.87	0.98	0.58	0.88
8	1461	1.40	0.91	1.56	0.94	1.76	1.03	1.86	0.98	1.43	1.07	1.50	1.00	1.34	1.01	0.80	0.95	0.55	0.87

Generalised Anxiety Disorder-7 Scale

		Question 1		***Question 2***		***Question 3***		***Question 4***		***Question 5***		***Question 6***		***Question 7***	
Appointment	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1	1634	2.56	0.68	2.49	0.75	2.54	0.73	2.26	0.86	1.38	1.08	1.93	0.97	2.02	1.04
2	772	2.32	0.82	2.27	0.89	2.35	0.84	2.06	0.91	1.27	1.08	1.88	0.95	1.84	1.05
3	799	2.14	0.86	2.09	0.92	2.18	0.89	1.90	0.96	1.18	1.05	1.74	0.98	1.65	1.09
4	838	2.05	0.88	2.00	0.92	2.06	0.91	1.77	0.98	1.10	1.02	1.63	0.98	1.55	1.08
5	849	2.01	0.87	1.91	0.95	1.98	0.92	1.72	0.96	1.05	1.02	1.57	0.97	1.47	1.04
6	837	1.87	0.88	1.77	0.95	1.82	0.95	1.58	0.97	0.99	0.97	1.48	0.97	1.33	1.00
7	838	1.74	0.90	1.62	0.97	1.71	0.95	1.49	0.99	0.89	0.97	1.40	0.96	1.30	1.02
8	810	1.66	0.89	1.52	1.00	1.57	0.99	1.40	1.00	0.86	0.96	1.30	0.96	1.21	1.02

**SD – standard deviation*

Supplementary Table D3. *Covariates for the model predicting the probability of having no depressive symptoms amongst patients with at least one full set of Patient Health Questionnaire-9 scores across eight appointments*

	Odds Ratio	**95 % Confidence Intervals**		**p-value**
Question				
1	0.38	0.32	0.44	< 0.001
2	0.17	0.14	0.21	< 0.001
3	0.64	0.56	0.75	< 0.001
4	0.15	0.13	0.19	< 0.001
5	1.84	1.63	2.07	< 0.001
6	0.33	0.28	0.39	< 0.001
7	1.13	0.99	1.28	0.064
8	10.95	9.92	12.08	< 0.001
9	20.58	17.65	24.00	-
Appointment	1.25	1.24	1.27	< 0.001
Male (ref: female)	0.86	0.76	0.96	0.009
Age	1.01	1.01	1.02	< 0.001
Black, Asian, and ethnic minority (ref: White)	0.96	0.80	1.15	0.659
Not working (ref: employed)	0.91	0.81	1.02	0.111
Index of Multiple Deprivation	1.00	0.99	1.00	0.049
Disability (ref: none)	0.99	0.83	1.17	0.901

Long-term Health condition (ref: none)	0.95	0.83	1.07	0.386
Recurrent depressive disorder (ref: Depressive episode)	0.95	0.84	1.07	0.376
Baseline Patient Health Questionnaire-9	0.82	0.81	0.84	0.000
Baseline Generalised Anxiety Disorder Scale-7	0.99	0.97	1.00	0.094
Baseline Work and Social Adjustment Scale	0.98	0.98	0.99	< 0.001
No medication (ref: taking medication)	1.18	1.05	1.33	0.005
Referral number	0.94	0.91	0.98	0.006
Referral Source (ref: Self)				
Primary Care	0.88	0.76	1.00	0.058
Other	0.86	0.68	1.09	0.216
Service (ref: A)				
B	1.17	0.99	1.39	0.064
C	1.38	1.06	1.79	0.017
D	1.17	0.88	1.57	0.277
E	1.54	1.18	2.00	0.001
F	1.09	0.78	1.52	0.614
G	1.25	0.95	1.65	0.108
H	1.48	1.16	1.89	0.002
I	1.55	1.23	1.95	< 0.001
J	1.13	0.89	1.43	0.314

Year (ref: 2014)				
2015	1.01	0.84	1.22	0.885
2016	0.99	0.81	1.22	0.958
2017	0.94	0.77	1.14	0.524
2018	0.88	0.72	1.07	0.202
2019	1.09	0.74	1.60	0.677
Question 1 x Appointment	1.04	1.01	1.07	0.014
Question 2 x Appointment	1.07	1.04	1.11	< 0.001
Question 3 x Appointment	0.94	0.91	0.96	< 0.001
Question 4 x Appointment	1.04	1.00	1.08	0.027
Question 5 x Appointment	0.93	0.91	0.95	< 0.001
Question 6 x Appointment	1.06	1.03	1.09	< 0.001
Question 7 x Appointment	0.99	0.97	1.01	0.335
Question 8 x Appointment	0.94	0.92	0.96	< 0.001
Question 9 x Appointment	1.00	0.98	1.03	-

**P-values for the reference group of the sum coding cannot be estimated.*

Supplementary Table D4. *Covariates for the model predicting the probability of having no generalised anxiety symptoms amongst patients with at least one full set of Generalised Anxiety Disorder-7 scores across eight appointments*

	Odds Ratio	**95% Confidence Intervals**		**p-value**
Question				
1	0.11	0.08	0.16	< 0.001
2	0.31	0.24	0.40	< 0.001
3	0.22	0.17	0.30	< 0.001
4	0.83	0.67	1.03	0.085
5	19.01	16.25	22.25	< 0.001
6	2.01	1.67	2.42	< 0.001
7	4.00	3.12	5.11	-
Appointment	1.35	1.32	1.38	< 0.001
Male (ref: female)	0.96	0.78	1.18	0.704
Age	1.01	1.00	1.02	0.002
Black, Asian, and ethnic minority (ref: White)	1.00	0.71	1.41	0.983
Not working (ref: employed)	0.79	0.65	0.97	0.027
Index of Multiple Deprivation	1.00	0.99	1.01	0.985
Disability (ref: none)	0.82	0.57	1.19	0.294
Long-term Health condition (ref: none)	0.89	0.71	1.11	0.306

Baseline Patient Health Questionnaire-9	0.95	0.93	0.97	< 0.001
Baseline Generalised Anxiety Disorder Scale-7	0.80	0.77	0.82	< 0.001
Baseline Work and Social Adjustment Scale	0.99	0.98	1.00	0.153
No medication (ref: taking medication)	1.10	0.91	1.33	0.323
Referral number	0.93	0.87	1.00	0.047
Referral Source (ref: Self)				
Other	0.76	0.47	1.23	0.266
Primary Care	0.95	0.76	1.19	0.646
Service (ref: A)				
B	0.72	0.54	0.98	0.034
C	0.85	0.49	1.47	0.568
D	0.92	0.57	1.49	0.747
E	0.97	0.66	1.43	0.877
F	0.91	0.48	1.72	0.772
G	1.30	0.92	1.85	0.139
H	0.88	0.60	1.29	0.516
I	1.07	0.75	1.54	0.694
J	0.80	0.50	1.29	0.363
Year (ref: 2014)				
2015	1.04	0.75	1.45	0.817
2016	1.22	0.86	1.73	0.259

2017	1.13	0.80	1.59	0.488
2018	1.12	0.79	1.58	0.530
2019	1.83	0.95	3.54	0.071
Question 1 x Appointment	1.07	1.00	1.13	0.044
Question 2 x Appointment	1.09	1.04	1.14	< 0.001
Question 3 x Appointment	1.09	1.04	1.15	< 0.001
Question 4 x Appointment	1.03	0.99	1.07	0.169
Question 5 x Appointment	0.89	0.87	0.92	< 0.001
Question 6 x Appointment	0.93	0.90	0.96	< 0.001
Question 7 x Appointment	0.93	0.89	0.97	-

Supplementary Table D5. *Covariates for the model predicting the probability of having no depressive symptoms amongst patients with a complete set of Patient Health Questionnaire-9 scores across eight appointments*

	Odds Ratio	**95 % Confidence Intervals**		**p-value**
Question				
1	0.37	0.28	0.50	< 0.001
2	0.17	0.12	0.25	< 0.001
3	0.61	0.47	0.80	< 0.001
4	0.18	0.13	0.26	< 0.001
5	1.54	1.24	1.91	< 0.001
6	0.35	0.26	0.47	< 0.001
7	1.00	0.79	1.26	0.988
8	11.75	9.82	14.05	< 0.001
9	22.28	16.92	29.36	-
Appointment	1.25	1.22	1.28	< 0.001
Male (ref: female)	0.79	0.60	1.04	0.088
Age	1.01	1.00	1.02	0.119
Black, Asian, and ethnic minority (ref: White)	0.96	0.58	1.60	0.886
Not working (ref: employed)	1.01	0.77	1.33	0.943
Index of Multiple Deprivation	1.00	0.99	1.01	0.511

Disability (ref: none)	1.18	0.78	1.81	0.432
Long-term Health condition (ref: none)	1.14	0.84	1.55	0.386
Recurrent depressive disorder (ref: Depressive episode)	0.87	0.64	1.18	0.367
Baseline Patient Health Questionnaire-9	0.84	0.81	0.87	< 0.001
Baseline Generalised Anxiety Disorder Scale-7	0.99	0.95	1.02	0.471
Baseline Work and Social Adjustment Scale	0.99	0.97	1.01	0.163
No medication (ref: taking medication)	1.30	0.98	1.71	0.065
Referral number	0.91	0.84	0.99	0.032
Referral Source (ref: Self)				
Primary Care	0.99	0.72	1.35	0.945
Other	0.64	0.32	1.28	0.212
Service (ref: A)				
B	1.27	0.85	1.90	0.242
C	1.71	0.76	3.85	0.193
D	1.85	0.61	5.60	0.274
E	2.40	0.92	6.27	0.073
F	2.19	0.39	12.34	0.375
G	1.20	0.49	2.95	0.690
H	2.09	1.05	4.18	0.037
I	1.22	0.67	2.25	0.518
J	1.01	0.56	1.84	0.975

Year (ref: 2014)				
2015	1.14	0.60	2.14	0.688
2016	0.84	0.45	1.59	0.603
2017	0.64	0.35	1.18	0.152
2018	0.63	0.34	1.15	0.131
2019	1.09	0.44	2.72	0.846
Question 1 x Appointment	1.03	0.98	1.09	0.194
Question 2 x Appointment	1.07	1.01	1.14	0.022
Question 3 x Appointment	0.95	0.90	1.00	0.031
Question 4 x Appointment	1.03	0.97	1.10	0.326
Question 5 x Appointment	0.96	0.92	1.00	0.035
Question 6 x Appointment	1.04	0.99	1.10	0.093
Question 7 x Appointment	1.00	0.96	1.05	0.913
Question 8 x Appointment	0.92	0.89	0.96	< 0.001
Question 9 x Appointment	0.99	0.95	1.04	-

**P-values for the reference group of the sum coding cannot be estimated.*

Supplementary Table D6. *Covariates for the model predicting the probability of having no generalised anxiety symptoms amongst patients with a complete set of Generalised Anxiety Disorder-7 scores eight appointments*

	Odds Ratio	**95% Confidence Intervals**		**p-value**
Question				
1	0.09	0.05	0.16	< 0.001
2	0.38	0.25	0.58	< 0.001
3	0.26	0.17	0.41	< 0.001
4	0.72	0.50	1.04	0.080
5	19.01	14.48	24.95	< 0.001
6	1.82	1.32	2.50	< 0.001
7	4.68	3.12	7.04	-
Appointment	1.37	1.32	1.41	< 0.001
Male (ref: female)	1.27	0.81	1.98	0.299
Age	1.00	0.99	1.02	0.531
Black, Asian, and ethnic minority (ref: White)	0.81	0.32	2.04	0.650
Not working (ref: employed)	0.65	0.42	1.01	0.056
Index of Multiple Deprivation	1.01	0.99	1.02	0.367
Disability (ref: none)	0.76	0.35	1.65	0.482
Long-term Health condition (ref: none)	1.28	0.78	2.10	0.327
Baseline Patient Health Questionnaire-9	0.93	0.89	0.97	0.002

Baseline Generalised Anxiety Disorder Scale-7	0.81	0.77	0.86	< 0.001
Baseline Work and Social Adjustment Scale	0.99	0.96	1.02	0.480
No medication (ref: taking medication)	1.20	0.82	1.77	0.356
Referral number	0.92	0.79	1.06	0.236
Referral Source (ref: Self)				
Other	0.93	0.36	2.45	0.888
Primary Care	0.86	0.54	1.35	0.502
Service (ref: A)				
B	0.93	0.44	1.98	0.859
C	0.14	0.01	2.20	0.160
D	1.43	0.38	5.39	0.600
E	1.51	0.57	4.06	0.409
G	1.15	0.50	2.61	0.744
H	0.61	0.18	2.12	0.438
I	1.68	0.73	3.90	0.224
J	0.66	0.20	2.15	0.491
Year (ref: 2014)				
2015	1.22	0.57	2.64	0.606
2016	1.86	0.71	4.87	0.206
2017	1.33	0.52	3.37	0.549
2018	1.48	0.60	3.66	0.393

2019	2.73	0.72	10.42	0.141
Question 1 x Appointment	1.11	1.01	1.23	0.038
Question 2 x Appointment	1.06	0.99	1.14	0.100
Question 3 x Appointment	1.09	1.01	1.17	0.030
Question 4 x Appointment	1.04	0.98	1.11	0.228
Question 5 x Appointment	0.87	0.83	0.92	< 0.001
Question 6 x Appointment	0.95	0.89	1.00	0.059
Question 7 x Appointment	0.91	0.85	0.97	-

**P-values for the reference group of the sum coding cannot be estimated.*

9.5 Supplementary Material E: Chapter 7

The Early Impact of COVID-19 on Primary Care Psychological Therapy Services: A Descriptive Time Series of Electronic Healthcare Records

Appointment attendance, consultation medium, and ethnicity were grouped based on theoretical grounds. Appointment attendance was classified based on patients having clinical contact, with cancellation groupings remaining the same. Consultation medium was classified based on appointments taking place in person or remotely, with a separate level when it was not possible to determine this distinction. Ethnicity was divided into majority and minority ethnic groups. Referral source was reduced on the basis on available data points for each level, retaining the original labels for the two primary referral sources. Missing data was labelled as unknown if there was no entry recorded or the data entry indicated that the measurement was not available.

Supplementary Table E1**.** *Variable definition*

Variable	**Class**	**Levels**	**Data Source**	**Measurement**
Referral-Level Data				
Referrals	Count	N.A.	IAPT MDS	System capture
Referral Source	Categorical	Self-Referral Primary Care Other *Local Authority Services* *Employer* *Justice System*	IAPT MDS	Clinician report

		Child Health *Independent/Voluntary Sector* *Acute Secondary Care* *Other Mental Health NHS Trust* *Internal referrals from Community Mental Health Team (within own NHS Trust)* *Internal referrals from Inpatient Service (within own NHS Trust)* *Transfer by graduation (within own NHS Trust)* *Other* *IAPT*		
Age	Continuous	N.A.	IAPT MDS	Self-report
Gender	Categorical	Female Male Unknown *Not known* *Not specified* *Missing*	IAPT MDS	Self-report
Ethnicity	Categorical	White Black, Asian, and ethnic minority	IAPT MDS	Self-report

		Black *Asian* *Mixed* *Other* Unknown *Not Stated* *Not Known* *Missing*		
Long-Term Condition Status	Categorical	Long-Term Condition No Long-Term Condition Unknown *Unknown (Person asked and does not know or is not sure)* *Not Stated (Person asked but declined to provide a response)* *Missing*	IAPT MDS	Self-report
Previous Referrals	Continuous	N.A.	IAPT MDS	System capture
Baseline Patient Health	Continuous	N.A.	IAPT MDS	Self-report

Questionnaire-9 (PHQ-9)				
Baseline Generalised Anxiety Disorder Scale-7 (GAD-7)	Continuous	N.A.	IAPT MDS	Self-report
Index of Multiple Deprivation (IMD)	Continuous	N.A.	ONS database	Self-report
People per Km^2	Continuous	N.A.	ONS database	Self-report
Appointment-Level Data				
Appointments	Count	N.A.	IAPT MDS	System capture
Attendance		Attended *Attended on time or, if late, before the relevant care professional was ready to see the patient* *Arrived late, after the care professional was ready to see the patient, but was seen* Did not attend or late	IAPT MDS	Clinician report

		Patient arrived late and could not be seen		
		Did not attend - no advance warning given		
		Cancelled by patient		
		Cancelled by provider		
Consultation Medium	Categorical	Face-to-face	IAPT MDS	Clinician report
		Remote		
		Telephone		
		Telemedicine web camera		
		Email		
		Short Message Service (SMS)		
		Other		
		Other		
		Talk type for a Person unable to speak		
		Unknown		
		Missing		

**N.A. – Not Applicable; IAPT MDS - Improving Access to Psychological Therapies Minimum Dataset; ONS - Office of National Statistics*

Supplementary Table E2. *Characteristics of referrals from 1st January 2019 to 24th May 2020 stratified by NHS Trust*

	Trust 1	*Trust 2*	*Trust 3*	*Trust 4*	*Trust 5*
n	25,980	28,828	37,593	47,331	32,091
Age	38·13 (15·17)	41·09 (16·19)	35·95 (13·64)	37·47 (15·42)	38·01 (14·85)
Gender - n (%)					
Female	16,930 (65·2)	18,486 (64·1)	24,909 (66·3)	31,144 (65·8)	21,537 (67·1)
Male	9,022 (34·7)	10,331 (35·8)	12,636 (33·6)	16,108 (34·0)	10,546 (32·9)
Unknown	28 (0·1)	11 (0·0)	48 (0·1)	79 (0·2)	8 (0·0)
Ethnicity - n (%)					
White	22,386 (86·2)	24,908 (86·4)	16,195 (43·1)	3,9987 (84·5)	13,488 (42·0)
Black, Asian, and ethnic minority	1,049 (4·0)	1,329 (4·6)	14,576 (38·8)	2,491 (5·3)	13,155 (41·0)
Unknown	2,545 (9·8)	2,591 (9·0)	6,822 (18·1)	4,853 (10·3)	5,448 (17·0)
Index of Multiple Deprivation	17·70 (12·15)	15·79 (11·08)	25·28 (10·70)	21·41 (12·71)	22·82 (9·27)
People per Square Kilometre	2,878·67 (2653·96)	3,071·35 (2680·27)	14,676·56 (9303·41)	4,638·41 (3755·63)	10,005·72 (6163·44)
Long-Term Condition Status - n (%)					
Long-Term Condition	6,575 (25·3)	10,871 (37·7)	8,902 (23·7)	15,915 (33·6)	7,299 (22·7)
No Long-Term Condition	14,317 (55·1)	16,716 (58·0)	21,404 (56·9)	28,244 (59·7)	17,304 (53·9)
Unknown	5,088 (19·6)	1,241 (4·3)	7,287 (19·4)	3,172 (6·7)	7,488 (23·3)

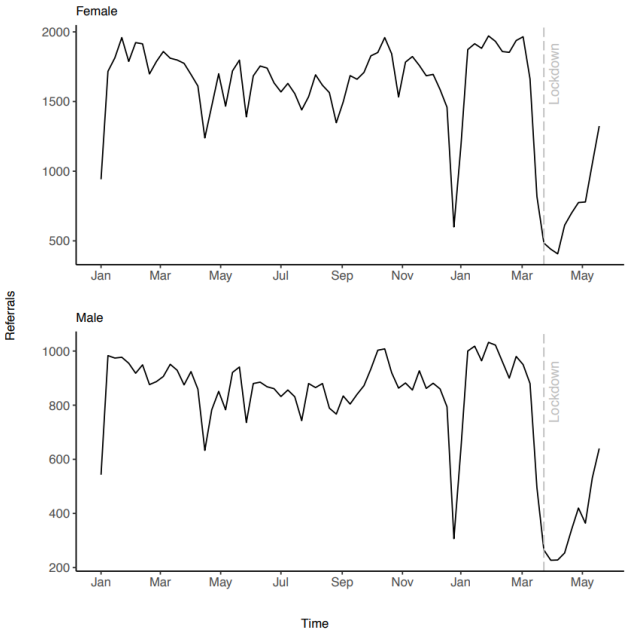
Referral Number	2·41 (2·16)	1·96 (1·50)	1·85 (1·57)	2·08 (1·58)	1·81 (1·56)
Baseline PHQ-9*	15·64 (6·29)	13·65 (6·60)	13·88 (6·56)	15·07 (6·03)	13·76 (6·48)
Baseline GAD-7*	13·85 (5·08)	12·12 (5·58)	12·50 (5·67)	13·09 (5·07)	12·51 (5·54)
Referral Source - n (%)					
Self	17,502 (67·4)	25,935 (90·0)	27,498 (73·1)	40,329 (85·2)	18,825 (58·7)
Primary Care	7,330 (28·2)	2,043 (7·1)	8,778 (23·4)	2,621 (5·5)	11,900 (37·1)
Other	1,148 (4·4)	850 (2·9)	1,317 (3·5)	4,381 (9·3)	1,366 (4·3)

**Data is presented as mean (standard deviation) unless otherwise specified. PHQ-9: Patient Health Questionnaire -9; GAD-7: Generalised Anxiety Disorder Scale -7. *Data until 17th May 2020.*

Supplementary Table E3. *Total weekly referrals 9 weeks after lockdown compared between 2019 and 2020*

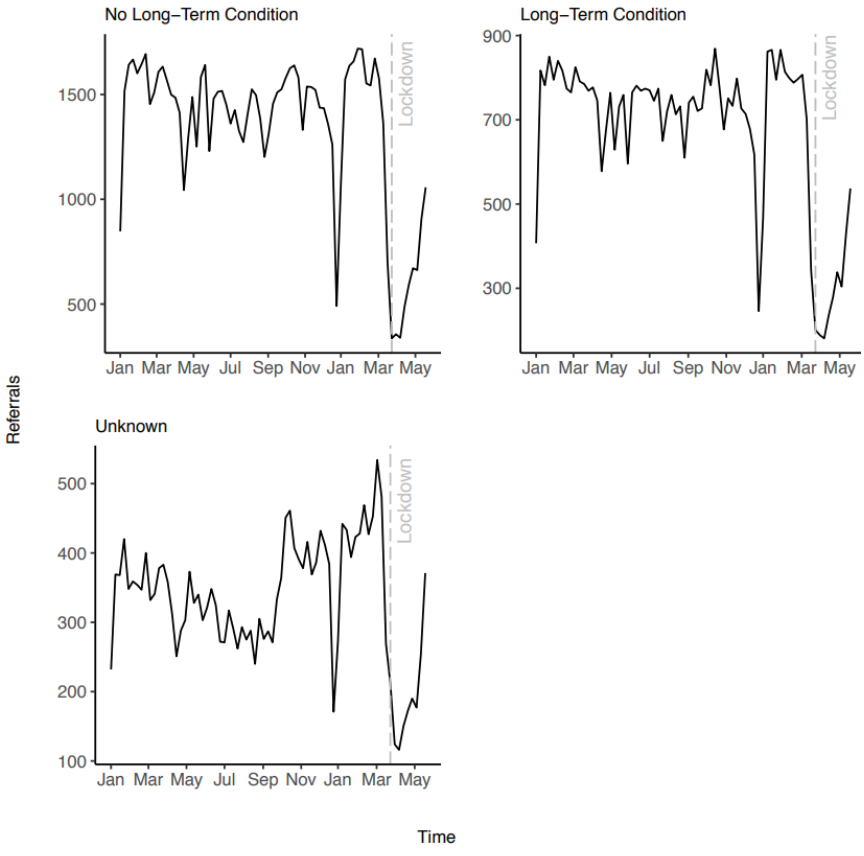
	Referrals 2019	**Referrals 2020**	**Difference (n)**	**Difference (%)**
	2,651	749	-1,902	-72
	2,620	669	-1,951	-74
	2,471	636	-1,835	-74
	1,873	870	-1,003	-54
	2,253	1,040	-1,213	-54
	2,555	1,199	-1,356	-53
	2,253	1,143	-1,110	-49
	2,642	1,588	-1,054	-40
	2,740	1,965	-775	-28
Sum/average			-12,199	-55

Supplementary Figure E1. *Average weekly referrals by gender from 1st January 2019 to 24th May*



**Missing data for gender is omitted as it was trivial*

Supplementary Figure E2. *Total weekly referrals by long-term condition status from 1st January 2019 to 24th May*



9.6 Supplementary Material F: Data Access Statement

For chapter 2 of this PhD, secondary data from two randomised controlled trials were used. Access to PANDA is restricted – to obtain access to the data a request must be made to the principal investigator, Glyn Lewis (glyn.lewis@ucl.ac.uk). The CoBalT dataset has a controlled access level. As such, a formal request to obtain the data must be made to the principal investigator, Nicola Wiles, via the University of Bristol's data request form (https://data.bris.ac.uk/data/dataset/iaar3yzlt7fp2llz8lb5terqz).

Data used for the present PhD in chapters 3,4 and 5 contains anonymous, individual-level secondary data in the form of electronic healthcare records from Improving Access to Psychological Therapies (IAPT) services. Data ownership lies with the NHS (the data holders). As such, the data cannot be made available by the author of this thesis. Access to individual-level data from the NHS requires an application through the Health Research Authority and individual NHS trusts (https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/). Further information on the process used in the present PhD can be found in the discussion section.