Understanding treatment-related phenotypes in depression and anxiety: genetic and longitudinal approaches

Skelton, Megan

Awarding institution:
King's College London

Understanding treatment-related phenotypes in depression and anxiety: genetic and longitudinal approaches

Megan Skelton

Social, Genetic and Developmental Psychiatry Centre

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Thesis submitted for the degree of

Doctor of Philosophy in Psychology & Statistical Genetics

October 2022

Abstract

Depression and anxiety are the most common mental health disorders and represent a substantial burden at both the societal and individual level. Effective treatments are available for depression and anxiety, however, a substantial proportion of patients do not recover. This is partly explainable by individual differences in response to treatment. Individual differences *during* psychological therapy have been overlooked in comparison with endpoint outcomes but may be especially informative for clinical decisions and expectations, as well as progress monitoring. The onset and prognosis of depression and anxiety are influenced by a complex interplay of genetic and environmental factors including a large number of genetic variants with individually very small effects. Genetic studies of depression and anxiety require vast sample sizes for sufficient statistical power to detect these effects. Identifying associated genetic variants can elucidate downstream biological pathways and could contribute to prediction models of disorder onset and treatment outcomes. This thesis explores two main themes. First, assessing the use of resource-saving 'brief phenotypes' of treatment-related variables to increase sample size and thus power for genetic studies. Second, investigating the existence of individual differences in longitudinal patterns of treatment outcomes during psychological therapy, as evidenced by multiple subgroups of trajectories. Chapter 1 provides an overview of literature and concepts relevant to this thesis. Chapter 2 is a study investigating self-reported medication use in the UK Biobank as a brief phenotype of depression and anxiety. The analysis presented in Chapter 3 is an investigation of the genetic overlap between symptom severity and functional impairment in a sample of patients with lifetime experience of depression or anxiety. This informs our understanding of using brief measures of symptom severity in genetic studies. The two studies in the second half of the thesis, Chapters 4 and 5, explore whether there are multiple subgroups of patients distinguished by similar outcome trajectories during psychological therapy. For this analysis, electronic treatment records from the NHS Improving Access to Psychological Therapies (IAPT) services were modelled. Chapter 4 uses data from in-person IAPT services, while Chapter 5 is an analysis of patients who received real-time therapy via the internet. The final chapter presents a discussion of findings from the study chapters in relation to one another, general strengths and limitations, and future directions.

Table of contents

Table of figures and tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Acknowledgements

The research presented in this thesis was only possible thanks to the study participants, patients and all the individuals involved in the design and management of the studies and data. I am also grateful to the NIHR Maudsley BRC for providing the opportunity and funding for me to undertake this PhD.

Thank you to my supervisors, Professors Thalia Eley and Gerome Breen. A special thanks to Thalia for being so kind, supportive and understanding during a very difficult personal time for me. I have learnt a lot from you and am always inspired by the combination of enthusiasm and pragmatism with which you approach research. I feel very lucky to have spent my PhD as part of the EDIT Lab, from the point of view of helping me develop as a researcher as well as all the rest, from silly games at summer and Christmas parties to yoga calendar updates.

I have thoroughly enjoyed being based at the SGDP centre, surrounded by great minds and, more importantly, lovely people. Seeing how everyone (virtually!) came together during the coronavirus pandemic was heart-warming. A huge thanks to Chris and Joni for all that you have taught me. I really appreciate you sharing your knowledge, time, and patience. I'm especially grateful for Kirstin, Tom, and Alicia, who have always been there to provide me with laughter, support, a motivational coworking session, and pet pics. Thank you to all my other amazing friends from both KCL and my life outside.

Thank you to my wonderful parents, Pip and Ned, and my sister, Corrina. And finally, thank you to Brian, who has supported me throughout my PhD from my first clumsy attempts at R coding, to moving to the middle of the Welsh countryside, and everything that's happened since. I'm ever grateful that you've been there to make me laugh, serenade me with drum solos and keep me fed. Although I'll never forgive you for making your own PhD look quite so easy.

Statement of authorship

All the work presented in this thesis is my own, except where acknowledged in the text. The investigations were completed by me, as first author, in collaboration with co-authors named in the author lists at the start of each chapter and acknowledged in the chapter-specific author contribution statements below. Data collection for all study samples was completed by the study research team or as part of routine patient data recording.

Megan Skelton

Chapter 2: M.S., T.C.E and G.B. conceived the research question. Quality control of genetic data was performed by the UK Biobank team and a second round by K.G and J.R.I.C. Medication codes were mapped by H.G and C.H prior to this study. Statistical analysis was performed by M.S. with guidance from C.R., J.R.I.C and K.L.P. M.S. wrote and revised the manuscript including all tables and figures. All authors critically reviewed the manuscript.

Chapter 3: M.S., T.C.E and G.B. conceived the investigation. Genetic data was processed by the NIHR Cambridge BRC and cleaned by co-authors as part of the data freeze. M.S. performed the statistical analysis with guidance from J.M, A.T.K and J.R.I.C. M.S. wrote and revised the manuscript along with all tables and figures. C.R., J.R.I.C and T.C.E. provided critical feedback of the manuscript.

Chapter 4: M.S. and T.C.E conceived the investigation. SLaM IAPT data was managed and extracted by the NIHR Maudsley BRC CRIS team. Clinical expertise was provided by J.E.J.B., C.R.H., K.A.R., R.S. and J.W. M.S. performed the data pre-processing and statistical analysis with advice from E.C., J.E.J.B., K.A.G and R.S. M.S. wrote and revised the manuscript and created all tables and figures. All authors critically reviewed the manuscript.

Chapter 5: M.S. and T.C.E conceived the investigation in agreement with A.C. Data was collected and managed by the ieso team. M.S. performed statistical analysis, wrote and revised the manuscript and created all tables and figures. All authors critically reviewed the manuscript.

Publications related to this thesis

Publications resulting from chapters of this thesis:

Chapter 2 has been peer-reviewed and published:

Skelton, M., Rayner, C., Purves, K. L., Coleman, J. R., Gaspar, H. A., Glanville, K. P., ... & Eley, T. C. (2021). Self-reported medication use as an alternate phenotyping method for anxiety and depression in the UK Biobank. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *186*(6), 389-398. https://doi.org/10.1002/ajmg.b.32878

Chapter 3 will be submitted for publication following approval from co-authors.

A version of Chapter 4 is available as a preprint on the PsyArXiv server. It is currently undergoing peer-review at *Psychological Medicine* following an invitation to revise and resubmit:

Skelton, M., Carr, E., Buckman, J. E. J., Davies, M. R., Goldsmith, K. A., Hirsch, C., … & Eley, T. C. (2021, November 16). Trajectories of depression and anxiety symptoms during psychological therapy for common mental health problems. https://doi.org/10.31234/osf.io/8scpx

A version of Chapter 5 is available as a preprint on the PsyArXiv server and has received an invitation to revise and resubmit from *Behaviour Research and Therapy*:

Skelton, M., Catarino, A., Brown, S., Carr, E., Davies, M. R., Peel, A., … & Eley, T. C. (2022, February 20). Trajectories of depression symptoms, anxiety symptoms and functional impairment during internet-enabled cognitive-behavioural therapy. https://doi.org/10.31234/osf.io/qxz2k

Other co-authored research articles published or under review during this PhD:

Published

1. Davies, M. R., Kalsi, G., Armour, C., Jones, I. R., McIntosh, A. M., Smith, D. J., ... **Skelton, M.,** … & Breen, G. (2019). The Genetic Links to Anxiety and Depression (GLAD) Study: Online recruitment into the largest recontactable study of depression and anxiety. *Behaviour Research and Therapy*, *123*, 103503. https://doi.org/10.1016/j.brat.2019.103503

2. Rayner, C., Coleman, J. R., Purves, K. L., Carr, E., Cheesman, R., Davies, M. R., ... **Skelton, M.,** ... & Eley, T. C. (2021). Sociodemographic factors associated with treatment-seeking and treatment receipt: cross-sectional analysis of UK Biobank participants with lifetime generalised anxiety or major depressive disorder. *BJPsych Open*, *7*(6), e216. https://doi.org/10.1192/bjo.2021.1012

3. Mundy, J., Hübel, C., Gelernter, J., Levey, D., Murray, R. M., **Skelton, M.,** ... & Coleman, J. R. (2021). Psychological trauma and the genetic overlap between posttraumatic stress disorder and major depressive disorder. *Psychological Medicine*, 1-10. https://doi.org/10.1017/S0033291721000830

4. Thompson, K. N., Hübel, C., Cheesman, R., Adey, B. N., Armour, C., Davies, M. R., ... **Skelton, M.,** … & Eley, T. C. (2021). Age and sex-related variability in the presentation of generalized anxiety and depression symptoms. *Depression and anxiety*, *38*(10), 1054-1065. https://doi.org/10.1002/da.23213

5. Peel, A. J., Armour, C., Buckman, J. E., Coleman, J. R., Curzons, S. C., Davies, M. R., ... **Skelton, M.,** … & Eley, T. C. (2021). Comparison of depression and anxiety symptom networks in reporters and non-reporters of lifetime trauma in two samples of differing severity. *Journal of affective disorders reports*, *6*, 100201. https://doi.org/10.1016/j.jadr.2021.100201

6. Buckman, J. E., Saunders, R., Arundell, L. L., Oshinowo, I. D., Cohen, Z. D., O'Driscoll, C., ... **Skelton, M.,** … & Pilling, S. (2021). Life events and treatment prognosis for depression: A systematic review and individual patient data meta-analysis. *Journal of affective disorders*, *299*, 298-308. https://doi.org/10.1016/j.jad.2021.12.030

7. Peel, A. J., Purves, K. L., Baldwin, J. R., Breen, G., Coleman, J. R., Pingault, J. B., ... **Skelton, M.,** … & Eley, T. C. (2022). Genetic and early environmental predictors of adulthood self-reports of trauma. *The British Journal of Psychiatry*, 1-8. https://doi.org/10.1192/bjp.2021.207

8. Davies, M. R., Buckman, J. E., Adey, B. N., Armour, C., Bradley, J. R., Curzons, S. C., ... **Skelton, M.,** … & Eley, T. C. (2022). Comparison of symptom-based versus self-reported diagnostic measures of anxiety and depression disorders in the GLAD and COPING cohorts. *Journal of Anxiety Disorders*, *85*, 102491. https://doi.org/10.1016/j.janxdis.2021.102491

Accepted

9. Rayner C., Coleman, J. R. I., **Skelton, M.**, Armour, C., Bradley, J., Buckman, J. E. J., … & Eley, T. C. Patient characteristics associated with retrospectively self-reported treatment outcomes following psychological therapy for anxiety or depressive disorders - a cohort of GLAD Study participants. *BMC Psychiatry*.

Under review

10. Parsons, C., Purves, K. L., **Skelton, M.**, Peel, A., Davies, M., Rijsdijk, F., … & Young, K. S. Different trajectories of depression, anxiety and anhedonia symptoms in the first 12 months of the COVID-19 pandemic in a UK longitudinal sample. Submitted to *Psychological Medicine*.

11. Davies, M. R., Glen, K., Mundy, J., ter Kuile, A. R., Adey, B. N., Armour, C., … **Skelton, M.,** … & Eley, T. C. Factors associated with anxiety disorder comorbidity. Submitted to *Journal of Affective Disorders*.

Chapter 1: Introduction

1.1 Depression and anxiety are common mental health disorders associated with substantial functional impairment

1.1.1 Depression, anxiety, and functional impairment

Depression and anxiety are debilitating mental health disorders that can affect not only one's emotional state, but behaviour, cognition, perception and physiology. Reference to depression here is to the most common form, major depressive disorder. A diagnosis of depression requires at least five of nine symptoms, which include decreased concentration, suicidal ideation, and sleeping too much or not enough (Diagnostic and Statistical Manual (DSM-5); American Psychiatric Association, 2013). At least one must be a core symptom of low mood or the loss of interest or pleasure in previously enjoyed activities (anhedonia) and symptoms must be present most days within a two-week period. The five core anxiety disorders in the DSM-5 are generalised anxiety disorder (GAD), social anxiety disorder (social phobia), agoraphobia, panic disorder and specific phobia. Anxiety disorders are primarily characterised by symptoms of excessive fear and anxiety, with the situations that elicit these varying between diagnostic subtypes. Diagnostic criteria for most anxiety disorders state that symptoms typically persist for at least six months.

Feelings of sadness, fear and anxiety are part of a normal range of emotions and are even posited to have adaptive roles (Badcock et al., 2017; Lee et al., 2006). A key distinction for clinical presentations is therefore their severity, frequency and impact on functioning. Diagnostic criteria for depression and most anxiety disorders require that the symptoms cause clinically significant distress or impairment in important areas of functioning, such as self-care, occupational or social activities. Although functional impairment is a diagnostic criterion for depression and anxiety, there is a notable lack of agreement on its definition within the literature, nor is it clearly operationalised within diagnostic manuals. In this thesis, functional impairment refers to difficulties performing tasks and roles, which are attributed to one's mental health symptoms. Measures can be objective (e.g., days out of work) or subjective (e.g., self-report rating of work performance), with the latter more

common. As measured by role limitations or days off work, substantial functional impairment is observed in depression and anxiety (Brenes, 2007; Kessler et al., 1999). For example, severe or very severe functional impairment, especially in the social role domain, is experienced by 60% of individuals with depression (Kessler et al., 2003). 'Functional impairment' has been used interchangeably with 'quality of life' (Brenes, 2007; Löwe et al., 2008) but they are somewhat distinct, with quality of life a more subjective perception of satisfaction with life goals and activities (Jacoby et al., 2014; Zimmerman et al., 2008).

1.1.2 Impact of depression and anxiety

Globally, depression and anxiety are the most prevalent mental health disorders and among the ten most burdensome illnesses, as measured by years lived with disability (i.e., functional impairment; GBD 2019 Mental Disorders Collaborators, 2022). Depression has an estimated lifetime prevalence of 11% internationally (Lim et al., 2018), with up to 17% in the United States (Kessler & Bromet, 2013). International lifetime prevalence of any anxiety disorder is roughly 17% (Somers et al., 2006), with estimates of up to a third of the population in Western cultures (Bandelow & Michaelis, 2015). These comparatively higher rates than other mental health disorders have led to depression and anxiety being collectively referred to as 'common mental health disorders'. Their prevalence is approximately twice as high for women as for men (Salk et al., 2017; Somers et al., 2006) and the median age of onset is 30 years for depression and 17 years for any anxiety disorder (Solmi et al., 2022).

There are vast direct and indirect costs to the individual experiencing depression or anxiety. Individuals are at risk of increased mortality (Pratt et al., 2016), lower quality of life (Rapaport et al., 2005) and death by suicide (Baxter et al., 2014; Ferrari et al., 2013). These disorders also create substantial cumulative economic costs to society and a high demand on healthcare services. In the United Kingdom (UK), depression and anxiety are conservatively estimated to account for 23% and 18% of the annual cost of mental health disorders, respectively, amounting to approximately £27 billion and £21 billion (McDaid et al., 2022). There is clearly an overwhelming need for effective prevention and treatment of common mental health disorders.

1.1.3 Functional impairment is an important yet often overlooked treatment-related outcome

Functional impairment is a clear target for intervention to help achieve diagnostic remission, improve patient quality of life, and minimise the economic impact of depression and anxiety. Crucially, returning to one's usual level of functioning is considered very important by patients (Zimmerman et al., 2006) and how they define remission (Zimmerman et al., 2008). Despite this, functional impairment is often overlooked in routine treatment and in clinical trials, with recovery often operationalised solely in terms of symptom improvement (Kamenov et al., 2015). This is possibly based on the conflation of functional impairment and symptom severity, and the resulting assumption that symptomatic improvement will simultaneously reduce functional impairment. These two concepts are undoubtedly closely related, indeed, functional impairment as defined here is not possible without experiencing symptoms. However, the minority of treatment outcome studies that have measured functional impairment suggest that they are partially independent. Symptoms and impairment contribute independently to explain remission (Jha et al., 2019; Zimmerman et al., 2008) and symptoms explain only part of the variance in impairment (Brenes, 2007; Denninger et al., 2011; Pedersen et al., 2017; Rapaport et al., 2005). Therefore, individuals with the same diagnosis and level of symptom severity can experience different levels of functional impairment. There is also evidence that symptoms and functional impairment do not change synchronously, rather, impairment lags behind, persisting after symptomatic remission and perhaps improving later (K. I. Howard et al., 1986; Sacchetti et al., 2015; Saris et al., 2017). In an antidepressant trial, 93% of patients were not within normal functioning range prior to treatment, and at the point of leaving treatment over 60% remained clinically impaired (IsHak et al., 2016). This included 20-40% (across treatment arms) of patients who had reached *symptomatic* remission, highlighting that treatment efficacy could be overestimated if functional impairment is omitted. This is particularly problematic as residual or worsening functional impairment predicts recurrence (Hardeveld et al., 2010; Rodriguez et al., 2005). Besides symptom severity, proposed contributors to functional impairment include comorbidities, social isolation, self-efficacy, cognitive ability and socioeconomic status (Chow et al., 2022), as well as personality factors (Verboom et al., 2011).

1.1.4 Comorbidity

Co-occurrence of depression and anxiety or of multiple anxiety disorders is frequently observed (Wittchen et al., 1994), with depression and anxiety comorbidity in around 60% of cases (Mineka et al., 1998; Moffitt et al., 2007). Depression and anxiety share many first-line treatments (NICE, 2011) and have notable symptom overlap (L. A. Clark & Watson, 1991). Depression and anxiety also often co-occur with other diagnoses, including eating disorders, substance use disorders, and personality disorders (Mineka et al., 1998; Plana-Ripoll et al., 2019). In addition to worse functional impairment, comorbidity is associated with poorer treatment outcomes (Amati et al., 2018; Kessler et al., 2017).

1.2 Depression and anxiety are moderately heritable and are therefore caused by a combination of genetic and environmental factors

It is commonly observed that relatives show greater behavioural resemblance than random members of the population. Indeed, mental health disorders are acknowledged to 'run in families', which has been formally investigated via family studies assessing the degree of risk in relatives of an affected individual (Hettema et al., 2001; Sullivan et al., 2000). Studies of twins are more informative as they can disentangle the influence of shared genetics from shared environment, using structural equation models to partition variance between sources. The proportion of variation in a phenotype within a population that can be attributed to genetics is known as 'heritability'. Heritability is a population-level statistic, therefore it is not descriptive at the individual level. Furthermore, it depends on the environment and the population being studied, for example, differing across life stages (Nivard et al., 2015). Twin-based heritability estimates are around 30-45% for depression and anxiety (Hettema et al., 2001; Sullivan et al., 2000). These are considered moderate, compared with, for example, 81% heritability of schizophrenia (Sullivan et al., 2003). There have been far fewer studies investigating functional impairment, however, impairment in an individual with depression appears to confer additional risk to their relatives (Sullivan et al., 2000). Twin heritability estimates for social dysfunction (20%; Rijsdijk et al., 2003) and the impact of emotional problems on daily activities (around 30%; Romeis et al., 2005) suggest that functional impairment is also moderately heritable. Other studies have used

scales that include questions about emotional symptoms, making it difficult to interpret the results specifically for functional impairment (Steenstrup et al., 2013).

Liability to mental health traits is thought to fall on a normally distributed continuum, with clinical ranges at the upper tail. The 'liability-threshold model' (Gottesman & Shields, 1967) proposes that the same genetic influences are at play across this continuum, but individuals presenting with a mental health diagnosis have exceeded a certain threshold of genetic load at the extreme of the distribution. The environment also contributes to this liability; both nature and nurture play a role in all behavioural and psychological traits, including mental health disorders (Plomin et al., 2016; Turkheimer, 2000). Therefore, mental health disorders do not arise from deterministic genetic effects but from numerous genetic factors both independently and in combination with the environment, as well as independent environmental influences. Environmental factors associated with increased risk of developing depression and anxiety include childhood abuse, stressful life events such as financial problems and the personality trait neuroticism (Fryers & Brugha, 2013; Kendler et al., 1999; McKay et al., 2021).

1.3 Genome-wide association studies reveal genetic variants that are robustly associated with depression and anxiety

1.3.1 Genome-wide methods - the why and how

Twin studies have enabled us to establish the overall contribution of genetics to depression and anxiety. This is important for our understanding of their aetiologies and can be used to empower patients (Austin, 2020), inform psychiatric nosology, and design prevention studies using environmental modification for individuals at familial risk. To understand the genetic architecture of a disorder, that is, the number, effect size and frequency of specific genetic variants associated with risk of a disorder (Sullivan & Geschwind, 2019) investigations are required at the level of genomic data (deoxyribonucleic acid (DNA)). We currently know little about the biological underpinnings of depression and anxiety. Identifying genetic variants involved in disorder onset can reveal the downstream causal pathways, offering targets for treatment. Genetic variants could also provide biomarkers to aid diagnosis, which is useful as familial risk may be unknown, and nevertheless is

uncertain due to random segregation of genes. Identifying specific variants associated with disorder course and treatment response could also enable personalised care.

Early attempts at variant identification tested for associations with a-priori selected variants based on theories of pathophysiology and pathways targeted by effective medications. The most famous of these 'candidate gene studies' is a polymorphism in the serotonin transporter gene, "5-HTTLPR", thought to increase risk of depression (Collier et al., 1996) and anxiety (Lesch et al., 1996). However, candidate gene studies were plagued by non-replication (Border et al., 2019). Several key advancements, including the creation of reference catalogues of genetic variants (International HapMap Consortium, 2003), permitted the move to genome-wide association studies (GWAS). GWAS have produced robust findings using an unbiased, hypothesis-free method for investigating trait-variant associations. Individuals, who are typically not closely related, are genotyped for common genetic variation occurring in at least 1-5% of the population. The type of genetic variation primarily explored is the most frequent, the single nucleotide polymorphism (SNP) consisting of a different DNA base at a specific location. Hundreds of thousands or millions of common variants are captured on a SNP genotyping array (or 'chip'). Each variant is tested for association with the phenotype of interest therefore a strict significance threshold is used to reduce the chance of false positives (typically $p < 5x10^{-8}$). The variance in the phenotype explained by all the significant SNPs provides a heritability estimate. Determining whether a significantly associated variant ('hit') has a causal role in liability to a disorder, and the nature of that role, requires detailed further analyses. 'Linkage disequilibrium', refers to correlations between SNPs in a given genomic region. It is not necessary that a specific causal variant is genotyped as long as it is in linkage disequilibrium with another variant that *is* on the array and can therefore 'tag' its effect.

1.3.2 Depression and anxiety are complex traits associated with a large number of common genetic variants individually of very small effect size

Attempts to identify variants associated with depression and anxiety at a genome-wide significant level met with limited success for years. Prior to the Psychiatric Genomics Consortium (PGC) depression mega-analysis in 2013, the eight existing GWAS of depression

had identified only one locus of possible significance (Kohli et al., 2011; MDD Working Group of the PGC, 2013). A locus is a known chromosomal position of a gene or genetic marker, at which more than one independently associated SNP can be present. The PGC 2013 study was the largest depression GWAS of its time: over 9,000 cases primarily phenotyped via clinician-led structured interviews, and more than 9,000 controls, yet not a single significant SNP was identified (MDD Working Group of the PGC, 2013). GWAS of anxiety also struggled to identify significant variants and received fewer efforts in comparatively small samples (e.g., under 2,000 cases; Otowa et al., 2012). A GWAS of schizophrenia in 2014 identified 108 genome-wide significant loci with almost 40,000 cases and over 110,000 controls (Ripke et al., 2014). This was pertinent because depression and anxiety, being less heritable than schizophrenia, were likely associated with variants of smaller effect that would require even larger samples to be detected. A major finding of early GWAS of depression and anxiety was therefore a lack of genetic variants of large effect size. Importantly, GWAS of other traits (e.g., cholesterol) have shown that small genetic effects do not negate the potential for large effects in downstream pathways that can be therapeutically targeted (Sullivan et al., 2018). Furthermore, aggregating risk variants of small effects may have clinical utility, as will be discussed later.

Significant SNP heritability estimates of depression and anxiety were identified following a move to methods that considered all common genotyped variants, regardless of significance level (e.g., genome-wide complex trait analysis (GCTA; Yang et al., 2011)). Using this method, the estimate for the PGC 2013 depression mega-analysis was 21% (Cross-Disorder Group of the PGC, 2013). This provided further evidence that the heritability of these traits is due to the cumulative influence of many variants with individually very small effects, which had been missed due to the stringent genome-wide significance threshold. Depression and anxiety are therefore described as highly polygenic disorders. Notably, SNP heritability from these methods will fall short of twin estimates as they are limited to the main effects of common, causal variants that are directly genotyped or tagged by SNP arrays, or imputed from a reference panel, as opposed to all genetic factors (Yang et al., 2017). The remaining 'missing heritability' between twin and SNP estimates is likely due to imperfect linkage disequilibrium, such that SNPs on the genotyping arrays cannot capture the effects of all the true causal SNPs. This is especially

likely if a causal SNP is rare (present in < 1% of the population) as these are not widely used on arrays, and the likelihood of linkage disequilibrium with a common variant is low (Yang et al., 2017). Besides only moderate heritability, small effect sizes, and possible rare variant effects, efforts to identify depression- and anxiety-associated variants have been further hindered by highly heterogeneous symptom presentations. A depression diagnosis can be met by over 1,000 different combinations of symptoms (Fried & Nesse, 2015), which could have different genetic profiles. Different environmental exposures (e.g., history of childhood trauma) and comorbidities (Levinson et al., 2014) further exacerbate the issue of heterogeneity. The prevalence of depression and anxiety also means that cases and controls are more phenotypically similar (assuming a normal distribution of traits), which reduces statistical power to identify genetic differences (Mullins & Lewis, 2017). Due to the complex nature of these disorders, it is also likely that many participants who do not meet diagnostic criteria have at least some genetic and environmental risk factors (Mullins & Lewis, 2017). To increase statistical power to detect differences in variant frequency associated with depression and anxiety, two main approaches have been proposed: reduce sample heterogeneity or increase sample size.

1.4 Increasing sample size for greater statistical power to identify genetic variants

The quest for far greater sample sizes led to a culture change within psychiatric genetics. Individual research groups would not have the time and money to phenotype and genotype sufficient participants, therefore large international consortia formed, such as the PGC. By performing meta-analyses and mega-analyses (individual-level data meta-analyses), samples considered unfeasibly large in most domains of psychological research have become the norm. For example, one of the first successes in anxiety variant discovery was a meta-analysis of studies totalling over 18,000 participants that had undergone clinical interviews to diagnose anxiety subtypes (Otowa et al., 2016).

To further help achieve large samples, depression and anxiety case definitions have been relaxed, using resource-saving methods to define case status rather than clinical ascertainment. These methods are referred to as ‘brief’ or ‘minimal’ phenotyping and examples include self-report of a diagnosis or symptoms. The population-based cohort UK

Biobank (Bycroft et al., 2018) largely relies on these methods and constitutes a noteworthy part of depression and anxiety genetic research. The rationale is that vast samples can outweigh the increased heterogeneity of a less precise measure to detect meaningful genetic signal. Some of the first support for brief phenotyping came from a GWAS combining the PGC 2013 depression samples and self-reported depression diagnosis or treatment from the commercial DNA testing service 23andme (Hyde et al., 2016). The resultant sample of approximately 26,000 cases and 232,000 controls revealed 15 significant loci and a significant SNP heritability estimate of 6%. Several depression GWAS have since capitalised on this approach, reporting an increasing number of significant variants. For example, an analysis of over 800,000 individuals reported 102 significant variants, 87 of which replicated in an independent sample, and a SNP heritability of 9% (D. M. Howard et al., 2019). As an example of successful brief phenotyping in anxiety disorders, 25,453 cases and 58,113 controls were identified using UK Biobank data on self-reported diagnosis or a more detailed definition based on a lifetime diagnostic questionnaire (Purves et al., 2020). Five loci reached genome-wide significance and SNP heritability was estimated at 26%. Analyses of brief phenotypes have also implicated genes involved in brain regions and pathways widely held to be involved in these disorders, including targets of antidepressant treatments (D. M. Howard et al., 2019; Purves et al., 2020; Wray et al., 2018). Many variants have replicated in independent samples both within and between studies. Evidence in favour of brief phenotyping also comes from high genetic overlap with more detailed clinical phenotypes (e.g., D. M. Howard et al., 2019).

Scores from symptom-based questionnaires have also been used as brief phenotypes. For GWAS of common traits, quantitative phenotypes can provide more statistical power than case-control (Yang et al., 2010). Symptom scores appear to identify much of the same genetic signal as more clinically phenotyped samples (Direk et al., 2017; Levey et al., 2020) and in combination with these samples can boost heritability estimates, consistent with a liability-threshold model of psychopathology (Direk et al., 2017). However, in isolation brief phenotypes typically produce lower heritability estimates than detailed phenotypes (e.g., depression symptoms 4% (Direk et al., 2017); clinical interview 21% (MDD Working Group of the PGC, 2013)). Critics of brief phenotyping argue that this weaker genetic signal, alongside the incomplete genetic overlap between brief and detailed phenotypes, might

reflect a different phenotype such that identified variants are not specific to the diagnosis (Cai et al., 2020). Critics therefore contend that statistical power should be improved by reducing heterogeneity and, supporting this, an early successful GWAS of depression took this approach. By using low coverage whole genome sequencing and selecting on Han Chinese ancestry, female sex and depression recurrence, two significant loci were identified, despite having approximately half of the PGC 2013 sample size (Cai et al., 2015). Conversely, no notable differences in heritability estimates were found between GWAS of depression stratified by sex or recurrence, and analysis across these strata (Hall et al., 2018). The authors of the study concluded that until we better understand what determines genetic heterogeneity and therefore relevant phenotypic groups, larger sample sizes will likely be more fruitful. Furthermore, lower heritability estimates from brief phenotypes may be due to increased noise, alongside relevant signal, rather than a separate phenotype (Schwabe et al., 2019). Using multiple measures to triangulate case status might improve the reliability of brief phenotypes (Glanville et al., 2021).

Brief phenotyping is therefore a pragmatic approach with the caveats that some of the variants specific to full diagnostic presentations may be missed and in case-control designs some controls, or cases for other disorders, are probably classed as depression or anxiety cases (Cai et al., 2020). The identified variants might therefore be more relevant for furthering our understanding of the biology of mental illness broadly whilst also informing follow-up analyses with more clinically defined phenotypes. The optimal way forward is in parallel efforts to increase sample size through brief phenotyping *and* to perform clinically precise phenotyping to investigate heterogeneity (Schwabe et al., 2019).

1.4.1 Genome-wide association studies of functional impairment

As in the broader literature, functional impairment has been under-researched in GWAS. One GWAS investigated functional impairment as a transdiagnostic phenotype across 2,246 patients with bipolar disorder, schizophrenia and depression but no variants were found (McGrath et al., 2013). The authors proposed that differences in functional impairment at the same level of symptom severity might represent a risk-resilience trait independent of diagnosis and could be used to increase sample sizes in genetic research.

1.5 Genetic correlations and polygenic risk scores

1.5.1 Genetic correlations describe genetic overlap between traits

Genetic overlap between two traits can be described by calculating a genetic correlation (r_g). This reflects the extent of pleiotropy, whereby the same variant influences two traits. Values vary between -1 and 1, where 1 indicates that all the same variants are influencing both traits in the same direction. Genetic correlations can be estimated from both twin and genomic data. They can help to explain phenotypic relationships, such as comorbidity, identify shared biological pathways and inform how we think about diagnostic categories. Knowledge of genetically similar but diagnostically distinct disorders can also be used to increase GWAS sample sizes. Although genetic correlations can arise from a variety of processes besides true biological pleiotropy, if the aim is to use genetic information from both traits for prediction, then the mechanism is ignorable (van Rheenen et al., 2019).

1.5.2 Polygenic risk scores represent individual-level genetic liability for a phenotype

Polygenic risk scores (PRS) represent the aggregate, additive effect of trait-associated SNPs in an individual. To calculate a PRS, the first step is to acquire information on trait-variant associations from a GWAS in a 'discovery' sample. In a non-overlapping 'target' sample, individual PRS are then created by determining the presence of SNPs that were identified in the discovery GWAS, weighting each one by its GWAS effect size, and summing. An individual's PRS indicates their genetic liability for the discovery trait. Analyses can then be performed in the target sample, such as a regression of PRS and a phenotypic measure of the same trait as the discovery GWAS, or a different trait. The proportion of variance explained by the PRS represents the ability of the genetic variants to identify independent, known, cases, thereby indicating whether the PRS could predict risk in samples of *unknown* affected status. When another trait is the outcome, the variance explained indicates genetic overlap, whether the variants and the size of their associations with the discovery GWAS trait can explain differences in this second trait. The target sample in which the PRS are calculated does not need to be as large as those required for GWAS (Choi et al., 2020).

Currently, PRS do not have clinical value for predicting risk of depression or anxiety at an individual level. The proportion of variance explained, whilst significant, is small. In two studies, depression PRS explained up to 3.2% of the variance in depression case status (D. M. Howard et al., 2019) and anxiety PRS explained 0.5% of variance in anxiety case status (Purves et al., 2020). Due to the complex nature of depression and anxiety and their moderate heritability, PRS will never conclusively predict disorder risk, but could improve predictive models alongside other clinically relevant variables. PRS also show promise in terms of stratification. One study reported that the odds of being a depression case in the tenth depression PRS decile were over twice as high as in the first decile (Wray et al., 2018). Furthermore, PRS were positively associated with severity across a range of indices including recurrence (Wray et al., 2018). PRS could therefore be used to identify at-risk individuals for prevention studies, aid formal diagnosis when unclear based on clinical presentation and indicate risk of severe prognosis within case samples (Wray et al., 2021). PRS could also have value in personalised medicine by predicting a patient's treatment outcome or revealing treatment-specific associations, with implications for treatment choice. Significant heritability estimates of antidepressant response have been reported, indicating the feasibility of this (Pain et al., 2022). Thus far, PRS have been used more widely as a research tool, rather than for prediction (Maier et al., 2018). Larger samples using brief phenotypes, will slowly increase the predictive power of PRS. Incorporating multiple PRS of correlated phenotypes into a model also provides predictive gains (Krapohl et al., 2018). PRS are among the first robust, valid, biomarkers for psychopathology (Sullivan et al., 2018) and exhibit the utility of GWAS, despite the small effect sizes of individual variants.

1.6 Phenotypic comorbidity partially arises from genetic overlap

As previously described, the co-occurrence of depression and anxiety is common. In fact, in mental health, comorbidity is the rule rather than the exception (Plana-Ripoll et al., 2019). A substantial part of this phenotypic covariance can be explained by pleiotropy (Eley & Stevenson, 1999). Twin studies showed that the genetic influences on depression and anxiety appeared to be largely the same, whilst environmental influences were primarily disorder-specific (Kendler et al., 1992; Roy et al., 1995). Twin estimates of genetic correlations often do not differ from unity (r_g = 1; Eley & Stevenson, 1999; Kendler, 1996;

Kendler et al., 1992; Roy et al., 1995) and SNP estimates also show high genetic overlap (r_g approximately 0.8; Levey et al., 2020; Ohi et al., 2019; Purves et al., 2020). Furthermore, GWAS of each disorder have identified a number of the same specific loci (Ask et al., 2021) and PRS show cross-disorder prediction (Meier et al., 2019; Otowa et al., 2016). Despite their similarities, depression and anxiety do not appear to either represent phases of the same disorder or only occur due to the presence of the other (Middeldorp et al., 2005). Furthermore, research suggests that their genetic correlation is not driven by overlapping symptoms (e.g., sleep and concentration problems; Eley & Stevenson, 1999). Conditional analyses have also revealed depression-associated genetic enrichment in brain tissues not observed for anxiety (Grotzinger et al., 2022) and GAD specific variants independent of depression (Levey et al., 2020). Therefore, depression and anxiety appear to be distinct disorders but with notable commonalities. This led to proposals to leverage their commonality by creating a broader phenotype capturing both, thereby increasing sample size for genetic studies (Hettema et al., 2005; Kendler et al., 2007).

Consistent with observations that depression is more often comorbid with GAD than other anxiety disorders (Kessler et al., 2005), phenotypic factor analyses show that depression and GAD load strongly onto a 'distress' factor and other anxiety disorders onto a 'fear' factor (L. A. Clark & Watson, 2006; Krueger, 1999; Mineka et al., 1998). These two factors are correlated and represented by an 'internalising' factor. There is evidence that this model is also observed genetically, with stronger genetic correlations between depression and GAD than with fear-based anxiety disorders (Hettema et al., 2005; Mineka et al., 1998; Morneau-Vaillancourt et al., 2020; Waszczuk et al., 2014).

Genetic correlations have also been reported for depression and anxiety with numerous co-occurring mental health disorders and symptoms, as well as various other phenotypes. For example, positive genetic correlations have been observed with schizophrenia, personality disorders, neuroticism, coronary artery disease and smoking (Luciano et al., 2018; Meier et al., 2019; Purves et al., 2020; Witt et al., 2017; Wray et al., 2018). Negative correlations have been identified with traits such as years of education and age at menarche (Meier et al., 2019; Wray et al., 2018).

The ‘general genes’ (Eley, 1997), or ‘p-factor’ theory (Caspi et al., 2014; Lahey et al., 2012) postulates a general underlying genetic predisposition to psychopathology captured by shared risk variants across disorders, in addition to disorder-specific genetic effects. Genetic correlations across putatively diverse disorders support this theory (Brainstorm Consortium et al., 2018; Selzam et al., 2018) and call into question diagnostic categories; “Our genes don’t seem to have read the DSM” (p.411, Smoller et al., 2019). This highlights the potential for transdiagnostic investigations of genetic risk. However, a recent genomic structural equation model study recommended biological insights would be better obtained by focusing on genetic correlations within and between lower-level factors, such as internalising disorders (Grotzinger et al., 2022).

1.7 Treatment of depression and anxiety

1.7.1 Recommended treatments

Consistent with the phenotypic and genetic overlap between depression and anxiety, these disorders share several successful treatment approaches. In the UK, the National Institute for Health and Care Excellence (NICE) recommends cognitive-behavioural therapy (CBT) and selective-serotonin reuptake inhibitors as the first-line psychological and pharmacological treatments. CBT is a collection of interventions with diagnostic-specific protocols and manuals focusing on changing maladaptive cognitions and behaviours. For example, for anxiety disorders there will often be a behavioural exposure component. The therapist and patient discuss specific problems in the session, and between sessions the patient completes ‘homework’ practising the skills and techniques they are learning. For common mental health disorders, NICE recommends a stepped care approach whereby patients first receive the least intensive treatment and can be stepped up or down to maximise the chance of response whilst conserving resources (NICE, 2011). The main steps are presented in a flowchart below (Figure 1). Self-help groups, befriending, education and employment support services are offered across disorders and steps.

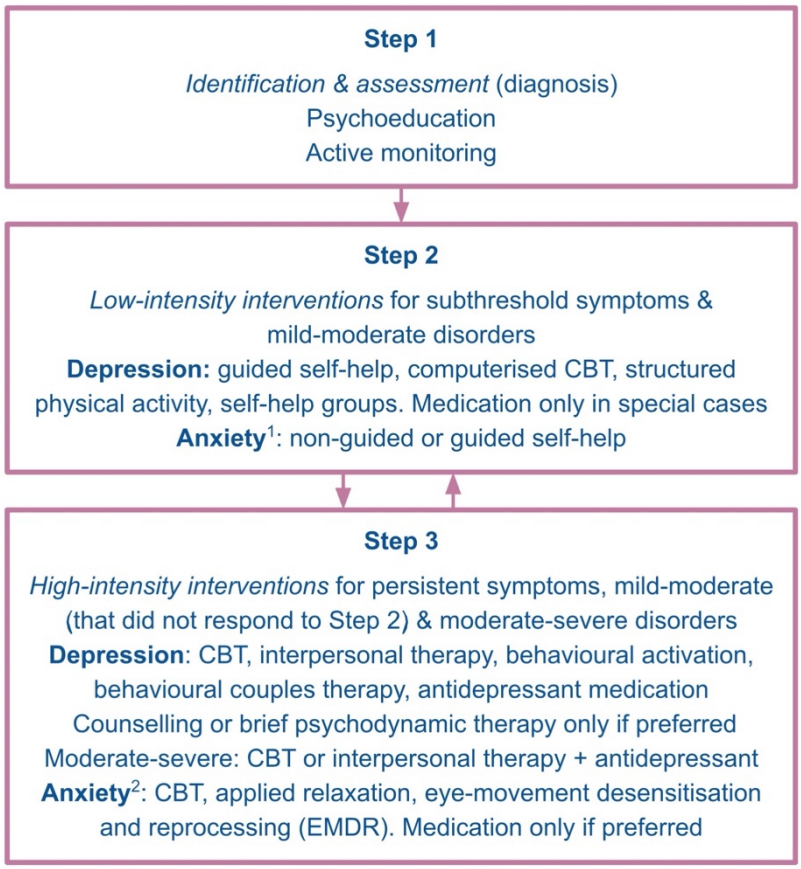


Figure 1. NICE recommended stepped care for common mental health disorders

*[1] GAD and panic disorder. Only **guided** self-help is recommended for obsessive-compulsive disorder (OCD). [2] GAD, panic disorder, OCD, social anxiety disorder and post-traumatic stress disorder (PTSD). Applied relaxation is only recommended for GAD and EMDR only for PTSD.*

1.7.2 Approximately half of patients improve following treatment

The average remission rate across treatments is around 50% in trials of depression (Cuijpers, Karyotaki, et al., 2021; Undurraga & Baldessarini, 2012), and anxiety (Loerinc et al., 2015; Springer et al., 2018), as well as in routine treatment (IAPT Team, NHS Digital, 2021). Despite continued efforts to produce more effective treatments, on average, treatments do not appear to significantly differ. This has been shown across psychological therapies (Barth et al., 2013; Cuijpers, Quero, et al., 2021) and between psychological and pharmacological treatments (Carl et al., 2020; Cuijpers et al., 2020; Cuijpers, Sijbrandij, et al., 2013). Combined psychological therapy and medication does appear to result in slightly better outcomes than monotherapy (Cuijpers et al., 2015, 2020; Cuijpers, Sijbrandij, et al., 2014). Some studies have conversely reported superior effects for specific treatments

(Bandelow et al., 2015; Cipriani et al., 2018; Mayo-Wilson et al., 2014). Inconsistencies may be partly attributable to definitions of treatment outcome. A review of CBT trials for anxiety found response rates ranged between 0 and 100% depending on how response was operationalised (Loerinc et al., 2015). Inconsistent findings can also arise due to the lack of an appropriate, active control group, especially in psychotherapy trials (Carl et al., 2020; Cuijpers et al., 2016). In favour of psychotherapy there is evidence of a more sustained long-term effect, decreasing the chance of depression relapse or recurrence compared with medication (Cuijpers, Hollon, et al., 2013; Karyotaki et al., 2016). Moreover, drug-induced side-effects are a concern. Patients with depression or anxiety generally report a preference for psychological treatment over pharmacological (McHugh et al., 2013), yet medication use is far more common (Kendrick, 2021; McManus et al., 2016).

1.8 NHS Improving Access to Psychological Therapies (IAPT) services

The National Health Service (NHS) England Improving Access to Psychological Therapies (IAPT) services were launched in 2008. This initiative was a response to NICE recommendations of evidence-based psychological therapies for common mental health disorders as well as reported economic benefits of investing in psychological healthcare. Specifically, this was proposed to decrease welfare support and increase work productivity in those who have received treatment (D. M. Clark, 2011; D. M. Clark et al., 2009). IAPT services primarily treat adults with depression and/or anxiety, including post-traumatic stress disorder (PTSD) and obsessive-compulsive disorder (OCD). NICE-recommended low-intensity treatments such as guided self-help are offered by Psychological Wellbeing Practitioners, usually up to eight sessions. High-intensity treatments are delivered by accredited therapists and consist of more traditional, structured individual and group-based therapies, over a higher number of sessions (e.g., 15 to 20 for CBT). Referrals can be made by a GP or other health or community services. Additionally, self-referral is available to increase treatment access, especially for underrepresented groups. No formal diagnosis is made, and no symptom thresholds are used for accessing treatment. The patient's primary presenting mental health problem to be targeted in treatment is identified through discussion with a clinician. The 'problem descriptor' is recorded using an International Classification of Diseases-10 (ICD-10) code, although patients do not have to meet all

diagnostic criteria. This assessment informs allocation to treatment, ensuring that disorder-specific protocols are followed (Roth & Pilling, 2007). Some patients do not go on to have a full course of treatment, instead receiving a session of assessment, psychoeducation and possible signposting to other services. IAPT offers a more representative sample of individuals experiencing depression and anxiety than is available from clinical trials, which often have strict inclusion criteria. Substance misuse, suicidal ideation, self-harm and psychotic or manic symptoms do not automatically exclude an individual from IAPT, therefore services treat complex cases (Hepgul et al., 2016). However, where these features are present to the extent that IAPT treatments are considered unsuitable, patients are referred to specialist services.

In 2020 to 2021, IAPT services received almost 1.5 million referrals, of which a million accessed treatment (IAPT Team, NHS Digital, 2021). Of the 634,649 patients that completed a course of treatment (minimum two treatment sessions), 51% recovered, reflecting rates from clinical trials (IAPT Team, NHS Digital, 2021). A recent meta-analysis of IAPT studies showed large effects for reductions in depression and anxiety symptoms, and medium for improved functional impairment (Wakefield et al., 2021). Several service-level factors are associated with higher rates of recovery, including provision of accurate problem descriptors and a higher average number of treatment sessions (D. M. Clark et al., 2018; Gyani et al., 2013; Saunders et al., 2020). Improvements in some of these areas and higher quality data recording may underlie increased national recovery rates (Saunders et al., 2020).

From a research point of view, IAPT is an excellent resource of a vast amount of treatment data, beyond the scale feasible from a clinical trial. The key value of IAPT as a research resource is in its routine outcome monitoring at every clinical contact. The measures include three principal self-report questionnaires: the Patient Health Questionnaire 9-item version (PHQ9; Kroenke et al., 2001) which assesses depression symptoms, Generalized Anxiety Disorder 7-item Scale (GAD7; Spitzer et al., 2006) for anxiety symptoms, and the Work and Social Adjustment Scale (WSAS; Marks, 1986) to measure functional impairment. Recovery in IAPT is calculated as moving from above to below case thresholds on both the

PHQ9 and GAD7. The PHQ9 and GAD7 are also widely used in research settings and are the outcomes in three of the studies in this thesis.

1.9 Internet-delivered psychological therapy

Over the last couple of decades, technological advances have resulted in the development of a host of computer- and internet-based psychological therapy programmes to treat depression and anxiety. The majority are now delivered via the internet and are built on CBT principles (Burger et al., 2020). Formats include websites with only text-based information or with added interactive content and exercise modules. A key point of variation is in the level of provided support. Unguided self-help versions offer only autonomous feedback or technical or administrative support. Guided programmes assign patients a therapist. These programmes can be *guided self-help* whereby the therapist provides *asynchronous* support and feedback, such as weekly emails or brief calls, outside of self-help therapeutic sessions. Alternatively, they can be more akin to traditional therapy and offer a course of regular *synchronous* therapy sessions via real-time video or messaging platforms (Burger et al., 2020; Webb et al., 2017). Unguided formats show comparatively inferior outcomes and higher rates of dropout to guided treatments, especially for patients with more severe symptoms (Andersson, Titov, et al., 2019; Karyotaki et al., 2021). IAPT services offer internet-delivered interventions, both as low-intensity guided self-help and high-intensity synchronous therapy sessions (Thew, 2020). Attitudes towards internet-delivered interventions seem to be favourable (Andrews et al., 2018), although more so among patients than therapists (Schröder et al., 2017).

Meta-analytic evidence shows large pre- to post-treatment symptom improvement for guided internet CBT compared to waitlist or treatment as usual controls (Andersson, Carlbring, et al., 2019; Andersson, Titov, et al., 2019; Karyotaki et al., 2021). These effects are sustained at long-term follow up (Andersson et al., 2018). Stronger evidence for the efficacy of guided internet CBT comes from comparisons to face-to-face CBT showing equivalent effects (Andersson, Titov, et al., 2019), and similar rates of attrition (Carlbring et al., 2018), although larger sample sizes are required for more robust evidence.

It is plausible that certain aspects of therapy benefit from an in-person format, such as exposure to phobic situations. Furthermore, internet-delivered therapies require a degree of computer literacy, the ability to read and write in a specified language (e.g., English), and internet availability, and therefore are not universally suitable. On the other hand, internet-delivered CBT has a number of advantages that could help to increase treatment seeking and access. For example, internet-delivered CBT can be used wherever one has suitable technology and internet access and providers can offer appointments outside of normal working hours. To meet treatment needs, the use of internet-delivered therapies will likely continue to grow (Thew, 2020).

1.10 Individual differences in outcomes following psychological therapy

Individual differences in outcomes following treatment are not observable when group-based averages are used, as is often the case in clinical trials (Hollon et al., 2002). Larger treatment effect sizes may therefore exist for some subgroups. This is thought to at least partly explain the average response rate of 50% across treatments, rather than poorly developed treatments (Cohen & DeRubeis, 2018; Lorenzo-Luaces et al., 2021). The likelihood of a universally effective treatment seems low when considering the heterogeneity of depression and anxiety in terms of possible aetiological influences and clinical presentations. Evidence for treatment response heterogeneity comes from patients who show treatment-specific recovery, and associations between certain patient characteristics and treatment outcomes (Cohen & DeRubeis, 2018). Stronger support for the existence of treatment response heterogeneity comes from a meta-analysis of depression treatment trials showing that variance was 9% higher in intervention groups than control groups (Kaiser et al., 2022). Heterogeneity is the basis for personalised (or 'precision') medicine which aims to optimise treatment outcomes by identifying 'what works for whom' and allocating patients accordingly (Paul, 1967).

A host of 'prescriptive' and 'prognostic' predictors and correlates have been identified for psychological treatment response. The former refers to variables that interact with treatment type to produce different outcomes, providing information for optimum treatment allocation. A number of prescriptive variables have been identified (e.g.,

Boschloo et al., 2022; Chekroud et al., 2021). However, studying prescriptive variables requires data on multiple treatments or a control group, that are ideally randomised, and of sufficient sample size to detect moderator effects. Prognostic variables are associated with outcome across treatments, or within one treatment when there is no comparison group. They can guide clinician and patient expectations regarding treatment outcomes (Cohen & DeRubeis, 2018), identify patients to target in further treatment trials and inform interventions targeting modifiable prognostic correlates. As prognostic variables are more relevant to this thesis and the patient samples analysed, they will be further described. Table 1 provides examples of prognostic variables reported to be associated with poorer outcomes following treatment in clinical trials and in IAPT. For clinical trials, only evidence from meta-analyses or systematic reviews is reported, and for IAPT, only variables that have been investigated in more than one study. Individual differences following internet-delivered CBT for depression and anxiety have been comparatively less researched. Meta-analyses report conflicting findings regarding baseline depression severity, which may be due to the restricted, milder symptom range in trials compared with routine treatment (Andersson, Carlbring, et al., 2019; Karyotaki et al., 2018; Rozental et al., 2019). However, studies of internet-delivered CBT provide some of the only evidence for genetic influences on treatment outcomes to date. Analysing the same sample of individuals via different methods, one study reported associations with PRS for depression and intelligence (Wallert et al., 2022), and another with PRS for autism spectrum disorder (E. Andersson et al., 2019).

Table 1. Examples of prognostic variables associated with poorer outcomes following psychological treatment

Variable	Study references [1]
Strong evidence of an association	
Higher baseline symptom severity	Amati et al., 2018; Buckman, Cohen, et al., 2021; Buckman, Saunders, et al., 2021; Kessler et al., 2017 *Delgadillo, Dawson, et al., 2017; Delgadillo et al., 2016; Goddard et al., 2015*
Unemployment	Buckman et al., 2022; Kessler et al., 2017 *Buckman, Stott, et al., 2021; Delgadillo, Dawson, et al., 2017; Delgadillo et al., 2016*
Chronic course	Buckman, Saunders, et al., 2021; Kessler et al., 2017; Lorenzo-Luaces et al., 2021
Comorbid depression or anxiety disorder	Amati et al., 2018; Buckman, Saunders, et al., 2021; Kessler et al., 2017; Springer et al., 2018
Higher functional impairment	Jha et al., 2019; Kessler et al., 2017 *Delgadillo, Dawson, et al., 2017; Delgadillo et al., 2016*
Substance or alcohol use	Buckman, Cohen, et al., 2021; Springer et al., 2018
Modest evidence of an association	
Lower outcome expectancy	Constantino et al., 2018
Single relationship status	Kessler et al., 2017
Presence of long-term physical health condition or disability	*Delgadillo, Dawson, et al., 2017; Delgadillo et al., 2016*
Use of psychotropic medication	Springer et al., 2018
Lower socioeconomic status (e.g., lower income or educational attainment)	Kessler et al., 2017
Strong evidence of small or no association	
Age	Kessler et al., 2017; Springer et al., 2018 *Delgadillo et al., 2016*
Gender	Cuijpers, Weitz, et al., 2014 *Delgadillo, Dawson, et al., 2017; Springer et al., 2018*

[1]Non-italicised references indicate meta-analyses and systematic reviews of clinical trials, italicised references represent studies performed with data from IAPT. Strength of evidence base was not formally assessed.

Information about associations between patient variables and treatment outcomes can be implemented within statistically informed clinical decision tools. Statistics-based decisions produce better outcomes compared with clinician decisions (Ægisdóttir et al., 2006; Grove & Meehl, 1996). Nevertheless, clinicians will likely have additional information that a tool will not and therefore the aim is only to provide recommendations which support, not replace, clinicians' decision making.

1.11 Beyond endpoint outcomes: Trajectories of response

The previous section described research investigating individual differences in outcomes at the end of treatment, such as response and remission. Endpoint outcomes are undoubtedly important, reflecting the culmination of the therapeutic work. However, treating all recovered patients as a single group and non-recovered as another ignores possible differences between patients *during* treatment, that is, in trajectories of outcomes over therapy sessions. Heterogeneity in patterns of change may be meaningful and clinically informative. Rather than all patients following a single average trajectory with only some quantitative variation, there may be multiple subgroups of patients with qualitatively different trajectories. Some patients could then be at risk of inaccurate expectations, continuation of unsuitable treatment, or treatment ending early due to apparent non-response. Trajectory analyses could therefore identify patients at risk of non-response but also other clinically relevant subgroups. If patient characteristics can be identified that are associated with these subgroups, then a patient's most likely trajectory could be known at the start of treatment.

Structural equation modelling reveals unobserved 'latent' variables that explain relationships between observed variables (e.g., questionnaire scores). Longitudinal structural equation modelling techniques, such as growth mixture modelling, offer a person-centred approach to identify clusters of patients based on the similarity of their trajectories over treatment sessions. Using growth mixture modelling, the presence of these subpopulations can be captured by latent classes, whereby each class has its own 'latent growth curve' (trajectory; see Figure 2). The modelling process involves identifying the number of trajectory classes that best fits the observed data. Crucially, associations can be explored between potential predictors and trajectory class. Whilst more nuanced than binary response outcomes often used in endpoint outcome literature, it is important to note that the revealed classes are a simplified heuristic reflecting a more complicated reality with an underlying continuum (Lutz et al., 2014).

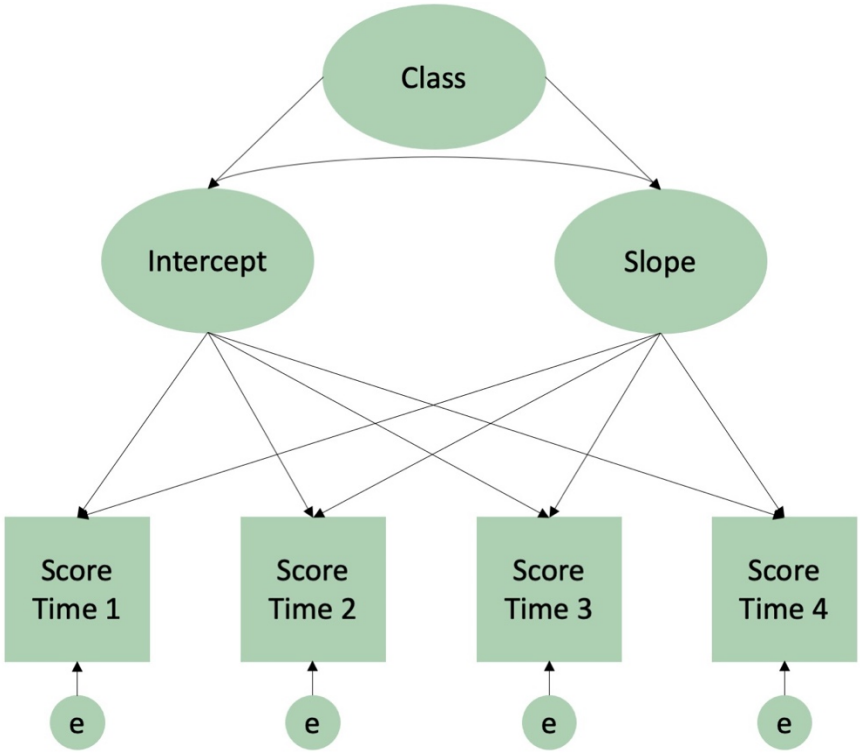


Figure 2. A growth mixture model of questionnaire scores over time

Latent variables are represented by circles and observed variables by squares. 'e' indicates error residuals. Class is a categorical latent factor. Slope and intercept are continuous latent factors, each with a mean and variance. 'Time' here could be therapy session.

1.11.1 Trajectories of depression and anxiety symptoms during psychological therapy

Existing studies of trajectory classes during psychological treatment are generally of small sample size and often focus only on 'early change patterns' in symptoms over the initial sessions. Across three of these studies in patients with depression or anxiety, optimal models had between three and five trajectory classes (Lutz et al., 2009, 2014; Stulz et al., 2007). The number of sessions included in each analysis ranged from three to six. Among the identified models, there were similar slopes of non-response plateau, fast (early) improvement and more gradual improvement. Samples had fewer than 350 patients and few, if any, associations were explored with baseline variables. Although early change patterns warrant attention as they show associations with endpoint recovery and dropout (Lutz et al., 2014), trajectories throughout a longer course of therapy have been under-researched and may reveal later relevant patterns. Furthermore, efforts are required to

identify associations with baseline characteristics that could inform prediction models of a patient's most likely trajectory at the start of treatment. One notable investigation of trajectories over a greater number of sessions was a study with over 4,000 IAPT patients (Saunders et al., 2019), which is described in more detail in Chapter 4. To my knowledge, trajectory classes of functional impairment have not been explored.

Information about trajectories could also be used for monitoring progress and providing feedback during treatment. Clinicians perform poorly at identifying patients at risk of treatment failure and overestimate their patients' rates of improvement (de Jong et al., 2021). Monitoring and feedback systems provide alerts when a patient is 'not on track' according to their expected trajectory. Factors that are possibly contributing to this can be highlighted so that the clinician can focus on them (e.g., aspects of therapeutic alliance). Monitoring and feedback tools have a well-established evidence basis for improving treatment outcomes. A meta-analysis of 58 studies covering almost 22,000 patients showed a small significant effect on symptom reduction and dropout rates, although not on the number of sessions (de Jong et al., 2021). Furthermore, a trial in IAPT services concluded that a feedback system increased the likelihood of reliable symptom improvement at a modest incremental cost (Delgadillo et al., 2021). It generally seems to be acceptable to patients and therapists to implement such systems in routine practice (Delgadillo, Overend, et al., 2017). A one-size-fits-all approach would be an unsuitable basis for these systems if, for example, a patient's expected trajectory is actually one of more gradual change than the average.

1.12 Aims and overview of this thesis

The aims of this thesis fall broadly into two halves. The first two studies focus on the genetics of brief treatment-related phenotypes and the second two studies on individual differences in longitudinal trajectories of treatment outcomes during psychological therapy. All outcomes are related to depression and anxiety. A key aim of the thesis as a whole was to better understand functional impairment, both in terms of genetics and of trajectories of change during therapy. Figure 3 provides a visualisation of how the study

chapters relate to one another, as well as the relevance of the themes from the literature I have presented to each chapter.

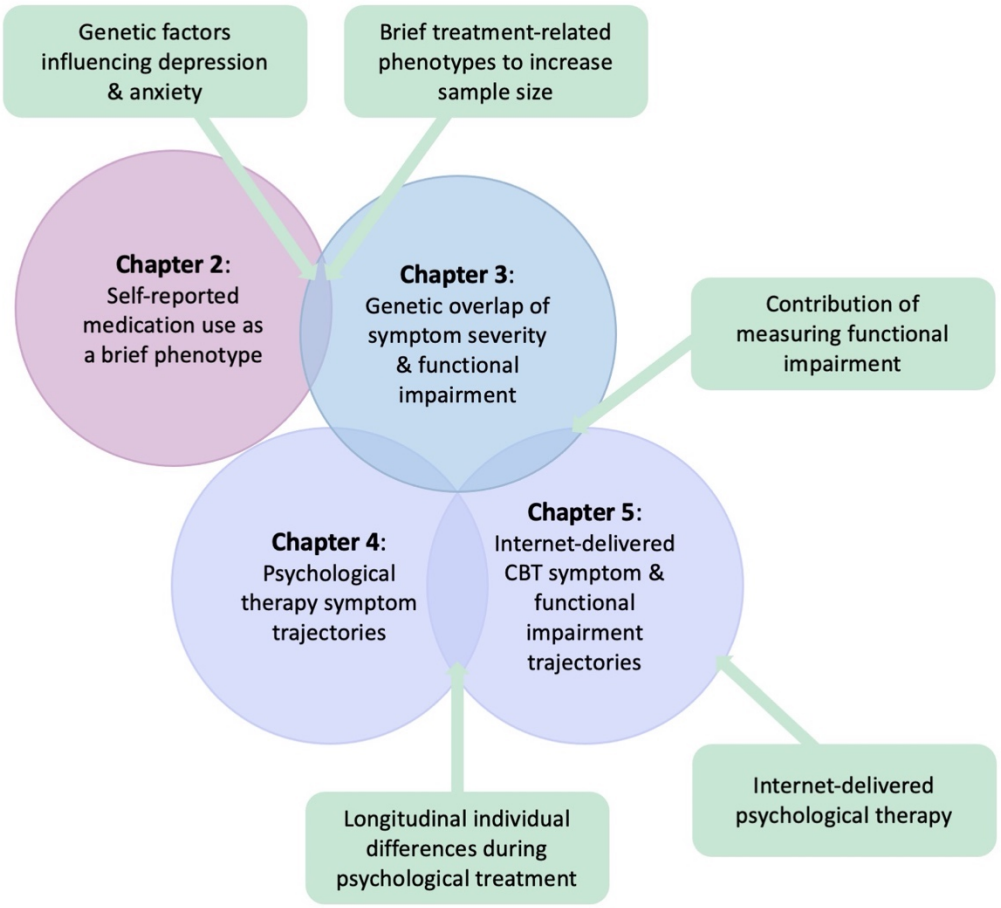


Figure 3. Visualisation of how the study chapters in this thesis relate to one another as well as the relevance of themes from the literature presented in this chapter

Chapter 2: One method of improving statistical power in genetic studies of depression and anxiety is to increase sample sizes. This can be done using 'brief phenotyping', a resource-efficient method of defining a phenotype, which often relies on self-report on a single item. Existing studies indicate that the genetic variants associated with brief phenotypes of depression and anxiety show high overlap with those from more clinical diagnostic measures. I aimed to assess self-reported medication use as an additional or alternate brief phenotype for depression and anxiety. I performed a genome-wide association study of this phenotype in the UK Biobank and calculated genetic overlap with more detailed case-control definitions using genetic correlation and polygenic risk scores.

Chapter 3: Functional impairment is often overlooked as a treatment outcome and conflated with symptom severity, despite showing only moderate phenotypic overlap. The genetic relationship between symptoms and functional impairment, a diagnostic criterion for depression and anxiety, is unknown. Establishing this would indicate whether brief phenotypes of symptom severity sufficiently capture genetic variants associated with diagnoses of depression and anxiety or if it would be valuable to also measure functional impairment. It would also reveal whether genetic information about functional impairment could contribute to prediction models of disorder prognosis, or if it is redundant when genetic data on symptom severity is available. I explored this question in a sample of participants with lifetime depression or anxiety from the Genetic Links to Anxiety and Depression (GLAD) Study. A genome wide association study was performed for each of the PHQ9 (depression symptoms), GAD7 (anxiety symptoms) and WSAS (functional impairment) measures. Genetic correlations were calculated to determine the extent of genetic overlap between symptoms and functional impairment.

Chapter 4: There is established heterogeneity in patient outcomes at the end of psychological treatment. Individual differences occurring *during* treatment have been comparatively under-researched, especially in sufficient sample sizes and beyond the initial few treatment sessions. I used growth mixture modelling to explore whether there were unobserved subgroups of symptom (PHQ9 and GAD7) trajectories during psychological therapy in the South London and Maudsley NHS IAPT services. I then explored associations between trajectory class and possible prognostic variables recorded at baseline, which could be used to indicate a patient's most likely trajectory at the start of therapy. This could guide treatment expectations and decisions, and signal whether a patient is 'off-track' according to their expected trajectory.

Chapter 5: Internet-delivered therapies offer an additional route to accessing psychological treatment for depression and anxiety. I explored whether similar trajectory subgroups and associations with patient characteristics existed in a sample of IAPT patients who had received real-time CBT sessions with an accredited therapist, via the internet. The same statistical method was used as in Chapter 4. The analysis was extended to model functional impairment (WSAS).

1.13 References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., … Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, *34*(3), 341–382.

Amati, F., Banks, C., Greenfield, G., & Green, J. (2018). Predictors of outcomes for patients with common mental health disorders receiving psychological therapies in community settings: a systematic review. *Journal of Public Health*, *40*(3), e375–e387.

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

Andersson, E., Crowley, J. J., Lindefors, N., Ljótsson, B., Hedman-Lagerlöf, E., Boberg, J., … Rück, C. (2019). Genetics of response to cognitive behavior therapy in adults with major depression: a preliminary report. *Molecular Psychiatry*, *24*(4), 484–490.

Andersson, G., Carlbring, P., & Rozental, A. (2019). Response and Remission Rates in Internet-Based Cognitive Behavior Therapy: An Individual Patient Data Meta-Analysis. *Frontiers in Psychiatry / Frontiers Research Foundation*, *10*, 749.

Andersson, G., Rozental, A., Shafran, R., & Carlbring, P. (2018). Long-term effects of internet-supported cognitive behaviour therapy. *Expert Review of Neurotherapeutics*, *18*(1), 21–28.

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry: Official Journal of the World Psychiatric Association*, *18*(1), 20–28.

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders*, *55*, 70–78.

Ask, H., Cheesman, R., Jami, E. S., Levey, D. F., Purves, K. L., & Weber, H. (2021). Genetic contributions to anxiety disorders: where we are and where we are heading. *Psychological Medicine*, *51*(13), 2231–2246.

Austin, J. C. (2020). Evidence-Based Genetic Counseling for Psychiatric Disorders: A Road Map. *Cold Spring Harbor Perspectives in Medicine*, *10*(6). https://doi.org/10.1101/cshperspect.a036608

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The Depressed Brain: An Evolutionary Systems Theory. *Trends in Cognitive Sciences*, *21*(3), 182–194.

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335.

Bandelow, B., Reitt, M., Röver, C., Michaelis, S., Görlich, Y., & Wedekind, D. (2015). Efficacy of treatments for anxiety disorders: a meta-analysis. *International Clinical Psychopharmacology*, *30*(4), 183–192.

Barth, J., Munder, T., Gerger, H., Nüesch, 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.

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine*, *44*(11), 2363–2374.

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *The American Journal of Psychiatry*, *176*(5), 376–387.

Boschloo, L., Hieronymus, F., Cuijpers, P., & ICECA Work Group. (2022). Clinical response to SSRIs relative to cognitive behavioral therapy in depression: a symptom-specific approach. *World Psychiatry: Official Journal of the World Psychiatric Association*, *21*(1), 152–153.

Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., … Murray, R. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395). https://doi.org/10.1126/science.aap8757

Brenes, G. A. (2007). Anxiety, depression, and quality of life in primary care patients. *Primary Care Companion to the Journal of Clinical Psychiatry*, *9*(6), 437–443.

Buckman, J. E. J., Cohen, Z. D., O'Driscoll, C., Fried, E. I., 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.

Buckman, J. E. J., Saunders, R., Cohen, Z. D., Barnett, P., Clarke, K., Ambler, G., … Pilling, S. (2021). The contribution of depressive "disorder characteristics" to determinations of prognosis for adults with depression: an individual patient data meta-analysis. *Psychological Medicine*, *51*(7), 1068–1081.

Buckman, J. E. J., Saunders, R., Stott, J., Cohen, Z. D., Arundell, L.-L., Eley, T. C., … Pilling, S. (2022). Socioeconomic Indicators of Treatment Prognosis for Adults With Depression: A Systematic Review and Individual Patient Data Meta-analysis. *JAMA Psychiatry* . https://doi.org/10.1001/jamapsychiatry.2022.0100

Buckman, J. E. J., Stott, J., Main, N., Antonie, D. M., Singh, S., Naqvi, S. A., … Saunders, R. (2021). Understanding the psychological therapy treatment outcomes for young adults who are not in education, employment, or training (NEET), moderators of outcomes, and what might be done to improve them. *Psychological Medicine*, 1–12.

Burger, F., Neerincx, M. A., & Brinkman, W.-P. (2020). Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research*, *22*(1), e12599.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209.

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., … Flint, J. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588–591.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*(4), 437–447.

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, *47*(1), 1–18.

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.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., … Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry: Official Journal of the World Psychiatric Association*, *20*(2), 154–170.

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772.

Chow, T. K., Bowie, C. R., Morton, M., Lalovic, A., McInerney, S. J., & Rizvi, S. J. (2022). Contributors of Functional Impairment in Major Depressive Disorder: a Biopsychosocial Approach. *Current Behavioral Neuroscience Reports*, *9*(2), 59–72.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., … Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366.

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.

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.

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

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336.

Clark, L. A., & Watson, D. (2006). Distress and fear disorders: an alternative empirically based taxonomy of the “mood” empirically based taxonomy of the “mood” and “anxiety” disorders. *British Journal of Psychiatry*, *189*, 481–483.

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review of Clinical Psychology*, *14*, 209–236.

Collier, D. A., Stöber, G., Li, T., Heils, A., Catalano, M., Di Bella, D., … Lesch, K. P. (1996). A novel functional polymorphism within the promoter of the serotonin transporter gene: possible role in susceptibility to affective disorders. *Molecular Psychiatry*, *1*(6), 453–460.

Constantino, M. J., Vîslă, A., Coyne, A. E., & Boswell, J. F. (2018). A meta-analysis of the association between patients’ early treatment outcome expectation and their posttreatment outcomes. *Psychotherapy*, *55*(4), 473–485.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., … International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry: Official Journal of the World Psychiatric Association*, *15*(3), 245–258.

Cuijpers, P., de Wit, L., Weitz, E., Andersson, G., & Huibers, M. J. H. (2015). The combination of psychotherapy and pharmacotherapy in the treatment of adult depression: A comprehensive meta-analysis. *Journal of Evidence-Based Psychotherapies*, *15*(2), 147–168.

Cuijpers, P., Hollon, S. D., van Straten, A., Bockting, C., Berking, M., & Andersson, G. (2013). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, *3*(4). https://doi.org/10.1136/bmjopen-2012-002542

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.

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry: Official Journal of the World Psychiatric Association*, *19*(1), 92–107.

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., … Furukawa, T. A. (2021). Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry: Official Journal of the World Psychiatric Association*, *20*(2), 283–293.

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons. *World Psychiatry: Official Journal of the World Psychiatric Association*, *12*(2), 137–148.

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F., III. (2014). Adding psychotherapy to antidepressant medication in depression and anxiety disorders: A meta-analysis. *World Psychiatry: Official Journal of the World Psychiatric Association*, *13*(1), 56–67.

Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., … Others. (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.

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

Delgadillo, J., Dawson, A., Gilbody, S., & Böhnke, J. R. (2017). Impact of long-term medical conditions on the outcomes of psychological therapy for depression and anxiety. *The British Journal of Psychiatry: The Journal of Mental Science*, *210*(1), 47–53.

Delgadillo, J., McMillan, D., Gilbody, S., de Jong, K., Lucock, M., Lutz, W., … Ali, S. (2021). Cost-effectiveness of feedback-informed psychological treatment: Evidence from the IAPT-FIT trial. *Behaviour Research and Therapy*, *142*, 103873.

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.

Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., … de Jong, K. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy*, *99*, 89–97.

Denninger, J. W., van Nieuwenhuizen, A. O., Wisniewski, S. R., Luther, J. F., Trivedi, M. H., Rush, A. J., … Fava, M. (2011). Changes in depressive symptoms and social functioning in the sequenced treatment alternatives to relieve depression study. *The Journal of Nervous and Mental Disease*, *199*(10), 807–810.

Dewar, M., Paradis, A., & Fortin, C. A. (2020). Identifying trajectories and predictors of response to psychotherapy for post-traumatic stress disorder in adults: A systematic review of literature. The Canadian Journal of Psychiatry, 65(2), 71-86.

Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., … Sullivan, P. F. (2017). An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biological Psychiatry*, *82*(5), 322–329.

Eley, T. C. (1997). General genes: A new theme in developmental psychopathology. *Current Directions in Psychological Science*, *6*(4), 90–95.

Eley, T. C., & Stevenson, J. (1999). Using genetic analyses to clarify the distinction between depressive and anxious symptoms in children. *Journal of Abnormal Child Psychology*, *27*(2), 105–114.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., … Whiteford, H. A. (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, *10*(11), e1001547–e1001547.

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.

Fryers, T., & Brugha, T. (2013). Childhood determinants of adult psychiatric disorder. *Clinical Practice and Epidemiology in Mental Health: CP & EMH*, *9*, 1–50.

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*(2), 137–150.

Glanville, K. P., Coleman, J. R. I., Howard, D. M., Pain, O., Hanscombe, K. B., Jermy, B., … Lewis, C. M. (2021). Multiple measures of depression to enhance validity of major depressive disorder in the UK Biobank. *BJPsych Open*, *7*(2), e44.

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy*, *73*, 1–7.

Gottesman, I. I., & Shields, J. (1967). A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(1), 199–205.

Grotzinger, A. D., Mallard, T. T., Akingbuwa, W. A., Ip, H. F., Adams, M. J., Lewis, C. M., … Nivard, M. G. (2022). Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics*, *54*(5), 548–559.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law: An Official Law Review of the University of Arizona College of Law and the University of Miami School of Law*, *2*(2), 293–323.

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.

Hall, L. S., Adams, M. J., Arnau-Soler, A., Clarke, T.-K., Howard, D. M., Zeng, Y., … McIntosh, A. M. (2018). Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational Psychiatry*, *8*(1), 9.

Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, *122*(3), 184–191.

Hepgul, N., King, S., Amarasinghe, M., Breen, G., Grant, N., Grey, N., … Cleare, A. J. (2016). Clinical characteristics of patients assessed within an Improving Access to Psychological Therapies (IAPT) service: results from a naturalistic cohort study (Predicting Outcome Following Psychological Therapy; PROMPT). *BMC Psychiatry*, *16*, 52.

Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American Journal of Psychiatry*, *158*(10), 1568–1578.

Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry*, *62*(2), 182–189.

Hollon, S. D., DeRubeis, R. J., Shelton, R. C., & Weiss, B. (2002). The emperor's new drugs: Effect size and moderation effects. *Prevention & Treatment*, *5*(1), 28.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, *41*(2), 159–164.

Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., … Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, *48*(9), 1031–1036.

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

International HapMap Consortium. (2003). The International HapMap Project. *Nature*, *426*(6968), 789–796.

IsHak, W. W., James, D. M., Mirocha, J., Youssef, H., Tobia, G., Pi, S., … Cohen, R. M. (2016). Patient-reported functioning in major depressive disorder. *Therapeutic Advances in Chronic Disease*, *7*(3), 160–169.

Jacoby, R. J., Leonard, R. C., Riemann, B. C., & Abramowitz, J. S. (2014). Predictors of quality of life and functional impairment in Obsessive–Compulsive Disorder. *Comprehensive Psychiatry*, *55*(5), 1195–1202.

Jha, M. K., South, C., Trivedi, J., Minhajuddin, A., Rush, A. J., & Trivedi, M. H. (2019). Prediction of Acute-Phase Treatment Outcomes by Adding a Single-Item Measure of Activity Impairment to Symptom Measurement: Development and Validation of

an Interactive Calculator from the STAR*D and CO-MED Trials. *International Journal of Neuropsychopharmacology*, *22*(5), 339–348.

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice* (3), 294–303.

Kamenov, K., Cabello, M., Coenen, M., & Ayuso-Mateos, J. L. (2015). How much do we know about the functional effectiveness of interventions for depression? A systematic review. *Journal of Affective Disorders*, *188*, 89–96.

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.

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. G., Furukawa, T. A., Cuijpers, P., … Forsell, Y. (2021). Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. *JAMA Psychiatry*, *78*(4), 361–371.

Karyotaki, E., Smit, Y., Holdt Henningsen, K., Huibers, M. J. H., Robays, J., de Beurs, D., & Cuijpers, P. (2016). Combining pharmacotherapy and psychotherapy or monotherapy for major depression? A meta-analysis on the long-term effects. *Journal of Affective Disorders*, *194*, 144–152.

Kendler, K. S. (1996). Major depression and generalised anxiety disorder: same genes, (partly) different environments—revisited. *British Journal of Psychiatry*, 68–75.

Kendler, K. S., Gardner, C. O., Gatz, M., & Pedersen, N. L. (2007). The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. *Psychological Medicine*, *37*(3), 453–462.

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *The American Journal of Psychiatry*, *156*(6), 837–841.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of General Psychiatry*, *49*(9), 716–722.

Kendrick, T. (2021). Strategies to reduce use of antidepressants. *British Journal of Clinical Pharmacology*, *87*(1), 23–33.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, *289*(23), 3095–3105.

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *34*, 119–138.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617–627.

Kessler, R. C., DuPont, R. L., Berglund, P., & Wittchen, H. U. (1999). Impairment in pure and comorbid generalized anxiety disorder and major depression at 12 months in two national surveys. *The American Journal of Psychiatry*, *156*(12), 1915–1923.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., … Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*(1), 22–36.

Kohli, M. A., Lucae, S., Saemann, P. G., Schmidt, M. V., Demirkan, A., Hek, K., … Binder, E. B. (2011). The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron*, *70*(2), 252–265.

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., … Plomin, R. (2018). Multi-polygenic score approach to trait prediction. *Molecular Psychiatry*, *23*(5), 1368–1374.

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.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*(10), 921–926.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, *121*(4), 971–977.

Lee, W. E., Wadsworth, M. E. J., & Hotopf, M. (2006). The protective role of trait anxiety: a longitudinal cohort study. *Psychological Medicine*, *36*(3), 345–351.

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., … Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, *274*(5292), 1527–1531.

Levey, D. F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., … Stein, M. B. (2020). Reproducible Genetic Risk Loci for Anxiety: Results From ∼200,000 Participants in the Million Veteran Program. *The American Journal of Psychiatry*, *177*(3), 223–232.

Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R., & Sullivan, P. F. (2014). Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biological Psychiatry*, *76*(7), 510–512.

Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. *Scientific Reports*, *8*(1), 2861.

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review*, *42*, 72–82.

Lorenzo-Luaces, L., Peipert, A., De Jesús Romero, R., Rutter, L. A., & Rodriguez-Quintana, N. (2021). Personalized Medicine and Cognitive Behavioral Therapies for Depression: Small Effects, Big Problems, and Bigger Data. *International Journal of Cognitive Therapy*, *14*(1), 59–85.

Löwe, B., Spitzer, R. L., Williams, J. B. W., Mussell, M., Schellberg, D., & Kroenke, K. (2008). Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *General Hospital Psychiatry*, *30*(3), 191–199.

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T. K., Shirali, M., … Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*. https://doi.org/10.1038/s41588-017-0013-8

Lutz, W., Arndt, A., Rubel, J., Berger, T., Schröder, J., Späth, C., … Moritz, S. (2017). Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression. *Journal of Medical Internet Research*, *19*(6), e206.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., … Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, *82*(2), 287–297.

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, *118*(1-3), 60–68.

Maier, R. M., Visscher, P. M., Robinson, M. R., & Wray, N. R. (2018). Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychological Medicine*, *48*(7), 1055–1067.

Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., … Sullivan, P. F. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497–511.

Marks, I. M. (1986). *Behavioural psychotherapy: Maudsley pocket book of clinical management*. Wright/Institute of Psychiatry Publishing.

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. *The Lancet. Psychiatry*, *1*(5), 368–376.

McDaid, D., Park, A.-L., Davidson, G., & John, A. (2022). *The economic case for investing in the prevention of mental health conditions in the UK*. https://www.mentalhealth.org.uk/sites/default/files/2022-06/MHF-Investing-in-Prevention-Full-Report.pdf

McGrath, L. M., Cornelis, M. C., Lee, P. H., Robinson, E. B., Duncan, L. E., Barnett, J. H., … Smoller, J. W. (2013). Genetic predictors of risk and resilience in psychiatric disorders: a cross-disorder genome-wide association study of functional impairment in major depressive disorder, bipolar disorder, and schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, *162B*(8), 779–788.

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. *The Journal of Clinical Psychiatry*, *74*(6), 595–602.

McKay, M. T., Cannon, M., Chambers, D., Conroy, R. M., Coughlan, H., Dodd, P., … Clarke, M. C. (2021). Childhood trauma and adult mental disorder: A systematic review and meta-analysis of longitudinal cohort studies. *Acta Psychiatrica Scandinavica*, *143*(3), 189–205.

McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. NHS Digital.

Meier, S. M., Trontti, K., Purves, K. L., Als, T. D., Grove, J., Laine, M., … Mors, O. (2019). Genetic Variants Associated With Anxiety and Stress-Related Disorders: A Genome-Wide Association Study and Mouse-Model Study. *JAMA Psychiatry*, *76*(9), 924–932.

Middeldorp, C. M., Cath, D. C., Van Dyck, R., & Boomsma, D. I. (2005). The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychological Medicine*, *35*(5), 611–624.

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, *49*, 377–412.

Moffitt, T. E., Harrington, H., Caspi, A., Kim-Cohen, J., Goldberg, D., Gregory, A. M., & Poulton, R. (2007). Depression and generalized anxiety disorder: cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years. *Archives of General Psychiatry*, *64*(6), 651–660.

Morneau-Vaillancourt, G., Coleman, J. R. I., Purves, K. L., Cheesman, R., Rayner, C., Breen, G., & Eley, T. C. (2020). The genetic and environmental hierarchical structure of anxiety and depression in the UK Biobank. *Depression and Anxiety*, *37*(6), 512–520.

Mullins, N., & Lewis, C. M. (2017). Genetics of Depression: Progress at Last. *Current Psychiatry Reports*, *19*(8), 43.

National Institute for Health and Care Excellence. (2011). *Common mental health problems: identification and pathways to care*. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/CG123/chapter/1-Guidance#steps-2-and-3-treatment-and-referral-for-treatment

Nivard, M. G., Dolan, C. V., Kendler, K. S., Kan, K. J., Willemsen, G., van Beijsterveldt, C. E. M., … Boomsma, D. I. (2015). Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological Medicine*, *45*(5), 1039–1049.

Ohi, K., Otowa, T., Shimada, M., Sasaki, T., & Tanii, H. (2019). Shared genetic etiology between anxiety disorders and psychiatric and related intermediate phenotypes. *Psychological Medicine*, 1–13.

Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., … Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry*, *21*(10), 1391–1399.

Otowa, T., Kawamura, Y., Nishida, N., Sugaya, N., Koike, A., Yoshida, E., … Sasaki, T. (2012). Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. *Translational Psychiatry*, *2*, e186.

Pain, O., Hodgson, K., Trubetskoy, V., Ripke, S., Marshe, V. S., Adams, M. J., … Lewis, C. M. (2022). Identifying the Common Genetic Basis of Antidepressant Response. *Biological Psychiatry Global Open Science*, *2*(2), 115–126.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109–118.

Pedersen, G., Kvarstein, E. H., & Wilberg, T. (2017). The Work and Social Adjustment Scale: Psychometric properties and validity among males and females, and outpatients with and without personality disorders. *Personality and Mental Health*, *11*(4), 215–228.

Plana-Ripoll, O., Pedersen, C. B., Holtz, Y., Benros, M. E., Dalsgaard, S., de Jonge, P., … McGrath, J. J. (2019). Exploring Comorbidity Within Mental Disorders Among a Danish National Population. *JAMA Psychiatry*, *76*(3), 259–270.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 Replicated Findings From Behavioral Genetics. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(1), 3–23.

Pratt, L. A., Druss, B. G., Manderscheid, R. W., & Walker, E. R. (2016). Excess mortality due to depression and anxiety in the United States: Results from a nationally representative survey. *General Hospital Psychiatry*, *39*, 39–45.

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A Major Role for Common Genetic Variation in Anxiety Disorders. *Molecular Psychiatry*, *25*(12), 3292–3303.

Rapaport, M. H., Clary, C., Fayyad, R., & Endicott, J. (2005). Quality-of-Life Impairment in Depressive and Anxiety Disorders. *American Journal of Psychiatry*, *162*(6), 1171–1178.

Richards, D., Richardson, T., Timulak, L., & McElvaney, J. (2015). The efficacy of internet-delivered treatment for generalized anxiety disorder: A systematic review and meta-analysis. *Internet Interventions*, *2*(3), 272–282.

Rijsdijk, F. V., Snieder, H., Ormel, J., Sham, P., Goldberg, D. P., & Spector, T. D. (2003). Genetic and environmental influences on psychological distress in the population: General Health Questionnaire analyses in UK twins. *Psychological Medicine*, *33*(5), 793–801.

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., … O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427.

Rodriguez, B. F., Bruce, S. E., Pagano, M. E., & Keller, M. B. (2005). Relationships among psychosocial functioning, diagnostic comorbidity, and the recurrence of generalized anxiety disorder, panic disorder, and major depression. *Journal of Anxiety Disorders*, *19*, 752–766.

Romeis, J. C., Heath, A. C., Xian, H., Eisen, S. A., Scherrer, J. F., Pedersen, N. L., … True, W. R. (2005). Heritability of SF-36 among middle-age, middle-class, male-male twins. *Medical Care*, *43*(11), 1147–1154.

Roth, A. D., & Pilling, S. (2007). The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders. *IAPT Program*.

Roy, M. A., Neale, M. C., Pedersen, N. L., Mathé, A. A., & Kendler, K. S. (1995). A twin study of generalized anxiety disorder and major depression. *Psychological Medicine*, *25*(5), 1037–1049.

Rozental, A., Andersson, G., & Carlbring, P. (2019). In the Absence of Effects: An Individual Patient Data Meta-Analysis of Non-response and Its Predictors in Internet-Based Cognitive Behavior Therapy. *Frontiers in Psychology*, *10*, 589.

Sacchetti, E., Frank, E., Siracusano, A., Racagni, G., Vita, A., & Turrina, C. (2015). Functional impairment in patients with major depression in clinical remission: results from the

VIVAL-D-Rem, a nationwide, naturalistic, cross-sectional survey. *International Clinical Psychopharmacology*, *30*(3), 129–141.

Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, *143*(8), 783–822.

Saris, I. M. J., Aghajani, M., van der Werff, S. J. A., van der Wee, N. J. A., & Penninx, B. W. J. H. (2017). Social functioning in patients with depressive and anxiety disorders. *Acta Psychiatrica Scandinavica*, *136*(4), 352–361.

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.

Saunders, R., Cape, J., Leibowitz, J., Aguirre, E., Jena, R., Cirkovic, M., … Buckman, J. E. J. (2020). Improvement in IAPT outcomes over time: are they driven by changes in clinical practice? *Cognitive Behaviour Therapy*, *13*, e16.

Schröder, J., Berger, T., Meyer, B., Lutz, W., Hautzinger, M., Späth, C., … Moritz, S. (2017). Attitudes Towards Internet Interventions Among Psychotherapists and Individuals with Mild to Moderate Depression Symptoms. *Cognitive Therapy and Research*, *41*(5), 745–756.

Schwabe, I., Milaneschi, Y., Gerring, Z., Sullivan, P. F., Schulte, E., Suppli, N. P., … Middeldorp, C. M. (2019). Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies. *Psychological Medicine*, *49*(16), 2646–2656.

Selzam, S., Coleman, J. R. I., Caspi, A., Moffitt, T. E., & Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Translational Psychiatry*, *8*(1), 205.

Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J., & Kendler, K. S. (2019). Psychiatric genetics and the structure of psychopathology. *Molecular Psychiatry*, 1–12.

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., … Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295.

Somers, J. M., Goldner, E. M., Waraich, P., & Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: a systematic review of the literature. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *51*(2), 100–113.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

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.

Steenstrup, T., Pedersen, O. B., Hjelmborg, J., Skytthe, A., & Kyvik, K. O. (2013). Heritability of health-related quality of life: SF-12 summary scores in a population-based nationwide twin cohort. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *16*(3), 670–678.

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, *75*(6), 864–874.

Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børglum, A. D., Breen, G., … Psychiatric Genomics Consortium. (2018). Psychiatric Genomics: An Update and an Agenda. *The American Journal of Psychiatry*, *175*(1), 15–27.

Sullivan, P. F., & Geschwind, D. H. (2019). Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*, *177*(1), 162–183.

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies. *Archives of General Psychiatry*, *60*(12), 1187–1192.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *The American Journal of Psychiatry*, *157*(10), 1552–1562.

Thew, G. R. (2020). IAPT and the internet: the current and future role of therapist-guided internet interventions within routine care settings. *Cognitive Behaviour Therapy*, *13*, e4.

Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*, *9*(5), 160–164.

Undurraga, J., & Baldessarini, R. J. (2012). Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *37*(4), 851–864.

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews. Genetics*. https://doi.org/10.1038/s41576-019-0137-z

Verboom, C. E., Sentse, M., Sijtsema, J. J., Nolen, N. A., Ormel, J., & Penninx, B. W. J. H. (2011). Explaining heterogeneity in disability with major depressive disorder: Effects of personal and environmental characteristics. *Journal of Affective Disorders*, *132*, 71–81.

Vittengl, J. R., Jarrett, R. B., Weitz, E., Hollon, S. D., Twisk, J., Cristea, I., … Cuijpers, P. (2016). Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *The American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2015.15040492

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *The British Journal of Clinical Psychology / the British Psychological Society*, *60*(1), 1–37.

Wallert, J., Boberg, J., Kaldo, V., Mataix-Cols, D., Flygare, O., Crowley, J. J., … Rück, C. (2022). Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Translational Psychiatry, 12*(1), 1-10. https://doi.org/10.1038/s41398-022-02133-3

Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M., & Eley, T. C. (2014). The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. *JAMA Psychiatry* , *71*(8), 905–916.

Webb, C. A., Rosso, I. M., & Rauch, S. L. (2017). Internet-Based Cognitive-Behavioral Therapy for Depression: Current Progress and Future Directions. *Harvard Review of Psychiatry*, *25*(3), 114–122.

Wittchen, H. U., Zhao, S., Kessler, R. C., & Eaton, W. W. (1994). DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *51*(5), 355–364.

Witt, S. H., Streit, F., Jungkunz, M., Frank, J., Awasthi, S., Reinbold, C. S., … Rietschel, M. (2017). Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Translational Psychiatry*. https://doi.org/10.1038/tp.2017.115

Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2021). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* , *78*(1), 101–109.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82.

Yang, J., Wray, N. R., & Visscher, P. M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology*, *34*(3), 254–257.

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*(9), 1304–1310.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*, *163*(1), 148–150.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Boerescu, D., & Attiullah, N. (2008). Remission in depressed outpatients: more than just symptom resolution? *Journal of Psychiatric Research*, *42*(10), 797–801.

Chapter 2: Self-reported medication use as an alternate phenotyping method for anxiety and depression in the UK Biobank

This chapter is presented as a published paper. It is an exact copy of the peer-reviewed publication. Supplementary materials for this chapter, as detailed in the text, are included in Appendix A.

Citation: **Skelton, M.**, Rayner, C., Purves, K. L., Coleman, J. R., Gaspar, H. A., Glanville, K. P., ... & Eley, T. C. (2021). Self-reported medication use as an alternate phenotyping method for anxiety and depression in the UK Biobank. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *186*(6), 389-398. https://doi.org/10.1002/ajmg.b.32878

Received: 9 December 2020 | Revised: 3 June 2021 | Accepted: 21 September 2021

DOI: 10.1002/ajmg.b.32878

RESEARCH ARTICLE

American Journal of Medical Genetics Part B: Neuropsychiatric Genetics | WILEY

Self-reported medication use as an alternate phenotyping method for anxiety and depression in the UK Biobank

Megan Skelton[1,2] | Christopher Rayner[1] | Kirstin L. Purves[1] | Jonathan R. I. Coleman[1,2] | Héléna A. Gaspar[1] | Kylie P. Glanville[1] | Avina K. Hunjan[1,2] | Christopher Hübel[1,3,4] | Gerome Breen[1,2] | Thalia C. Eley[1,2]

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, University of London, London, UK

[2]National Institute for Health Research (NIHR) Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK

[3]National Centre for Register-based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

[4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Correspondence
Thalia C. Eley, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK.
Email: thalia.eley@kcl.ac.uk

Funding information
Guy's and St Thomas' Charity, Grant/Award Number: TR130505; Maudsley Charity, Grant/Award Number: 980; Medical Research Council, Grant/Award Numbers: MR/M021475/1, MR/V012878/1

Abstract

The requirement for large sample sizes for psychiatric genetic analyses necessitates novel approaches to derive cases. Anxiety and depression show substantial genetic overlap and share pharmacological treatments. Data on prescribed medication could be effective for inferring case status when other indicators of mental health are unavailable. We investigated self-reported current medication use in UK Biobank participants of European ancestry. Medication Status cases reported using antidepressant or anxiolytic medication (n = 22,218), controls did not report psychotropic medication use (n = 168,959). A subset, "Medication Only," additionally did not meet criteria for any other mental health indicator (case n = 2,643, control n = 107,029). We assessed genetic overlap between these phenotypes and two published genetic association studies of anxiety and depression, and an internalizing disorder trait derived from symptom-based questionnaires in UK Biobank. Genetic correlations between Medication Status and the three anxiety and depression phenotypes were significant (r_g = 0.60–0.73). In the Medication Only subset, the genetic correlation with depression was significant (r_g = 0.51). The three polygenic scores explained 0.33% – 0.80% of the variance in Medication Status and 0.07% – 0.19% of the variance in Medication Only. This study provides evidence that self-reported current medication use offers an alternate or supplementary anxiety or depression phenotype in genetic studies where diagnostic information is sparse or unavailable.

KEYWORDS
genetic correlation, internalizing, polygenic score, UK Biobank

1 | INTRODUCTION

Anxiety and depression are commonly comorbid (Kendler, Neale, Kessler, Heath, & Eaves, 1992; Moffitt et al., 2007), have notable symptom overlap (American Psychiatric Association, 2013), and share first-line pharmaceutical treatments (National Institute of Health and Care Excellence, 2011). Anxiety and depression are heritable (twin h^2: 20–60% (Hettema, Neale, & Kendler, 2001; McIntosh, Sullivan, & Lewis, 2019; Meier & Deckert, 2019; Sullivan, Neale, & Kendler, 2000), SNP-based h^2: 10–28% (Hettema et al., 2001; McIntosh et al., 2019;

Meier & Deckert, 2019; Sullivan et al., 2000). They also share much of their underlying genetic influences (twin r_g: ~1.00 (Kendler et al., 1992; Purves et al., 2020; Roy, Neale, Pedersen, Mathé, & Kendler, 1995), SNP-based r_g: ~0.8 (Kendler et al., 1992; Purves et al., 2020; Roy et al., 1995). Decades of work have demonstrated that liability to anxiety and depression is influenced by numerous individual genetic variants, each associated with a very small effect. To detect these effects, genetic association studies of complex disorders require extremely large sample sizes. The difficulty of ensuring adequate statistical power is compounded by the clinical heterogeneity of internalizing disorders, and their high lifetime prevalence whereby the mean difference in phenotypic liability between cases and controls is smaller than for a rare disorder such as schizophrenia (Mullins & Lewis, 2017; Wray et al., 2018).

One approach to increase statistical power in genetic association studies is to focus participant ascertainment on more clinically and demographically homogeneous individuals (Cai et al., 2015). However, collecting high-quality diagnostic information such as that from structured clinical interviews is resource-intensive and thus data are rarely available on the required scale. Another approach is to use less in-depth phenotyping methods to produce very large samples which can outweigh the noise introduced by phenotypic heterogeneity. For example, brief self-report symptom-based questionnaires with scoring criteria can indicate current or lifetime diagnoses in genetic studies (Coleman et al., 2020; Direk et al., 2017; Purves et al., 2020). Medical records of diagnoses, admissions, and treatments have also shown utility (Hall et al., 2018; Howard et al., 2018). Another method is "broad" or "minimal" phenotyping, whereby case status is determined using single data points, such as self-report of having received a diagnosis from a clinician (Howard et al., 2019; Hyde et al., 2016; Purves et al., 2020; Wray et al., 2018). Broad depression phenotypes demonstrate significant genetic correlations (r_g = 0.64–0.79) with clinically defined major depressive disorder (Howard et al., 2018). The availability of multiple phenotyping methods presents the opportunity for combination and triangulation, which can maximize the likelihood of reliably identifying cases (Glanville et al., 2021). Combining indicators of mental health status is also useful in the presence of missing data on more in-depth measures and in meta-analyses of studies that used different phenotyping methods. This has facilitated the identification of numerous novel genetic variants associated with anxiety and depression (Howard et al., 2019; Purves et al., 2020).

An additional indicator of anxiety or depression case status that has not yet been thoroughly investigated for genetic studies is reported use of antidepressant or anxiolytic medication. Medication data could supplement existing symptom-based and broad phenotyping measures by identifying probable cases who are not otherwise captured, perhaps due to missing data or low sensitivity to specific diagnoses. In support of this, 11% of the UK Biobank participants who responded to a follow-up mental health questionnaire reported using antidepressant medication but did not meet case criteria for anxiety or depression in the symptom questionnaires, hospital records or self-reported diagnoses (Davis et al., 2019). As well as supplementing other mental health indicators, medication data could function as a proxy for anxiety and depression when alternate sources of information are unavailable due to study design. For example, in cohort studies where mental health was not of primary interest during study development, records of reported medication use could be used to infer a diagnosis. It may also be a practical phenotyping solution for studies using big data from electronic health records. There is some support for the use of medication data as an alternate phenotype in genetic analyses. One study reported an odds ratio of 1.7 for taking antidepressants in individuals in the top decile of a depression polygenic score, compared with the lowest decile (Wu et al., 2019), and another found high genetic correlations between depression and self-reported antidepressant use in common genetic variants (SNP-based $r_g = 1$) and pedigree associated variants (rare and common variants; kinship-based $r_g = 0.9$) (Wigmore et al., 2019). While these previous studies are indicative of the potential for antidepressants as a proxy phenotype for depression, they did not investigate whether there could be a gain in sample size from using medication data in addition to other commonly used phenotypes. Furthermore, it is unclear whether the same is true for anxiolytics and anxiety. It is pertinent to note that the same class of drugs, selective serotonin reuptake inhibitors, is prescribed as a first-line pharmaceutical treatment for both depression and anxiety, despite being categorized as antidepressants (National Institute of Health and Care Excellence, 2011). By combining reports of antidepressants and anxiolytics, we can create a single internalizing disorder phenotype and examine overlap with existing anxiety and depression phenotypes. The genetic overlap between anxiety and depression supports the study of these two disorders as a combined phenotype (Hettema, 2008), indeed, this has been shown to increase agreement between self-report and symptom-based indicators of mental health status (Davis et al., 2019). While knowledge of the genetic influences on specific phenotypes is required to investigate how symptoms and diagnostic subtypes are related, there are advantages to studying psychiatric genetics at the broader level of internalizing disorders (Slade & Watson, 2006). One benefit could be the identification of genetic influences on factors of general distress or negative affect that act transdiagnostically and are clear therapeutic targets (Clark & Watson, 1991). This is particularly relevant in the context of mixed anxiety and depressive disorder, characterized by the presence of symptoms of anxiety and of depression that do not meet diagnostic criteria for either diagnosis. Although a controversial diagnosis, mixed anxiety and depressive disorder is highly prevalent in primary care and is associated with pronounced distress and impairment (Möller et al., 2016).

1.1 | Current study

We used data on current medication use reported by participants from the UK Biobank population cohort in an interview with a nurse. This is hereafter referred to as "Medication Status." Our aim was to explore use of antidepressant and anxiolytic drugs as an alternate phenotyping method to identify anxiety and depression cases where data about diagnoses or symptoms are absent. To test the utility of our Medication Status phenotype, we determined the extent to which common genetic variants associated with Medication Status overlap with those of anxiety and depression cases identified using other measures.

We further tested this question using multiple indicators of mental health status available in the UK Biobank. Specifically, we identified a subset of Medication Status cases who did not meet case status on any other indicator and were thus referred to as "Medication Only." This subset might represent individuals who were experiencing internalizing disorders but have not previously been identified with other commonly used methods in genetic studies of anxiety and depression. If so, we would expect this group to confer reliable genetic signal similar to that of anxiety and depression cases phenotyped using other measures, demonstrated by high genetic overlap. This would indicate that leveraging medication data alongside symptom-based and diagnostic measures can provide a useful gain in sample size, even in samples well characterized for psychiatric disorders. Alternatively, Medication Only cases could be a particularly heterogeneous group of individuals, a substantial proportion of whom were prescribed medication for reasons other than anxiety or depression diagnoses. We would then expect this group to show low genetic overlap with internalizing disorder cases. As such, caution would be required using this phenotyping approach; the heterogeneity would reduce statistical power, and any genetic signal identified could be nonspecific to internalizing disorders.

2 | MATERIALS AND METHODS

2.1 | Sample

The UK Biobank is a health research resource of more than 500,000 volunteers from the UK population, aged between approximately 40 and 70 years old at recruitment. Between 2006 and 2010, participants provided biological samples and responded to a range of health and lifestyle questionnaires, including an interview with a nurse about medication use. In each of two follow-up visits, approximately 20,000 participants completed additional measures and could update information from their initial session (Sudlow et al., 2015), resulting in up to three measurement time points. An online Mental Health Questionnaire (MHQ), which asked more detailed information about psychiatric diagnoses and symptoms, was completed by almost one third of the cohort in 2016 (Davis et al., 2020). In the current study, we limited our analyses to participants of European ancestry who passed genetic quality control and had complete data on the phenotypes and covariates.

2.2 | Phenotype definitions

2.2.1 | Medication Status

Medication Status was obtained via a verbal interview item requesting the names of regular prescription medications that the participants were currently taking (UK Biobank Data Field 20003). The nurses conducting the interviews did not record medications that were short-term (e.g., 1-week course of antibiotics), historical, or prescribed but not being taken. We excluded individuals who responded to the MHQ to prevent sample overlap in polygenic comparisons with selected phenotypes that had been derived using MHQ responders (further details in Section 2.4). The sample for analysis consisted of individuals who supplied the name of any medication in the interview, at any of the three data collection time points. The medication interview field had 6,745 unique response values, which were mapped to psychotropic medication classes of antidepressant, anxiolytic, antipsychotic, and mood stabilizer using an existing list of codes (Davis et al., 2019; see Data S1, Supporting Information). Medication Status cases were individuals who reported taking a medication at the time of interview that was classed as an anxiolytic or antidepressant. Note that participants themselves did not need to report, or even know, that the medication was an antidepressant or anxiolytic. This variable thus had the potential to capture individuals who were prescribed medication for reasons other than core symptoms of anxiety or depression, such as sleep difficulties. Controls were defined as individuals who did not report taking any psychotropic medication but who reported taking at least one medication and therefore had a nonblank response (as per a previous UK Biobank study; Wu et al., 2019). As there was not a negative option for this item, we chose not to infer control status from a blank response. Although some blank responses would represent controls, we could not distinguish these from individuals who did not wish to share the information (i.e., possible cases or exclusions). There was a preceding binary question regarding the use of any medications, which could have helped to identify controls, but it had low agreement with the more detailed item in which participants named a medication. Case and control definitions and screening were performed solely using the medication data, so that the same sample would be drawn if only these data were available. We excluded individuals who reported using antipsychotics or mood stabilizers indicative of potential psychosis or bipolar disorder. These disorders are highly heritable and demonstrate strong genetic correlations with other psychiatric disorders, thus potentially producing misleading relationships if included in our analyses. It is worth noting that this may have excluded some of the more severe depression cases who were receiving adjunct antipsychotics. This could lead to decreased observed genetic overlap with other internalizing phenotypes, as more severe symptoms are associated with greater genetic liability (Wray et al., 2018); however, we chose to be conservative.

We also identified a subset of the Medication Status cases and controls for analysis who we will refer to as "Medication Only." Whereas Medication Status cases and controls were defined solely using the interview item regarding medication use, Medication Only incorporated information from other indicators of mental health. Cases were therefore defined in the same way as for Medication Status but additionally did not meet case criteria (i.e., were controls or missing sufficient information) for anxiety, depression, or other psychiatric conditions, on any other indicator of mental health status in the UK Biobank data. These diagnostic indicators included hospital records (primary and secondary International Classification of Diseases [ICD] 10 "F" codes), self-reported diagnosis in a nurse interview, and treatment seeking (see Data S2 for details and Tables S1 and S2 for data completeness and case overlap between indicators). Medication Only cases therefore represented a group of individuals who, across all these indicators of mental health, were only identified as anxiety or depression cases by their reported

medication use. Medication Only controls were participants who were not taking an antidepressant or anxiolytic (defined as per Medication Status controls) and furthermore did not meet case criteria for any other indicator of mental health status.

Medications can be prescribed for conditions other than those for which they are licensed. Therefore, we created versions of the Medication Status and Medication Only phenotypes excluding individuals with evidence of physical conditions for which antidepressants or anxiolytics are commonly prescribed (see Data S3 and Table S3 for details, including sample sizes).

2.2.2 | Lifetime Internalizing Disorder

To provide data for comparison with Medication Status, we used the anxiety and depression modules of the Composite International Diagnostic Interview Short Form (CIDI-SF) from the MHQ to derive a symptom-based case–control phenotype. This is closest to a "gold-standard" diagnostic approach available in UK Biobank (Cai et al., 2020). To ensure a comparable internalizing disorder phenotype, cases met criteria for either or both of the lifetime anxiety and lifetime depression modules and controls did not fulfill case criteria for either. We refer to this as the "Lifetime Internalizing" disorder diagnosis (see Data S2 for further details).

Phenotype derivation and descriptive analyses were performed in R version 3.6.0 (R Core Team, 2019).

2.3 | Genotyping and quality control

Genetic data were centrally processed as per the UK Biobank pipeline (Bycroft et al., 2018) and variants were limited to those genotyped or imputed (INFO score > 0.4 (Zheng et al., 2015) to the Haplotype Reference Consortium reference panel and the UK10K Consortium reference panel. Individuals of European ancestry were identified using four-means clustering on the first two genomic principal components available from UK Biobank. Single nucleotide polymorphisms (SNPs) were included in the analyses if they were common (minor allele frequency > 0.01), had low missingness (≤ 0.02), and a nonsignificant Hardy–Weinberg equilibrium test result ($p > 1 \times 10^{-8}$). Individuals were unrelated (more distant than third-degree relatives; KING < 0.044 as reported by UK Biobank), had low SNP missingness (≤ 0.02) and concordant chromosomal and phenotypic sex. SNPs were further filtered to INFO > 0.9 for analyses conducted with linkage disequilibrium score regression (LDSC; Bulik-Sullivan et al., 2015) and PRSice-2 (Choi & O'Reilly, 2019).

2.4 | Statistical analysis

2.5 | Genome-wide association

A genome-wide association study (GWAS) was performed for Medication Status in the full sample and in the Medication Only subset, as well as for the Lifetime Internalizing phenotype, using BGENIE software (version 1.2; Bycroft et al., 2018). GWAS were a necessary step to perform the main investigation of genetic overlap between self-reported medication data and other anxiety and depression phenotypes; see Data S4 and Figures S1 and S2 for GWAS details and results.

To determine how self-reported medication data compared to existing definitions of anxiety and depression, we selected two published genetic studies in addition to the UK Biobank Lifetime Internalizing phenotype. These were a lifetime anxiety GWAS (Purves et al., 2020) ("UKB-anxiety") and the Psychiatric Genomics Consortium's second major depressive disorder GWAS (Wray et al., 2018) ("PGC-depression"); see Table 1 for details. Of note, the medication phenotypes had been created in UK Biobank participants who did not complete the MHQ. UKB-anxiety and Lifetime Internalizing phenotypes were derived exclusively from MHQ responders, and we selected summary statistics from a GWAS of the PGC-depression data which excluded the UK Biobank sample. As such, known sample overlap was eliminated.

2.6 | Heritability and genetic correlations

The proportion of phenotypic variance explained by common genetic variants, defined as SNP-based heritability, was estimated using LDSC (Bulik-Sullivan, Loh, et al., 2015) of the GWAS summary statistics. For conversion to the liability scale, we assumed that the sample prevalence represented the true population prevalence, with calculations using ±10% of this value also performed.

We estimated genetic correlations using LDSC (Bulik-Sullivan et al., 2015; Bulik-Sullivan, Loh, et al., 2015). These were calculated between our medication phenotypes with each of UKB-anxiety (Purves et al., 2020), PGC-depression (Wray et al., 2018), and the Lifetime Internalizing phenotype. We used block jackknifing to determine if the correlations between Medication Status and the three comparison phenotypes were significantly different to those estimated with Medication Only. This method divides the summary statistics into a number of blocks (default in LDSC is 200) of contiguous SNPs and performs multiple estimations of the genetic correlation, omitting one block at a time. This provides *SE*s with which to determine statistically significant differences between pairs of genetic correlations.

2.7 | Polygenic scores

Genetic overlap between self-reported medication use and previously defined anxiety and depression was also assessed via polygenic scoring. GWAS summary statistics from UKB-anxiety (Purves et al., 2020), PGC-depression (Wray et al., 2018), and Lifetime Internalizing were used to compute polygenic scores for each participant in our sample. An individual's polygenic score is the sum of trait-associated variants they carry, weighted by the GWAS effect size. Polygenic scores were calculated across a range of *p*-value thresholds, with increasingly lenient thresholds capturing a greater number of the variants that

TABLE 1 Summary statistics from genome-wide association studies (GWAS) selected for comparison with self-reported current antidepressant or anxiolytic medication use in the UK Biobank (European ancestry, excluding Mental Health Questionnaire (MHQ) responders)

GWAS	Sample	*N* cases	*N* controls	Phenotyping methods
UKB-anxiety (Purves et al., 2020)	UK Biobank MHQ responders	25,453	58,113	Lifetime anxiety symptom-based questionnaire Self-report of anxiety diagnosis from a clinician
PGC-depression (Wray et al., 2018)	Meta-analysis, excluding UK Biobank	116,404	314,990	Various, including: Structured diagnostic interviews Self-report of depression diagnosis from a clinician Depression symptom-based questionnaire Depression diagnoses in medical records
Lifetime Internalizing	UK Biobank MHQ responders in the present study	32,160	91,732	Lifetime anxiety and depression symptom-based questionnaires

Note: To eliminate known overlap between the analytical sample and the comparison GWAS, MHQ responders were excluded from the medication phenotypes, and we selected summary statistics from a GWAS of PGC-depression which did not include UK Biobank data.

were tested in the GWAS. For each of the three trait polygenic scores, logistic regressions were performed with Medication Status and with Medication Only. Age, sex, genotyping batch, assessment center, and the first six genetic principal components were included as covariates. We previously determined that this number of principal components, alongside assessment center, is sufficient to control for the majority of genome-wide inflation associated with geographic location in UK Biobank. The optimum *p*-value threshold was defined as the one where the proportion of variance (Nagelkerke's R^2) explained in the medication phenotype was highest, as this is likely to represent the maximal amount of true variance explained (in addition to noise). Nagelkerke's R^2 is subject to bias when the sample prevalence does not reflect the population prevalence and therefore the variance explained was converted to the liability scale using a range of prevalence values (Lee, Goddard, Wray, & Visscher, 2012). To gain an empirical *p*-value, 10,000 permutations were performed. Polygenic score analyses were performed in PRSice version 2.3.1 (Choi & O'Reilly, 2019).

2.8 | Sensitivity analyses

The Medication Status and Medication Only GWAS and subsequent analyses were repeated excluding individuals with evidence of physical conditions, such as epilepsy and chronic pain, for which antidepressants or anxiolytics are commonly prescribed (see Data S3 and Table S3 for further details).

2.9 | Ethics

UK Biobank has Research Ethics Committee approval (11/NW/0382) and Research Tissue Bank approval. Participants provided written informed consent which included permission to access their medical records. The current study was performed under UK Biobank application 18177.

3 | RESULTS

3.1 | Sample and phenotype distribution

The genetic quality-control criteria resulted in a sample of 385,645 UK Biobank participants of European ancestry. Of these, 283,662 individuals supplied the name of any medication in response to the Medication Status verbal interview item. After excluding participants who completed the MHQ ($n = 89{,}801$) and who reported taking antipsychotic or mood stabilizer medication ($n = 2{,}684$), 191,177 participants remained for analysis. Of the 22,218 Medication Status cases, 20,399 (92%) reported taking an antidepressant and 3,047 (14%) reported taking an anxiolytic, with 1,228 (6%) reporting taking both. This left 168,959 controls who had responded to the medication interview item but had not named an antidepressant, anxiolytic, antipsychotic, or mood stabilizer. Following exclusion of individuals who met case criteria for any other indicator of mental health status, there were 2,643 Medication Only cases and 107,029 controls. Of the Medication Only cases, 2,083 (79%) reported taking an antidepressant drug, 618 (23%), an anxiolytic, and 58 (2%) reported taking both.

The majority of Medication Status cases were female (68%), with a mean age of 57 years ($SD = 7.8$), while 53% of controls were female, and had a mean age of 58 years ($SD = 8.0$). Medication Only cases were also predominantly female (67%), with a mean age of 60 years ($SD = 7.3$), and 49% of controls were female, mean age 58 years ($SD = 8.0$). Further descriptives about these groups and the Lifetime Internalizing phenotype are available in Table S4. Details of the overlap between these groups and other mental health indicators in UK Biobank are available in Tables S1 and S2.

3.2 | Genetic correlations

Results from genetic correlations are displayed in Figure 1. These were performed for Medication Status and the Medication Only subset with the comparison phenotypes; UKB-anxiety, PGC-depression,

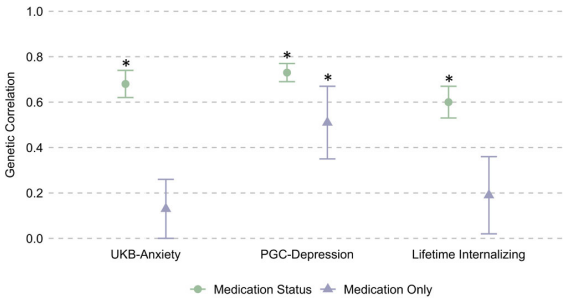


FIGURE 1 Genetic correlations between self-reported current antidepressant or anxiolytic medication use in the UK Biobank (European ancestry, excluding Mental Health Questionnaire (MHQ) responders) and anxiety and depression phenotyped using existing methods. Medication Status was defined using self-reported current use of antidepressant or anxiolytic medication in individuals who did not complete the MHQ ($N = 191{,}177$, cases = 22,218). Medication Only were a subset who additionally did not meet case criteria for any other indicators of mental health condition in UK Biobank ($N = 109{,}672$, cases = 2,643). UKB-anxiety is the Purves et al. (2020) study, which used data from the UK Biobank MHQ. PGC-depression is Wray et al. (2018), an international meta-analysis, excluding the UK Biobank sample. Lifetime Internalizing was created in the present study using lifetime anxiety and depression symptom-based questionnaires in UK Biobank MHQ responders. Standard errors are represented by error bars. * Indicates significance at the Bonferroni-adjusted significance level of 8.3×10^{-3} used for the six independent tests performed. Liability scale SNP-based h^2 assuming sample prevalence equals population prevalence: Medication Status 0.074 ($SE = 0.004$), Medication Only 0.053 ($SE = 0.010$)

and Lifetime Internalizing. All genetic correlations for Medication Status were significant, while for Medication Only the SEs were large; only the correlation with PGC-depression was significant.

The block jackknifing revealed that the correlation between Medication Status and UKB-anxiety significantly differed to the correlation between Medication Only and UKB-anxiety ($p = 2.0 \times 10^{-5}$), using a Bonferroni-adjusted significance level of 0.016 for three tests. A significant difference between the correlations was also found when Lifetime Internalizing was the comparison phenotype ($p = 0.012$). However, the difference between the genetic correlation for Medication Status and PGC-depression and the correlation for Medication Only and PGC-depression was not significantly different from zero ($p = 0.071$).

3.3 | Polygenic scores

The proportion of variance explained in the Medication Status and Medication Only phenotypes by polygenic scores created from the UKB-anxiety and PGC-depression studies, as well as the Lifetime Internalizing phenotype, are displayed in Table 2. For both medication use phenotypes, the PGC-depression polygenic score explained the greatest proportion of variance. For bar plots showing the proportion of variance explained across each *p*-value threshold tested, see Figure S3.

TABLE 2 Proportion of variance (Nagelkerke's R^2) in self-reported current antidepressant or anxiolytic medication use in the UK Biobank (European ancestry, excluding Mental Health Questionnaire (MHQ) responders) explained by polygenic scores of anxiety and depression

Discovery sample	Medication status	Medication only
UKB-anxiety (Purves et al., 2020, UK Biobank MHQ responders)	0.41% ($p = 9.99 \times 10^{-5}$[a], $p_T = 1$; 200,935 SNPs)	0.07% ($p = 7.10 \times 10^{-3}$[a], $p_T = 0.1$; 35,404 SNPs)
PGC-depression (Wray et al., 2018, excluding UK Biobank)	0.80% ($p = 9.99 \times 10^{-5}$[a], $p_T = 0.3$; 49,513 SNPs)	0.19% ($p = 9.99 \times 10^{-5}$[a], $p_T = 0.05$; 14,288 SNPs)
Lifetime Internalizing (Lifetime symptom-based questionnaires, UK Biobank MHQ responders in the present study)	0.33% ($p = 9.99 \times 10^{-5}$[a], $p_T = 1$; 201,839 SNPs)	0.09% ($p = 2.30 \times 10^{-3}$[a], $p_T = 0.5$; 129,634 SNPs)

Note: Proportion explained is presented on the liability scale, assuming population prevalence is equal to sample prevalence. Medication Status was defined using self-reported current use of antidepressant or anxiolytic medication in individuals who did not complete the MHQ (to provide independent discovery and target samples) ($N = 191{,}177$, cases = 22,218). Medication Only were a subset who additionally did not meet case criteria for any other indicators of mental health condition in UK Biobank ($N = 109{,}672$, cases = 2,643). *p* = empirical *p*-value resulting from 10,000 permutations.
Abbreviations: p_T = *p*-value threshold at which Nagelkerke's R^2 was highest; SNPs, single nucleotide polymorphisms.
[a]Significant at the Bonferroni-adjusted significance level of 8.3×10^{-3} used for the six independent tests performed.

3.4 | Sensitivity analyses

The sensitivity analysis excluded individuals with evidence of physical conditions that are commonly treated with antidepressants or anxiolytics. Of the Medication Status cases, 4% had evidence of one of these physical conditions, and 3.5% of the Medication Only cases (see Table S3 for sample size details). The sensitivity analysis for Medication Status and for Medication Only did not substantially change the results for the genetic correlations or polygenic scores (see Data S5, Figure S4, and Tables S5 and S6).

4 | DISCUSSION

4.1 | Overview

Using genetic correlations and polygenic score analysis, this study demonstrated that self-reported, current antidepressant, and

anxiolytic medication use can serve as an alternate phenotyping method for anxiety or depression. A substantial proportion of the UK Biobank European ancestry sample for genetic analysis had answered the medication verbal interview item (73%) and therefore had data to inform our Medication Status phenotype. We observed genetic overlap between Medication Status and anxiety and depression phenotypes (UKB-anxiety, PGC-depression, and UK Biobank Lifetime Internalizing) that were defined using other methods such as self-report of a diagnosis, symptom-based questionnaires, and clinical interviews.

The existence of multiple indicators of anxiety and depression in the UK Biobank allowed us to further explore the utility of this alternate phenotyping method for identifying anxiety and depression cases. The majority of Medication Status cases were the same individuals that would be identified if we had used other UK Biobank measures, such as reported treatment seeking or hospital codes. However, we did identify approximately 10% of the Medication Status cases who reported antidepressant or anxiolytic use but did not meet case criteria for any of the other indicators of a mental health condition. We named this subset "Medication Only." This group appeared to be more heterogeneous, demonstrating lower genetic overlap with the comparison phenotypes.

4.2 | Genetic correlations

The SNP-based genetic correlations between Medication Status and UKB-anxiety (Purves et al., 2020), PGC-depression (Wray et al., 2018), and Lifetime Internalizing ranged between 0.60 and 0.73. This indicates substantial common variant overlap between individuals who reported taking antidepressant or anxiolytic medication and participants who were identified using other indicators, such as symptom-based questionnaires or clinical interviews.

The correlations were comparable to those from the PGC-depression study, which reported a weighted mean genetic correlation of 0.76 between the seven contributing cohorts (Wray et al., 2018), and also to those reported for depression phenotypes from multiple indicators in a previous UK Biobank study (0.85–0.87; Howard et al., 2018). This is despite the potential presence of individuals with a broader range of psychiatric disorders in the medication phenotype than these depression studies. Antidepressant medication is used to treat generalized anxiety disorder as well as fear disorders, such as agoraphobia, but fear disorders are less genetically similar to depression than generalized anxiety disorder is (Mineka, Watson, & Clark, 1998; Morneau-Vaillancourt et al., 2020). The inclusion of individuals with fear disorders could therefore result in lower genetic correlations with the comparison depression phenotype, but also with UKB-anxiety and Lifetime Internalizing, as the anxiety module of the CIDI-SF questionnaire is most sensitive to generalized anxiety disorder.

We performed analyses in the Medication Only subset to further assess the utility and pitfalls of this approach to identifying cases. An issue we encountered was that, while representing 10% of the Medication Status cases, the Medication Only case group was just approximately 3,000 individuals, which meant that statistical power was attenuated for genetic correlations. The genetic correlation with PGC-depression in this subsample was lower than in the full analysis sample and had a large *SE* ($r_g = 0.51 \pm 0.16$). The block jackknife analysis revealed that this correlation was not significantly different compared to the Medication Status correlation with PGC-depression. The Medication Only genetic correlations with UKB-anxiety ($r_g = 0.13 \pm 0.13$) and with Lifetime Internalizing ($r_g = 0.19 \pm 0.17$) were not statistically significant. The block jackknife demonstrated that these correlations were significantly lower than the respective correlations with the Medication Status phenotype. The significant correlation with the depression phenotype but not anxiety and internalizing phenotypes could be due to the design of the PGC-depression study, which was a large meta-analysis of numerous cohorts. As such, it may capture a more heterogeneous, negative affect phenotype, than the UK Biobank MHQ derived anxiety and internalizing phenotypes. The higher, significant, genetic correlation may also be explained by the majority of the Medication Only sample taking medications for depression (79% reported an antidepressant, 23% an anxiolytic). However, the use of antidepressants for anxiety disorders renders this unknown. Due to the lack of power, we cannot be certain whether Medication Only cases represent previously unidentified true cases, or a group with noise introduced by heterogeneity due to individuals taking medications for other reasons.

The results of the sensitivity analyses excluding physical conditions known to be treated with antidepressants and anxiolytics were largely the same, with fewer than 4% of cases removed from Medication Status and Medication Only phenotypes. However, it is likely that some individuals' physical conditions were not recorded in the dataset.

4.3 | Polygenic scores

To further explore the genetic overlap of our self-reported medication use phenotypes with the other definitions of anxiety and depression, particularly in the Medication Only subset, we tested for polygenic associations. Polygenic scores created from the UKB-anxiety, PGC-depression, and Lifetime Internalizing GWAS significantly explained 0.33–0.80% of variance (Nagelkerke's R^2) in Medication Status in the full sample and 0.07–0.19% of variance (Nagelkerke's R^2) in the Medication Only subset.

Polygenic scores rarely account for large proportions of variance in psychiatric phenotypes. For example, with over 115,000 cases, the PGC-depression study reported that 1.9% of the variance in case status was explained by the polygenic score out-of-sample (Wray et al., 2018), and the UKB-anxiety polygenic score explained 0.5% of the variance using a within-sample leave-one-out approach (Purves et al., 2020). Our results therefore provide further evidence that to some extent Medication Status functioned as an alternate phenotype for anxiety or depression case status, although evidence was stronger for the latter. As well as previous reasons discussed, such as the

potential presence of other anxiety subtypes in the medication phenotype, this may be because the discovery sample for the depression polygenic score was larger and better powered than the anxiety discovery GWAS. It is worth noting that polygenic scoring methods are constantly advancing. As such, if researchers were to use this phenotyping approach in a predictive polygenic score analysis they might see predictive gains, in comparison to what we have reported here, by using alternative methods (Pain et al., 2021).

4.4 | Limitations

There are several limitations to the current study. Primarily, the verbal interview question from which the medication phenotypes were created was self-reported and included only current medication being taken. Medication Status cases and controls were defined solely using this item, and as such individuals with historical use of medications for anxiety or depression are likely to be included as controls, reducing the power of the GWAS. Furthermore, only a fraction of those with mental health conditions seek and receive treatment (McManus, Bebbington, Jenkins, & Brugha, 2016; Rayner et al., 2020) and other, nonpharmacological treatments are available. Similar to other methods of phenotyping mental health conditions, perceived stigma could also impact the reliability of this self-report data, assuming that the individual is aware that their medication is primarily prescribed for anxiety or depression. Recall bias is an additional concern, although somewhat mitigated as only current prescriptions were requested. However, these issues appear to be more relevant for some medications than others; self-report of antidepressants shows high agreement with medical records, whereas for mood stabilizers agreement is poor to moderate (Gnjidic, Du, Pearson, Hilmer, & Banks, 2017; Hafferty et al., 2018).

A further limitation is that the dataset did not contain data on the reason for prescription, or consistent information on duration or dosage with which to infer it, which would have enabled refinement of the sample (Wigmore et al., 2019). As mentioned previously, noise in the medication phenotype may have been introduced by including individuals who were prescribed medication for reasons other than anxiety or depression. This could result in the identification of nonspecific genetic variants in addition to variants associated with internalizing disorders, consistent with criticisms of the broad phenotyping approach (Cai et al., 2020). However, it is worth noting that high genetic correlations are often not observed even between samples that have similarly been assessed by a clinician using a DSM-based interview or checklist (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013).

UK Biobank participants significantly differ from the general population with regard to both socioeconomic status and long-term illnesses (Davis et al., 2020). The majority (85%) of UK Biobank participants were from urban areas (Fry et al., 2017) and geographical differences in prescription rates, due to factors such as the availability of psychological therapies, may have further impacted the utility of our phenotyping method. However, it has been illustrated that common mental health conditions in UK Biobank are reported at a similar prevalence as national surveys, suggesting that studies focusing on mental health have some generalizability (Davis et al., 2020).

4.5 | Future directions and conclusion

These findings suggest that self-reported current use of antidepressant and anxiolytic medications can offer a reasonable alternate approach to identifying clinical cases of anxiety and depression, where more detailed measures, such as diagnoses or questionnaires, are not available. However, the lower variance explained by the Medication Only group suggests that when other measures of anxiety or depression are present in a dataset, they should be preferentially used. The gain in sample size from using this additional information will not contribute a substantial amount of signal to analyses.

It is likely that prescription data from medical records will better approximate clinically diagnostic measures of anxiety and depression. As medical record linkage has recently become available in the UK Biobank cohort, this presents a future avenue of investigation. With increasing electronic health record data analysis, the use of medication data as an additional or alternate anxiety or depression phenotype could lead to unparalleled sample sizes. In conclusion, this study provides evidence that phenotyping anxiety and depression using medication data may be a useful and pragmatic approach when higher-quality diagnostic information is unavailable.

ACKNOWLEDGMENTS

This research was conducted using the UK Biobank Resource, under application number 18177. The authors would like to thank all individuals involved in UK Biobank, including the researchers, supporting organizations and of course all the participants who have made a vital contribution to improving health by sharing their experiences. T.C.E. and G.B. were part-funded during this work by program grants from the UK Medical Research Council (MR/M021475/1 and MR/V012878/1). C.R. is supported by a grant from Fondation Peters to T.C.E. and G.B. G. B. has received honoraria, research or conference grants and consulting fees from Illumina and Otsuka. The authors acknowledge use of the research computing facility at King's College London, Rosalind (https://rosalind.kcl.ac.uk), which was part-funded by capital equipment grants from Guy's and St Thomas' Charity (TR130505) and Maudsley Charity (Award 980). This study represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or King's College London.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

AUTHOR CONTRIBUTIONS

M.S., K.L.P., G.B., and T.C.E. were responsible for conceptualizing and designing the study. M.S. performed the data analysis with technical input from C.R., J.R.I.C., and K.L.P. K.P.G., J.R.I.C., C.H., and H.A.G. performed data preparation and quality control. M.S. wrote the original draft, and all other authors provided a critical review and edits to the draft, including interpretation of the results. All authors reviewed and approved the final manuscript prior to submission. T.C.E. and G.B. provided supervision of the study.

DATA AVAILABILITY STATEMENT

This research was conducted using the UK Biobank Resource, under application number 18177. UK Biobank data is available to researchers conducting health-related research, who register with UK Biobank.

ORCID

Megan Skelton https://orcid.org/0000-0002-7619-349X
Christopher Rayner https://orcid.org/0000-0003-2950-343X
Kirstin L. Purves https://orcid.org/0000-0002-8110-5554
Jonathan R. I. Coleman https://orcid.org/0000-0002-6759-0944
Héléna A. Gaspar https://orcid.org/0000-0002-5540-2707
Kylie P. Glanville https://orcid.org/0000-0001-8321-9435
Avina K. Hunjan https://orcid.org/0000-0002-5743-9597
Christopher Hübel https://orcid.org/0000-0002-1267-8287
Gerome Breen https://orcid.org/0000-0003-2053-1792
Thalia C. Eley https://orcid.org/0000-0001-6458-0700

REFERENCES

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

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P., … Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B., Loh, P., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., … Flint, J. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588–591. https://doi.org/10.1038/nature14659

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., … Levinson, D. F. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*(4), 437–447. https://doi.org/10.1038/s41588-020-0594-5

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*, *8*(7), giz082. https://doi.org/10.1093/gigascience/giz082

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336. https://doi.org/10.1037//0021-843x.100.3.316

Coleman, J. R. I., Peyrot, W. J., Purves, K. L., Davis, K. A. S., Rayner, C., Choi, S. W., … Breen, G. (2020). Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Molecular Psychiatry*, *25*(7), 1430–1446. https://doi.org/10.1038/s41380-019-0546-6

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., … International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994. https://doi.org/10.1038/ng.2711

Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., … Hotopf, M. (2020). Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: A reanalysis. *BJPsych Open*, *6*(2), e18. https://doi.org/10.1192/bjo.2019.100

Davis, K. A. S., Cullen, B., Adams, M., Brailean, A., Breen, G., Coleman, J. R. I., … Hotopf, M. (2019). Indicators of mental disorders in UK Biobank-A comparison of approaches. *International Journal of Methods in Psychiatric Research*, *28*(3), e1796. https://doi.org/10.1002/mpr.1796

Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., … Sullivan, P. F. (2017). An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biological Psychiatry*, *82*(5), 322–329. https://doi.org/10.1016/j.biopsych.2016.11.013

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, *186*(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Glanville, K. P., Coleman, J. R. I., Howard, D. M., Pain, O., Hanscombe, K. B., Jermy, B., … Lewis, C. M. (2021). Multiple measures of depression to enhance validity of Major Depressive Disorder in the UK Biobank. *BJPsych Open*, *7*(2), e44. https://doi.org/10.1192/bjo.2020.145

Gnjidic, D., Du, W., Pearson, S. A., Hilmer, S. N., & Banks, E. (2017). Ascertainment of self-reported prescription medication use compared with pharmaceutical claims data. *Public Health Research & Practice*, *27*(4), e27341702. https://doi.org/10.17061/phrp27341702

Hafferty, J. D., Campbell, A. I., Navrady, L. B., Adams, M. J., MacIntyre, D., Lawrie, S. M., … McIntosh, A. M. (2018). Self-reported medication use validated through record linkage to national prescribing data. *Journal of Clinical Epidemiology*, *94*, 132–142. https://doi.org/10.1016/j.jclinepi.2017.10.013

Hall, L. S., Adams, M. J., Arnau-Soler, A., Clarke, T.-K., Howard, D. M., Zeng, Y., … McIntosh, A. M. (2018). Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational Psychiatry*, *8*, 1–12. https://doi.org/10.1038/s41398-017-0034-1

Hettema, J. M. (2008). What is the genetic relationship between anxiety and depression? *American Journal of Medical Genetics Part C Seminars in Medical Genetics*, *148*(2), 140–146. https://doi.org/10.1002/ajmg.c.30171

Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American Journal of Psychiatry*, *158*(10), 1568–1578. https://doi.org/10.1176/appi.ajp.158.10.1568

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the

importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., … McIntosh, A. M. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature Communications*, *9*(1), 1470. https://doi.org/10.1038/s41467-018-03819-3

Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., … Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, *48*(9), 1031–1036. https://doi.org/10.1038/ng.3623

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of General Psychiatry*, *49*(9), 716–722. https://doi.org/10.1001/archpsyc.1992.01820090044008

Lee, S. H., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2012). A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*, *36*(3), 214–224. https://doi.org/10.1002/gepi.21614

McIntosh, A. M., Sullivan, P. F., & Lewis, C. M. (2019). Uncovering the genetic architecture of major depression. *Neuron*, *102*(1), 91–103. https://doi.org/10.1016/j.neuron.2019.03.022

McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Leeds, England: NHS Digital.

Meier, S. M., & Deckert, J. (2019). Genetics of anxiety disorders. *Current Psychiatry Reports*, *21*(3), 16. https://doi.org/10.1007/s11920-019-1002-7

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, *49*(1), 377–412. https://doi.org/10.1146/annurev.psych.49.1.377

Moffitt, T. E., Harrington, H., Caspi, A., Kim-Cohen, J., Goldberg, D., Gregory, A. M., & Poulton, R. (2007). Depression and generalized anxiety disorder: Cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years. *Archives of General Psychiatry*, *64*(6), 651–660. https://doi.org/10.1001/archpsyc.64.6.651

Möller, H. J., Bandelow, B., Volz, H. P., Barnikol, U. B., Seifritz, E., & Kasper, S. (2016). The relevance of 'mixed anxiety and depression' as a diagnostic category in clinical practice. *European Archives of Psychiatry and Clinical Neuroscience*, *266*(8), 725–736. https://doi.org/10.1007/s00406-016-0684-7

Morneau-Vaillancourt, G., Coleman, J. R. I., Purves, K. L., Cheesman, R., Rayner, C., Breen, G., & Eley, T. C. (2020). The genetic and environmental hierarchical structure of anxiety and depression in the UK Biobank. *Depression and Anxiety*, *37*(6), 512–520. https://doi.org/10.1002/da.22991

Mullins, N., & Lewis, C. M. (2017). Genetics of depression: Progress at last. *Current Psychiatry Reports*, *19*(8), 43. https://doi.org/10.1007/s11920-017-0803-9

National Institute of Health and Care Excellence. (2011). Common mental health problems: Identification and pathways to care. CG123. Retrieved from https://www.nice.org.uk/guidance/CG123/chapter/1-Guidance#steps-2-and-3-treatment-and-referral-for-treatment

Pain, O., Glanville, K. P., Hagenaars, S. P., Selzam, S., Fürtjes, A. E., Gaspar, H. A., … Lewis, C. M. (2021). Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genetics*, *17*(5), e1009021. https://doi.org/10.1371/journal.pgen.1009021

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry*, *25*(12), 3292–3303. https://doi.org/10.1038/s41380-019-0559-1

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.r-project.org/

Rayner, C., Coleman, J. R. I., Purves, K. L., Carr, E., Cheesman, R., Davies, M., … Eley, T. C. (2020). Sociodemographic inequalities in lifetime treatment seeking and receipt for common mental health problems. *PsyArXiv Preprints*. 1–18. https://doi.org/10.31234/osf.io/vsu4g

Roy, M. A., Neale, M. C., Pedersen, N. L., Mathé, A. A., & Kendler, K. S. (1995). A twin study of generalized anxiety disorder and major depression. *Psychological Medicine*, *25*(5), 1037–1049. https://doi.org/10.1017/s0033291700037533

Slade, T., & Watson, D. (2006). The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. *Psychological Medicine*, *36*(11), 1593–1600. https://doi.org/10.1017/S0033291706008452

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *The American Journal of Psychiatry*, *157*(10), 1552–1562. https://doi.org/10.1176/appi.ajp.157.10.1552

Wigmore, E. M., Hafferty, J. D., Hall, L. S., Howard, D. M., Clarke, T.-K., Fabbri, C., … McIntosh, A. M. (2019). Genome-wide association study of antidepressant treatment resistance in a population-based cohort using health service prescription data and meta-analysis with GENDEP. *The Pharmacogenomics Journal*, *20*(2), 329–341. https://doi.org/10.1038/s41397-019-0067-3

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Wu, Y., Byrne, E. M., Zheng, Z., Kemper, K. E., Yengo, L., Mallett, A. J., … Wray, N. R. (2019). Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nature Communications*, *10*(1), 1891. https://doi.org/10.1038/s41467-019-09572-5

Zheng, H.-F., Rong, J.-J., Liu, M., Han, F., Zhang, X.-W., Richards, J. B., & Wang, L. (2015). Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. *PLoS One*, *10*(1), e0116487. https://doi.org/10.1371/journal.pone.0116487

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Skelton, M., Rayner, C., Purves, K. L., Coleman, J. R. I., Gaspar, H. A., Glanville, K. P., Hunjan, A. K., Hübel, C., Breen, G., & Eley, T. C. (2021). Self-reported medication use as an alternate phenotyping method for anxiety and depression in the UK Biobank. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *186*B: 389–398. https://doi.org/10.1002/ajmg.b.32878

Chapter 3: The genetic overlap between depression and anxiety symptom severity and functional impairment

This chapter is a manuscript that will be submitted for peer-review following approval from co-authors. Supplementary materials for this chapter, as detailed in the text, are included in Appendix B.

Authors: **Megan Skelton**[1,2], Jessica Mundy[1,2], Abigail R ter Kuile[1,2], Brett N Adey[1], Chérie Armour[3], Joshua E J Buckman[4,5], Jonathan R I Coleman[1,2], Molly R Davies[1,2], Colette R Hirsch[1,6], Matthew Hotopf[1,2,6], Ian R Jones[7], Gursharan Kalsi[1,2], Georgina Krebs[1,6], Sang Hyuck Lee[1,2], Yuhao Lin[1], Andrew M McIntosh[8], Alicia J Peel[1], Christopher Rayner[1], Katharine A Rimes[1,6], Daniel J Smith[9], Katherine N Thompson[1], David Veale[1,2,6], James T R Walters[7], Janet Wingrove[1,6], Christopher Hübel[1,2,10,11], Gerome Breen[1,2], Thalia C Eley[1,2]

1. Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
2. National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK
3. School of Psychology, Queen's University Belfast, Belfast, Northern Ireland, UK
4. Centre for Outcomes Research and Effectiveness (CORE), Research Department of Clinical, Educational and Health Psychology, University College London, London, UK
5. iCope – Camden and Islington Psychological Therapies Services, Camden and Islington NHS Foundation Trust, London, UK
6. South London and Maudsley NHS Foundation Trust, London, UK
7. National Centre for Mental Health, Division of Psychiatry and Clinical Neuroscience, Cardiff University, Cardiff, UK
8. Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
9. Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
10. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
11. National Centre for Register-based Research, Aarhus Business and Social Sciences, Aarhus University, Aarhus, Denmark

3.1 Abstract

Impairment in the ability to perform one's roles and engage in activities such as work and socialising is a diagnostic criterion of major depressive disorder (MDD) and most anxiety disorders. Evidence suggests that symptom severity and functional impairment are partially distinct, however, functional impairment is often overlooked in treatment studies. Symptoms of depression and anxiety are heritable and share much of the same genetic risk but their genetic relationship with functional impairment is unknown. In a sample of 17,130 participants with lifetime depression or anxiety from the Genetic Links to Anxiety and Depression (GLAD) Study we analysed total scores from the Patient Health Questionnaire 9 (PHQ9; depression symptoms), Generalized Anxiety Disorder 7 (GAD7; anxiety symptoms) and Work and Social Adjustment Scale (WSAS; functional impairment). A genome-wide association study of each phenotype was performed with REGENIE software. Heritability was estimated with GCTA-GREML and genetic correlations calculated with bivariate-GREML. Phenotypic correlations were moderate (Pearson's *r* across traits = 0.50 - 0.69). No significant genetic variants were identified. Heritability estimates were significant (SNP h^2 = 0.11 - 0.19) and genetic correlations were high between all three measures (r_g = 0.77 - 0.87). Our results suggest that, within individuals with lifetime depression or anxiety, the genetic variants that underlie symptom severity are largely the same as those influencing functional impairment. This suggests that symptom scales can sufficiently capture genetic influences on MDD and anxiety disorders, which primarily differ from symptom severity by incorporating functional impairment.

3.2 Background

Major depressive disorder (MDD) and anxiety disorders are characterised by the presence of symptoms such as sadness, worry and fear that cause clinically significant distress or impairment in important areas of functioning. Functional impairment is a key consideration when determining diagnostic status (American Psychiatric Association, 2013) and therefore differentiates normal symptom variation from disorder levels. Furthermore, patients rate a return to normal functioning as a very important treatment outcome (Zimmerman et al., 2006). Despite this, it has been widely overlooked in both treatment trials and routine treatment outcome monitoring (Kamenov et al., 2015). Instead, remission and recovery are often defined using only symptom measures. An individual experiencing no symptoms of depression or anxiety will, by extension, not experience functional impairment due to symptoms. However, the relationship in the presence of symptoms is not perfectly correlated, such that individuals with the same level of symptom severity can experience different levels of functional impairment (Brenes, 2007; Denninger et al., 2011; Jha et al., 2019; Rapaport et al., 2005; Zimmerman et al., 2008). There is also evidence of asynchronous change, with some patients experiencing persistent impairment following symptomatic remission (Howard et al., 1986; IsHak et al., 2016; Sacchetti et al., 2015; Saris et al., 2017). This highlights the importance of assessing functional impairment alongside symptoms for an accurate reflection of patient wellbeing and treatment efficacy.

MDD and anxiety disorders result from the complex effects of both genetic and environmental factors. They show moderate heritability (Hettema et al., 2001; Sullivan et al., 2000), which is explained by many genetic variants each with a very small effect size (Purves et al., 2020; Wray et al., 2018). Genetic overlap between traits can be determined by estimating a genetic correlation (r_g) which ranges between -1 and 1. The high degree of genetic overlap between MDD and anxiety disorders is well-established, and most of the genetic associations are shared across these aetiologies (approximate r_g 0.8 - 1; Kendler et al., 1992; Purves et al., 2020; Roy et al., 1995; Thorp et al., 2021). The genetics of functional impairment have scarcely been researched (McGrath et al., 2013; Ordonana et al., 2013) but there are indications of a moderate heritable component from twin studies (Rijsdijk et al., 2003; Romeis et al., 2005). Furthermore, a twin study of MDD and associated functional

impairment concluded substantial but incomplete genetic overlap (Foley et al., 2003). However, thus far, genomic data has not been used to explore the genetic overlap between current depression and anxiety symptom severity, and functional impairment.

In this study, we analysed total scores of current depression symptoms (Patient Health Questionnaire 9-item version (PHQ); Kroenke et al., 2001), anxiety symptoms (Generalized Anxiety Disorder 7-item Scale (GAD7); Spitzer et al., 2006) and functional impairment (Work and Social Adjustment Scale (WSAS); Marks, 1986). These measures are widely used within research and clinical settings, including routine outcome monitoring in the National Health Service (NHS) England Improving Access to Psychological Treatment (IAPT) services (described in the introduction, section 1.8). A study in a sample of patients receiving IAPT treatment for symptoms of depression or anxiety found moderate phenotypic correlations between the WSAS and both the PHQ9 (0.58) and GAD7 (0.43) (Zahra et al., 2014). This aligns with the wider literature showing significant but incomplete phenotypic overlap of symptom severity and functional impairment. Further evidence comes from the original validation studies of the PHQ9 and GAD7, which were primarily conducted in primary care samples (Kroenke et al., 2001; Spitzer et al., 2006) . To assess the construct validity of total scores, an extra item was added to measure functional impairment: “If you checked off any problems … how difficult have these … made it for you to do your work, take care of things at home, or get along with other people?”. Self-reported disability days were recorded for the same purpose. Correlations with each of these two items (i.e., impairment and disability days) were 0.55 and 0.39 for the PHQ9 score and 0.63 and 0.27 for the GAD7 score. Overall, estimates from the existing literature indicate moderate phenotypic associations between measures of functional impairment and both the PHQ9 and GAD7, but these are weaker when a more objective measure of impairment is used. In practice, the PHQ9 and GAD7 are used as the nine- and seven-item symptom scales, such as in IAPT services, where they are the primary treatment outcome measures.

Similarity between phenotypic and genetic correlations has been reported for both depression and anxiety symptoms (Waszczuk et al., 2014). If this pattern extended to their relationship with functional impairment, it would indicate both a shared genetic factor as well as symptom-specific and impairment-specific genetic influences, reflecting a

moderate phenotypic association. High genetic correlations would be consistent with overlap between quantitative symptom score and diagnostic case-control phenotypes that incorporate impairment (e.g., r_g = 1 (SE = 0.2) for depression phenotypes; Direk et al. (2017); also see Purves et al. (2020)). However, incomplete overlap could align with, and help to explain, reported genetic variation specific to diagnoses (Cai et al., 2020; Kendler et al., 2018). Genetic factors associated with impairment may capture something beyond genetic influences on symptoms which is relevant to a full diagnostic presentation. An estimate of genetic overlap between symptom severity and functional impairment would therefore indicate whether there is value in supplementing symptom scales with measures of functional impairment in genetic studies. This would also inform whether existing genetic information on symptom severity is sufficient for prediction models of disorder prognosis, including treatment outcomes, or whether genetics on impairment are an important addition. If genetic overlap is high, this would lend support to the existing evidence that symptom scales sufficiently capture genetic effects on MDD and anxiety disorders. The practical advantages here are that symptom scales provide a 'brief phenotyping' instrument for detecting genetic effects on disorders, enabling vast increases in sample size, which are difficult to obtain via clinician derived diagnostic instruments.

In this study, we aimed to investigate the genetic influences on quantitative phenotypes of depression symptoms, anxiety symptoms and functional impairment, and the genetic overlap between symptom severity and functional impairment. Analysis was performed in a sample with lifetime experience of depression or anxiety. We expected that genetic correlations between symptom severity and functional impairment would be significantly different from zero, and moderate but incomplete, reflecting the phenotypic correlations (0.4 - 0.7, based on previous literature).

3.3 Methods

3.3.1 Sample

The Genetic Links to Anxiety and Depression (GLAD) Study is an online study that recruits individuals with lifetime experience of depression and/or anxiety, primarily from the

general population (Davies et al., 2019). Participants provide genetic data via a postal saliva sample and are required to meet case criteria on diagnostic questionnaires or self-report a diagnosis by a medical professional (hence the use of the broad terms 'depression' and 'anxiety'). Our analysis was limited to those with phenotypic data for at least one outcome of interest, covariate information and genotype data that passed quality control (N = 17,130). Ethical approval for the GLAD Study was granted by the London-Fulham Research Ethics Committee (REC reference: 18/LO/1218).

3.3.2 Phenotype measures

The analysis was centred around three phenotypes. **Depression symptoms** were assessed using the PHQ9. The PHQ9 measures the recent frequency of nine depression symptoms using the stem question, "Over the last two weeks, how often have you been bothered by any of the following problems?". Each item has a four-point response scale from 'not at all' (scored 0) to 'nearly every day' (scored 3). The summed scores indicate severity from 0 to 27. The PHQ9 has good internal ($\alpha = 0.89$) and test-retest (0.84) reliability (Kroenke et al., 2001).

Anxiety symptoms were measured by the GAD7, which has a very similar format to the PHQ9. It has the same overarching question regarding frequency of recent problems, and the seven items are answered using the four-point scale, yielding total scores from 0 to 21. Internal consistency ($\alpha = 0.92$) and test-retest reliability (0.83) are good (Spitzer et al., 2006).

The WSAS assesses the impact of symptoms on **functional impairment** in five life domains: ability to work, home management, social leisure activities, private leisure activities and ability to form and maintain close relationships. Each item is worded as, "because of my problem, my <domain> is impaired". A nine-point response scale of no (scored 0) to severe (scored 8) impairment gives total scores from 0 to 40. Previous analyses of patients with depression or anxiety have shown that the WSAS captures a single factor, has good test-retest reliability (0.73) and at least acceptable internal consistency ($\alpha = 0.72 - 0.94$) (Mataix-Cols et al., 2005; Mundt et al., 2002). One limitation is that the work item is aimed at

individuals in employment and is otherwise answered 'not applicable', resulting in missing data. Some (e.g., NHS Digital) have approached this by imputation using the mean of the individual's four non-missing WSAS items. This can create bias and spurious positive results if the data are missing not at random (missingness is related to the true value; Little & Rubin, 2002). We explored phenotypic and genetic relationships between the total scores from four WSAS items (without the work item) and all five items, as well as the four-item total and the work item. Subsequently, we present results from an individual mean imputed WSAS score, with complete case results in the supplementary.

3.3.3 Genotyping and quality control

Genotyping was performed by the National Institute for Health and Care Research Cambridge Biomedical Research Centre on the Affymetrix UK Biobank Axiom Array. The data from freeze 2.0 were used. Variants were imputed using the TOPMed reference panel (version r2 on GRCh38) and filtered to quality (R^2) > 0.3. Quality control exclusions were individuals with missingness > 5%, non-European ancestry (due to low sample size) or signs of contaminated data (unusual identity by descent or discordant sex statistics). Genetic variant exclusions were missingness > 2%, minor allele frequency < 1% and Hardy-Weinberg equilibrium $p < 10^{-8}$. This resulted in 564,245 genotyped and 7,027,957 imputed variants.

3.3.4 Statistical analyses

A genome-wide association study (GWAS) was performed with each phenotype using REGENIE version 2.2.4 (Mbatchou et al., 2021). Covariates were age, age squared, sex, four levels of genotyping batch, and the first 10 genetic principal components. Single nucleotide polymorphism (SNP) heritability was estimated with genomic-relatedness- based restricted maximum-likelihood in 'genome-wide complex trait analysis' software version 1.94 (GCTA-GREML; Yang et al., 2011).

Genetic overlap between the three phenotypes was calculated using genetic correlations in bivariate-GREML (S. H. Lee et al., 2012). These methods use individual-level data on

common SNPs identifiable from the genetic array to produce a genomic relatedness matrix. One of each pair of participants exceeding a relatedness threshold of 0.05 was removed (373 individuals). As well as the default test of whether the genetic correlation significantly differed from 0, we tested whether each genetic correlation of symptoms and functional impairment differed from 1. Furthermore, we estimated the bivariate SNP-heritability, performing calculations and simulating standard errors as per Morris et al. (2020). Bivariate heritability represents the proportion of the phenotypic correlation attributable to genetic overlap (de Vries et al., 2021; Morris et al., 2020). For example, a weak genetic correlation and high bivariate heritability would indicate that although the genetics of each phenotype are more trait-specific than shared, the genetic influences they have in common underlie much of the observed phenotypic overlap.

We also calculated genetic correlations with summary statistics from five a-priori selected phenotypes using linkage-disequilibrium score regression (LDSC, version 1.0.1; Bulik-Sullivan, Finucane, et al., 2015; Bulik-Sullivan, Loh, et al., 2015). First, we selected three case-control phenotypes of mental health diagnoses: MDD (Wray et al., 2018), anxiety disorders (Purves et al., 2020), and schizophrenia (Pardiñas et al., 2018). These would help us to determine the pattern of genetic relationships between each of symptom severity and functional impairment with depression and anxiety phenotypes from samples incorporating diagnostic measures that account for impairment. It would also reveal whether impairment attributed to depression and anxiety symptoms (WSAS) is associated with a diagnostically distinct disorder (schizophrenia). Second, two quantitative traits were chosen: years of education (J. J. Lee et al., 2018) and self-rated health (Harris et al., 2017). Education reflects socioeconomic status as well as cognitive ability, which each might have genetic associations with functional impairment. Self-rated health is associated with both mental and physical health outcomes and might therefore capture a somewhat different but nonetheless relevant phenotype to functional impairment. See Supplementary Table 1 for further details of these studies. To correct for multiple testing, a Bonferroni adjustment was applied to the significance threshold ($p < 0.01$ for five tests). Data preparation and visualisation were performed in R version 4.1.2 (R Core Team, 2021).

3.4 Results

3.4.1 Sample characteristics

Sample characteristics (N = 17,130) are presented in Supplementary Table 2 and Supplementary Figure 1. Briefly, participants were aged between 16 and 93 years (mean 39.5, SD = 14.6) and 78% were female. On average, participants had moderate current depression symptoms (mean PHQ9 11.2 (SD = 6.9); Kroenke et al., 2001), mild anxiety symptoms (mean GAD7 8.9 (SD = 5.9); Spitzer et al., 2006) and moderate functional impairment (mean WSAS 17.2 (SD = 9.2); Mataix-Cols et al., 2005; Mundt et al., 2002). Phenotypic and genetic explorations of the WSAS with and without the work item showed similar results (Supplementary Information 1).

3.4.2 Heritability estimates

No significant variants were identified in the GWAS of any of the three traits (at $p < 5x10^{-8}$; see Supplementary Figure 2). SNP heritability estimates were significant ($p < 0.05$) for depression symptoms (0.19, SE = 0.04, $p = 6x10^{-9}$), anxiety symptoms (0.17, SE = 0.03, $p = 2x10^{-7}$) and functional impairment (0.11, SE = 0.03, $p = 2x10^{-4}$).

3.4.3 Phenotypic and genetic overlap between traits

Phenotypic and genetic correlations between depression symptoms, anxiety symptoms and functional impairment are presented in Figure 1 (also Supplementary Tables 3 and 4). Phenotypic correlations between traits were moderate (Pearson's *r* range: 0.50 - 0.69), and genetic overlap was consistently higher (r_g range: 0.77 - 0.87). Phenotypic correlations were highest between depression and anxiety symptoms and lowest between anxiety symptoms and functional impairment. The same pattern was observed in the genetic correlations. The genetic correlations of functional impairment with depression symptoms and with anxiety symptoms were both significantly different from 1 ($p = 0.01$ and $= 7x10^{-3}$, respectively). Using LDSC to estimate heritabilities and genetic correlations produced similar results (Supplementary Information 2).

The proportion of phenotypic correlation due to shared genetics (bivariate SNP-heritability) between functional impairment and depression and anxiety symptoms, respectively, was 0.24 (SE = 0.04) and 0.26 (SE = 0.05). The high genetic correlations suggest that the majority of variants involved in symptom severity and functional impairment are the same, and the bivariate heritability indicates that this genetic overlap explains a quarter of the phenotypic overlap between traits.

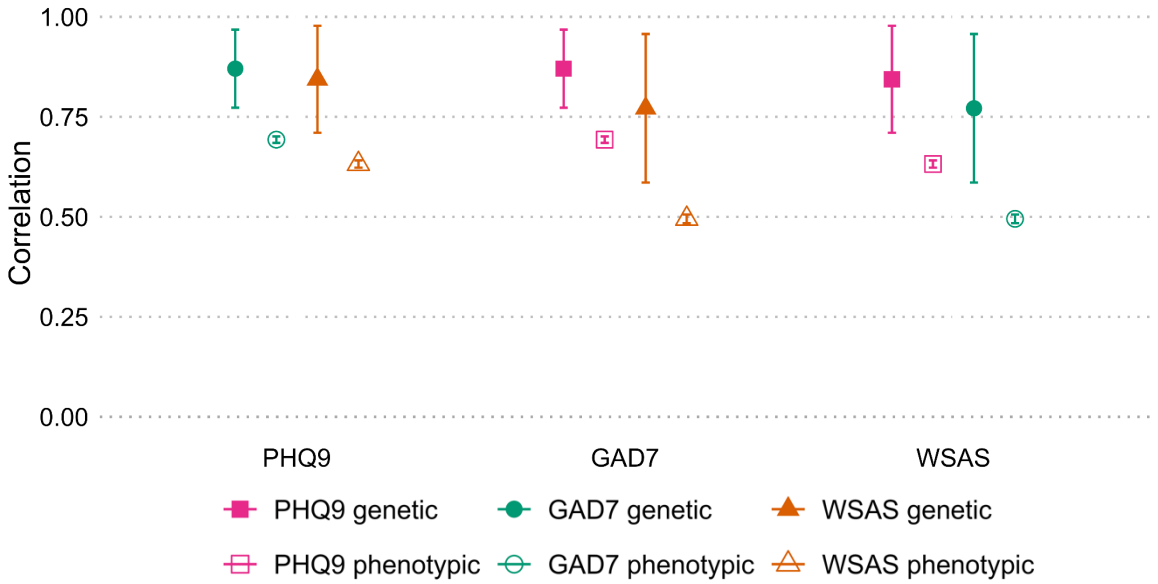


Figure 1. Genetic and phenotypic correlations between depression symptoms, anxiety symptoms and functional impairment in a sample from the GLAD Study (N = 17,130)

Error bars represent 95% confidence intervals. All estimates are significant ($p < 0.05$). PHQ9 = depression symptoms score, GAD7 = anxiety symptoms score, WSAS = functional impairment score. Genetic correlations were estimated using GCTA bivariate-GREML and phenotypic correlations using Pearson's r. For ease of comparability, both sides of the correlations are presented, therefore information is duplicated. For example, the PHQ9-WSAS genetic correlation is presented both by the filled orange triangle above PHQ9 on the x-axis and the filled pink square above WSAS.

3.4.4 Genetic correlations with existing phenotypes

LDSC estimates of genetic correlations between each of the measures and five existing phenotypes are shown in Figure 2. All three phenotypes had non-zero estimates with MDD, years of education and self-rated health, which remained significant after correcting for multiple testing. Negative correlations with years of education and self-rated health

indicated that, for example, genetic variants associated with higher PHQ9 scores were associated with fewer years of education. No significant associations were found with anxiety disorder or schizophrenia.

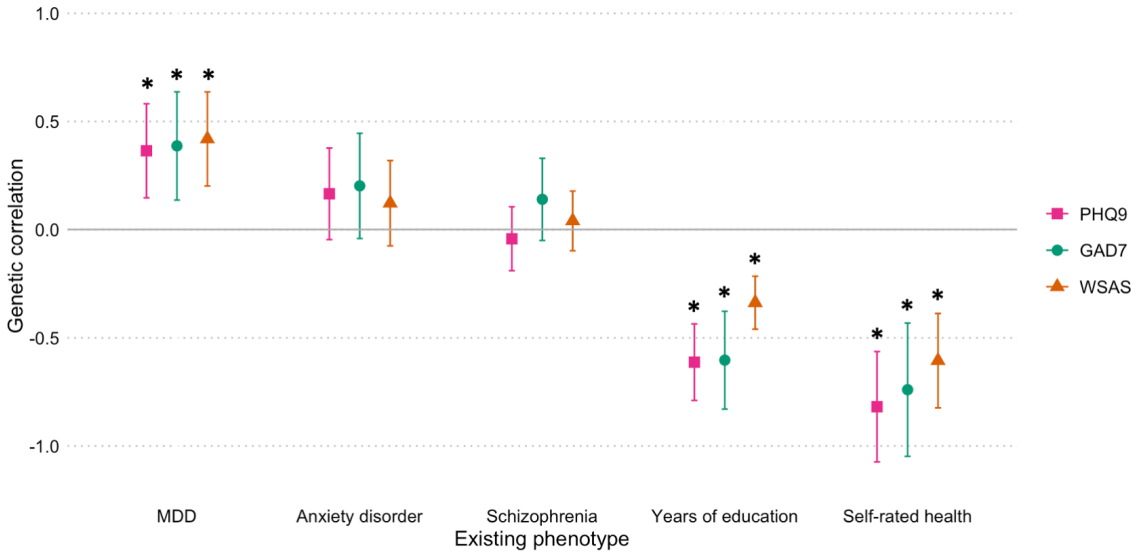


Figure 2. Genetic correlations between the three traits analysed in a sample from the GLAD Study (N = 17,130) and five existing phenotypes

*Error bars represent 95% confidence intervals. * Significant at p < 0.01. PHQ9 = depression symptoms score, GAD7 = anxiety symptoms score, WSAS = functional impairment score. MDD = major depressive disorder (Wray et al., 2018 without 23andme or UKBiobank), anxiety disorder (Purves et al., 2020), schizophrenia (Pardiñas et al., 2018), years of education (J. J. Lee et al., 2018), self-rated health (Harris et al., 2017). See Supplementary Table 1 for further details of these phenotypes. Genetic correlations were estimated using LDSC.*

3.5 Discussion

This study investigated the genetic influences on quantitative measures of depression symptoms (PHQ9), anxiety symptoms (GAD7), and functional impairment (WSAS) in 17,130 individuals with lifetime depression or anxiety. Heritability estimates of these traits and the genetic correlations between them were all significant and indicated substantial genetic overlap between symptom severity and functional impairment.

3.5.1 Heritability

SNP heritability estimates showed that the PHQ9, GAD7 and WSAS are under the influence of common genetic variants, with 11-19% of the phenotypic variance in this sample attributable to genetic factors. The estimates for symptoms were comparable to those reported from SNP-based analyses of case-control MDD and anxiety disorder phenotypes (e.g. 9% (SE < 0.01) (Wray et al., 2018) and 26% (SE = 0.01) (Purves et al., 2020), respectively). The heritability of functional impairment was similar to that of symptoms, as expected from existing twin-based estimates.

3.5.2 Phenotypic and genetic overlap between traits

Phenotypic correlations were consistent in size and pattern with previous estimates (Kroenke et al., 2001; Spitzer et al., 2006; Zahra et al., 2014) and existing evidence that symptom severity is not synonymous with functional impairment. The stronger phenotypic correlations between functional impairment and depression symptoms than anxiety symptoms might indicate that functioning is more impaired by symptoms of depression. However, this could also arise if the PHQ9 items are more conceptually similar to WSAS items. For example, 'little interest or pleasure in doing things' and 'private and social leisure activities'. Furthermore, although sleep difficulties, low energy and impaired concentration symptoms feature in diagnostic criteria for both MDD and generalised anxiety disorder (American Psychiatric Association, 2013), they only appear in the PHQ9, not the GAD7. There is evidence that these symptoms are especially relevant to functional impairment (Fried & Nesse, 2014), which could drive the higher correlation. Further work is required to assess the structural relationship between these measures. One approach could be to use network analysis to reveal item-level associations. A factor analysis of PHQ9 and GAD7 items in the GLAD Study identified four factors (Thompson et al., 2021); how these relate to or change with the addition of WSAS items might reveal clinically useful presentations.

From a genetic perspective it appears that symptom severity and functional impairment are highly similar. That is, many of the genetic variants that contribute to symptoms of depression and anxiety also play a role in functional impairment. This genetic overlap was found to underlie approximately a quarter of the phenotypic correlation. The high genetic

overlap suggests that the majority of the genetic factors associated with functional impairment, which is clinically relevant for diagnosis and treatment outcomes, will be identifiable using symptom-based measures in genetic studies of depression or anxiety. For genetic variant discovery using symptom scales or prediction models using information from symptom based GWAS, it is therefore not crucial to supplement with information on functional impairment. This is consistent with a UK Biobank study showing that utilising additional components of a diagnostic questionnaire for depression, including a binary item assessing functional impairment, had little impact on heritability or relevant genetic correlations beyond the core symptoms (Jermy et al., 2021). The high genetic correlations provide further support for the use of symptom scales as brief phenotypes for genetic studies of MDD and anxiety disorders, which differ from symptom severity by the incorporation of functional impairment. However, it is necessary to measure functional impairment from a clinical perspective as it is highly important to patients and cannot be inferred from symptom scales. Furthermore, there was evidence of modest unique genetic variance to functional impairment, as shown by genetic correlations significantly different from $r_g = 1$. This requires further investigation. The genetic correlations were higher than expected based on the phenotypic correlations but remain in line with the proposed significant yet incomplete genetic overlap.

Genetic correlations with existing phenotypes revealed similar results for all three traits, reflecting the high genetic overlap among them. The strongest were negative correlations with years of education and self-rated health. This is consistent with previous reports for case-control MDD and anxiety disorders (Harris et al., 2017; Purves et al., 2020; Wray et al., 2018), although point estimates in the present analyses were stronger. Genetic correlations with MDD were all significant and in the expected direction, although the correlation with depression symptoms was much weaker than estimated in the existing literature ($r_g = 1$; Direk et al., 2017). Surprisingly, correlations with anxiety disorders and schizophrenia were not significantly different from zero. The differences observed between our analyses and the literature are likely due to selection bias, which is discussed in the limitations section.

3.5.3 Limitations

This is the first study to perform a genomic analysis of the relationship between depression and anxiety symptoms and functional impairment. It is one of the only genetic studies of functional impairment, a highly relevant clinical outcome. The measures employed are widely used in clinical and research settings and have been validated in various cultures and patient groups (e.g., Mughal et al., 2020). However, several limitations should be highlighted when interpreting these results.

Our sample was restricted to individuals with lifetime depression and/or anxiety. Depression and anxiety are common, complex disorders, influenced by both genetic and environmental risk factors. Individuals with diagnoses therefore have different underlying aetiologies, and not everyone will have a high genetic loading of risk variants. However, this sample was likely overrepresented for individuals at the upper end of the genetic risk distribution, which would have reduced relevant genetic variation overall. As such, the observed relationships may be specific to clinical samples and not generalise at the population level. It is likely that this selection bias underlies the weak and non-significant genetic correlations observed with case-control phenotypes. On the other hand, scores for symptoms and functional impairment were normally distributed, as opposed to frequently observed zero-inflated estimates in population cohort studies. Arguably, investigating overlap between symptom severity and functional impairment requires a clinical sample who have non-zero levels of these traits, as it is only with symptoms that impairment becomes relevant. The high genetic overlap between depression and anxiety symptoms, while not our focus, was consistent with existing findings from diagnostic case-control phenotypes (Kendler et al., 1992; Purves et al., 2020; Roy et al., 1995) and an analysis of the PHQ9 and GAD7 in a population-based sample (Thorp et al., 2021). This provides tentative support for the generalisation of our findings regarding functional impairment to the general population, but further work is required to confirm this.

A common limitation of GWAS is low statistical power to detect genetic associations after Bonferroni correction. Analyses were thus underpowered for variant identification that could have been used to further investigate genetic overlap. However, GLAD Study

recruitment is ongoing so the sample size may increase. This will also hopefully lead to a cohort that is more demographically representative of the general population. The sample here was disproportionately female and highly educated and was restricted to European ancestry due to few participants of other ancestries.

Finally, over 10% of the sample were missing for the WSAS work item. We handled this using individual mean imputation and conducted complete case sensitivity analyses that revealed similar results. As GLAD is a recontactable resource one option is to request 'not applicable' respondents to repeat these measures, answering the work item in terms of their main working role, for example, looking after the family. Additionally, we could direct those too impaired to work to endorse severe impairment. As well as more complete data, this would help to assess whether remaining missingness is random and therefore suitable for imputation. Ideally, we would have a measure without gated questions. Objective measures of impairment might also reveal more about the relationship with symptoms, given lower phenotypic correlations reported with these in the PHQ9 and GAD7 validation papers (Kroenke et al., 2001; Spitzer et al., 2006).

3.5.4 Future directions

As statistical power for GWAS increases through larger sample sizes as well as detailed phenotyping, the question of genetic overlap between symptoms and impairment should be revisited. Investigations within the general population are required to determine the generalisability of our results. Explorations of symptoms and functional impairment in other mental health disorders might also be illuminating. Impairment has been proposed as a transdiagnostic phenotype that could be used to maximise sample sizes across disorders (McGrath et al., 2013) and perhaps plays a role in the common genetic factor underlying all mental health disorders (Caspi et al., 2014). Functional impairment may therefore offer an additional method of brief phenotyping when other information, including symptom severity, is unavailable. Furthermore, to ensure that findings are applicable across the patient population, there may be value in exploring group differences in the phenotypic and genetic relationships of these measures. Sex differences are reported phenotypically for depression and anxiety symptoms as well as impairment

(Romeis et al., 2005; Thandi et al., 2017; Thompson et al., 2021; Ulbricht et al., 2016), and also genetically for depression (Kendler, 2001).

Finally, genetic correlations can result from multiple mechanisms. A variant can influence both traits, or one trait which then drives the other (van Rheenen et al., 2019), and correlations can arise from genetically similar subgroups. Further work could elucidate the mechanisms at play here, using tools such as BUHMBOX (Han et al., 2016), mtCOJO (Zhu et al., 2018) and Mendelian Randomisation (Davey Smith & Hemani, 2014). For example, conditional analyses could reveal whether there are variants specific to functional impairment, independent of symptom severity, as they have for depression and anxiety themselves (Levey et al., 2020; Grotzinger et al., 2022).

3.5.6 Conclusion

Functional impairment is often overlooked in treatment studies despite only moderate phenotypic overlap with symptom severity. We found high but incomplete genetic overlap with symptoms, indicating that the genetic variants associated with functional impairment will be largely captured by using symptom severity measures. This supports the notion that brief measures of symptom severity, which differ from diagnoses by lacking assessment of impairment, are likely to continue to be useful measures in genetic studies of depression and anxiety. Genetic information about symptom severity may also be sufficient for prognostic prediction models.

3.6 References

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

Brenes, G. A. (2007). Anxiety, depression, and quality of life in primary care patients. *Primary Care Companion to the Journal of Clinical Psychiatry*, *9*(6), 437–443. https://doi.org/10.4088/pcc.v09n0606

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*(4), 437–447. https://doi.org/0.1038/s41588-020-0594-5.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98. https://doi.org/10.1093/hmg/ddu328

Davies, M. R., Kalsi, G., Armour, C., Jones, I. R., McIntosh, A. M., Smith, D. J., … Breen, G. (2019). The Genetic Links to Anxiety and Depression (GLAD) Study: Online recruitment into the largest recontactable study of depression and anxiety. *Behaviour Research and Therapy*, *123*, 103503. https://doi.org/10.1016/j.brat.2019.103503

Denninger, J. W., van Nieuwenhuizen, A. O., Wisniewski, S. R., Luther, J. F., Trivedi, M. H., Rush, A. J., … Fava, M. (2011). Changes in depressive symptoms and social functioning in the sequenced treatment alternatives to relieve depression study. *The Journal of*

Nervous and Mental Disease, *199*(10), 807–810. https://doi.org/10.1097/NMD.0b013e31822fcbe2

de Vries, L. P., van Beijsterveldt, T. C. E. M., Maes, H., Colodro-Conde, L., & Bartels, M. (2021). Genetic Influences on the Covariance and Genetic Correlations in a Bivariate Twin Model: An Application to Well-Being. *Behavior Genetics*, *51*(3), 191–203. https://doi.org/10.1007/s10519-021-10046-y

Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., … Sullivan, P. F. (2017). An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biological Psychiatry*, *82*(5), 322–329. https://doi.org/10.1016/j.biopsych.2016.11.013

Foley, D. L., Neale, M. C., Gardner, C. O., Pickles, A., & Kendler, K. S. (2003). Major Depression and Associated Impairment: Same or Different Genetic and Environmental Risk Factors? *Am J Psychiatry*, *160*, 2128–2133. https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.160.12.2128

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

Han, B., Pouget, J. G., Slowikowski, K., Stahl, E., Lee, C. H., Diogo, D., … Raychaudhuri, S. (2016). A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nature Genetics*, *48*(7), 803–810. https://doi.org/10.1038/ng.3572

Harris, S. E., Hagenaars, S. P., Davies, G., David Hill, W., Liewald, D. C. M., Ritchie, S. J., … Deary, I. J. (2017). Molecular genetic contributions to self-rated health. *International Journal of Epidemiology*, *46*(3), 994–1009. https://doi.org/10.1093/ije/dyw219

Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American Journal of Psychiatry*, *158*(10), 1568–1578. https://doi.org/10.1176/appi.ajp.158.10.1568

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, *41*(2), 159–164. https://www.ncbi.nlm.nih.gov/pubmed/3516036

IsHak, W. W., James, D. M., Mirocha, J., Youssef, H., Tobia, G., Pi, S., … Cohen, R. M. (2016). Patient-reported functioning in major depressive disorder. *Therapeutic Advances in*

Chronic Disease, *7*(3), 160–169. https://doi.org/10.1177/2040622316639769

Jermy, B. S., Glanville, K. P., Coleman, J. R. I., Lewis, C. M., & Vassos, E. (2021). Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Molecular Psychiatry*, *26*(12), 7337–7345. https://doi.org/10.1038/s41380-021-01231-w

Jha, M. K., South, C., Trivedi, J., Minhajuddin, A., Rush, A. J., & Trivedi, M. H. (2019). Prediction of Acute-Phase Treatment Outcomes by Adding a Single-Item Measure of Activity Impairment to Symptom Measurement: Development and Validation of an Interactive Calculator from the STAR*D and CO-MED Trials. *International Journal of Neuropsychopharmacology*, *22*(5), 339–348. https://doi.org/10.1093/ijnp/pyz011

Kamenov, K., Cabello, M., Coenen, M., & Ayuso-Mateos, J. L. (2015). How much do we know about the functional effectiveness of interventions for depression? A systematic review. *Journal of Affective Disorders*, *188*, 89–96. https://doi.org/10.1016/j.jad.2015.08.035

Kendler, K. S. (2001). Twin Studies of Psychiatric Illness: An Update. *Archives of General Psychiatry*.

Kendler, K. S., Gardner, C. O., Neale, M. C., Aggen, S., Heath, A., Colodro-Conde, L., … Gillespie, N. A. (2018). Shared and specific genetic risk factors for lifetime major depression, depressive symptoms and neuroticism in three population-based twin samples. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S003329171800377X

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of General Psychiatry*, *49*(9), 716–722. https://www.ncbi.nlm.nih.gov/pubmed/1514877

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

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived

genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540–2542. https://doi.org/10.1093/bioinformatics/bts474

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119013563.ch4

Marks, I. M. (1986). *Behavioural psychotherapy: Maudsley pocket book of clinical management*. Wright/Institute of Psychiatry Publishing.

Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., … Marks, I. M. (2005). Reliability and validity of the work and social adjustment scale in phobic disorders. *Comprehensive Psychiatry*, *46*(3), 223–228. https://doi.org/10.1016/j.comppsych.2004.08.007

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., … Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, *53*(7), 1097–1103. https://doi.org/10.1038/s41588-021-00870-7

McGrath, L. M., Cornelis, M. C., Lee, P. H., Robinson, E. B., Duncan, L. E., Barnett, J. H., … Smoller, J. W. (2013). Genetic predictors of risk and resilience in psychiatric disorders: a cross-disorder genome-wide association study of functional impairment in major depressive disorder, bipolar disorder, and schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, *162B*(8), 779–788. https://doi.org/10.1002/ajmg.b.32190

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*(16), eaay0328. https://doi.org/10.1126/sciadv.aay0328

Mughal, A. Y., Devadas, J., Ardman, E., Levis, B., Go, V. F., & Gaynes, B. N. (2020). A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries. *BMC Psychiatry*, *20*(1), 338. https://doi.org/10.1186/s12888-020-02753-3

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

Ordonana, J. R., Bartels, M., Boomsma, D. I., Cella, D., Mosing, M., Oliveira, J. R., … Others. (2013). Biological pathways and genetic mechanisms involved in social functioning. *Quality of Life Research: An International Journal of Quality of Life Aspects of*

Treatment, Care and Rehabilitation, *22*(6), 1189–1200. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s11136-012-0277-5&casa_token=ySgsLNJMAP0AAAAA:GP0N7n-dMuz3ll0-luonlzdAOGSAB73Zcr61-TH7wSaC-jtoLdo9OqmJqWDgvPRJ000kiwEkXYSZ

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., … Walters, J. T. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. https://doi.org/10.1038/s41588-018-0059-2

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A Major Role for Common Genetic Variation in Anxiety Disorders. *Molecular Psychiatry*, *25*(12), 3292–3303. https://doi.org/10.1038/s41380-019-0559-1

Rapaport, M. H., Clary, C., Fayyad, R., & Endicott, J. (2005). Quality-of-Life Impairment in Depressive and Anxiety Disorders. *American Journal of Psychiatry*, *162*(6), 1171–1178. https://doi.org/10.1176/appi.ajp.162.6.1171

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rijsdijk, F. V., Snieder, H., Ormel, J., Sham, P., Goldberg, D. P., & Spector, T. D. (2003). Genetic and environmental influences on psychological distress in the population: General Health Questionnaire analyses in UK twins. *Psychological Medicine*, *33*(5), 793–801. https://doi.org/10.1017/s0033291703007451

Romeis, J. C., Heath, A. C., Xian, H., Eisen, S. A., Scherrer, J. F., Pedersen, N. L., … True, W. R. (2005). Heritability of SF-36 among middle-age, middle-class, male-male twins. *Medical Care*, *43*(11), 1147–1154. https://doi.org/10.1097/01.mlr.0000183217.11811.bd

Roy, M. A., Neale, M. C., Pedersen, N. L., Mathé, A. A., & Kendler, K. S. (1995). A twin study of generalized anxiety disorder and major depression. *Psychological Medicine*, *25*(5), 1037–1049. https://www.ncbi.nlm.nih.gov/pubmed/8588001

Sacchetti, E., Frank, E., Siracusano, A., Racagni, G., Vita, A., & Turrina, C. (2015). Functional impairment in patients with major depression in clinical remission: results from the VIVAL-D-Rem, a nationwide, naturalistic, cross-sectional survey. *International Clinical*

Psychopharmacology, *30*(3), 129–141. https://doi.org/10.1097/YIC.0000000000000074

Saris, I. M. J., Aghajani, M., van der Werff, S. J. A., van der Wee, N. J. A., & Penninx, B. W. J. H. (2017). Social functioning in patients with depressive and anxiety disorders. *Acta Psychiatrica Scandinavica*, *136*(4), 352–361. https://doi.org/10.1111/acps.12774

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, 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

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *The American Journal of Psychiatry*, *157*(10), 1552–1562. https://doi.org/10.1176/appi.ajp.157.10.1552

Thandi, G., Fear, N. T., & Chalder, T. (2017). A comparison of the Work and Social Adjustment Scale (WSAS) across different patient populations using Rasch analysis and exploratory factor analysis. *Journal of Psychosomatic Research*, *92*, 45–48. https://doi.org/10.1016/j.jpsychores.2016.11.009

Thompson, K. N., Hübel, C., Cheesman, R., Adey, B. N., Armour, C., Davies, M. R., … Eley, T. C. (2021). Age and sex-related variability in the presentation of generalized anxiety and depression symptoms. *Depression and Anxiety*, *38*(10), 1054–1065. https://doi.org/10.1002/da.23213

Thorp, J. G., Campos, A. I., Grotzinger, A. D., Gerring, Z., An, J., Ong, J.-S., … Derks, E. M. (2021). Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nature Human Behaviour*, *5*(10), 1432–1442. https://doi.org/10.1101/2020.04.08.20057653

Ulbricht, C. M., Rothschild, A. J., & Lapane, K. L. (2016). Functional Impairment and Changes in Depression Subtypes for Women in STAR*D: A Latent Transition Analysis. *Journal of Women's Health* , *25*(5), 464–472. https://doi.org/10.1089/jwh.2015.5361

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews. Genetics*. https://doi.org/10.1038/s41576-019-0137-z

Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M., & Eley, T. C. (2014). The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. *JAMA Psychiatry* , *71*(8), 905–916.

https://doi.org/10.1001/jamapsychiatry.2014.655

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., … Byng, R. (2014). The work and social adjustment scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, *18*(2), 131–138. https://doi.org/10.3109/13651501.2014.894072

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., … Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*, *9*(1), 224. https://doi.org/10.1038/s41467-017-02317-2

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*, *163*(1), 148–150. https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.163.1.148

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Boerescu, D., & Attiullah, N. (2008). Remission in depressed outpatients: more than just symptom resolution? *Journal of Psychiatric Research*, *42*(10), 797–801. https://doi.org/10.1016/j.jpsychires.2007.09.004

Chapter 4. Trajectories of depression and anxiety symptom severity during psychological therapy for common mental health problems

A version of this chapter is available as a preprint on the PsyArXiv server. It is currently undergoing peer-review at *Psychological Medicine* following an invitation to revise and resubmit. Supplementary materials for this chapter, as detailed in the text, are included in Appendix C.

Preprint citation: **Skelton, M.**, Carr, E., Buckman, J. E. J., Davies, M. R., Goldsmith, K. A., Hirsch, C., … & Eley, T. C. (2021, November 16). Trajectories of depression and anxiety symptoms during psychological therapy for common mental health problems. https://doi.org/10.31234/osf.io/8scpx

Authors: **Megan Skelton**[1,2], Ewan Carr[1], Joshua E. J. Buckman[3,4], Molly R Davies[1], Kimberley A Goldsmith[1], Colette R Hirsch[1,5], Alicia J Peel[1], Christopher Rayner[1], Katharine A Rimes[1], Rob Saunders[3], Janet Wingrove[6], Gerome Breen[1,2], Thalia C Eley [1,2]

1. Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
2. National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK
3. Centre for Outcomes Research and Effectiveness (CORE), Research Department of Clinical, Educational and Health Psychology, University College London, London, UK
4. iCope – Camden and Islington Psychological Therapies Services, Camden and Islington NHS Foundation Trust, London
5. South London and Maudsley NHS Foundation Trust, London, UK
6. Talking Therapies Southwark, South London and Maudsley NHS Foundation Trust, London, UK

4.1 Abstract

Background: There is substantial variation in patient symptoms following psychological therapy for depression and anxiety. However, reliance on endpoint outcomes ignores additional interindividual variation during therapy. Knowing a patient's likely symptom trajectories could guide clinical decisions. We aimed to identify latent classes of patients with similar symptom trajectories over the course of psychological therapy and explore associations between baseline variables and trajectory class.

Methods: Patients received high-intensity psychological treatment for common mental health problems at NHS Improving Access to Psychological Therapies (IAPT) services in South London (N = 16,258). To identify trajectories, we performed growth mixture modelling of depression and anxiety symptoms over a baseline assessment and ten treatment sessions. We then ran multinomial regressions to identify baseline variables associated with trajectory class membership.

Results: Trajectories of depression and anxiety symptoms were highly similar and best modelled by four classes. Three classes started with moderate-severe symptoms and showed (1) no change, (2) gradual improvement, and (3) fast improvement. A final class (4) showed initially mild symptoms and minimal improvement. Within the moderate-severe baseline symptom classes, patients in the two showing improvement as opposed to no change tended not to report a disability or prescribed medication and were in employment. Patients showing fast improvement additionally reported lower baseline functional impairment on average.

Conclusions: Multiple trajectory classes of depression and anxiety symptoms were associated with baseline characteristics. Identifying the most likely trajectory for a patient at the start of treatment could inform decisions about the suitability and continuation of therapy, ultimately improving patient outcomes.

4.2 Background

Routinely collected patient information can be used to explain some of the variability in outcomes following psychological therapies for depression and anxiety (Delgadillo et al., 2016; Goddard et al., 2015; Kessler et al., 2017; Saunders, Buckman, et al., 2020). Most studies to date have focused on outcomes measured at the end of treatment, such as remission or recovery. Although clinically informative, reliance on endpoint outcomes may obscure interindividual differences in trajectories of symptoms occurring *during* treatment. Subgroups of patients may follow distinct trajectories, such as initially slow responders who nevertheless show clinically significant improvement by the end of treatment. Without a good understanding of these trajectories, treatments might be ended or altered early because patients, or their clinicians, consider the treatment unsuitable.

Person-centred structural equation modelling techniques can be used to reveal unobserved subgroups ('latent classes') of individuals who exhibit similar longitudinal trajectories. It is then possible to explore how baseline variables are associated with trajectory class membership. Relevant variables could be inspected at the start of treatment to identify a patient's most likely pattern of symptoms throughout therapy, and therefore inform patient and clinician expectations and decisions. Such information could also help to monitor whether patients are doing less well than expected and thereby increase positive outcomes (de Jong et al., 2021; Delgadillo et al., 2021; Lambert, 2007).

Studies that have employed these methods with psychological treatment data include analyses of symptoms of depression (Lutz, Stulz & Köck, 2009) and panic disorder (Lutz et al., 2014). As described in the introduction (section 1.11.1), these have reported between three and five trajectory classes, but have been limited in terms of statistical power (Ns < 350) as well as investigations of associations with baseline variables. Moreover, most existing studies have focused on change patterns during only the initial few sessions.

One study investigated symptom trajectory classes with a larger sample of patients (N = 4,394) from two North London National Health Service (NHS) Improving Access to Psychological Therapies (IAPT) services (Saunders et al., 2019). As part of the NHS in

England, the IAPT services provide psychological therapy for adults experiencing common mental health problems, primarily, depression and anxiety (described in the introduction, section 1.8). A stepped care framework is followed, offering evidence based low- and high-intensity treatments that primarily differ in terms of therapist involvement and number of sessions. The North London study analysed data from patients who received high-intensity therapy between 2008 and 2013 (Saunders et al., 2019). Four classes of depression symptom trajectories were identified, and five classes of anxiety symptom trajectories. Compared with trajectory classes showing improvement, classes with initially moderate-severe symptoms and limited change over time were associated with higher baseline levels of functional impairment, depression, anxiety, and phobia symptoms. No associations were found with age, gender, ethnicity, or reporting prescribed psychotropic medication. In the present study, we extended these findings in a larger and more ethnically diverse sample of IAPT patients and considered additional baseline variables to test for association. We also used more recent data which is likely of higher quality due to improvements in the recording of information in IAPT (Community and Mental Health Team, 2014; Saunders, Cape, et al., 2020).

4.3 Methods

4.3.1 Sample

The sample came from anonymised patient treatment records of routinely collected data from the four IAPT services of the South London and Maudsley (SLaM) NHS Foundation Trust. Using the Clinical Record Interactive Search (Stewart et al., 2009), we extracted treatment records of patients who had received psychological therapy for symptoms of depression or anxiety. Patients had started and ended treatment between July 2014 and September 2020, irrespective of the reason for ending. From the extracted records of 115,304 patients, we identified a high-intensity treatment sample of 16,258 patients and 110,773 sessions suitable for analysis (see Figure 1 and Supplementary Information 1). IAPT high-intensity therapies are formal psychological treatments consistent with National Institute for Health and Care Excellence guidelines (NICE, 2011), including disorder-specific cognitive-behavioural therapy (CBT; Roth & Pilling, 2008) and counselling for depression

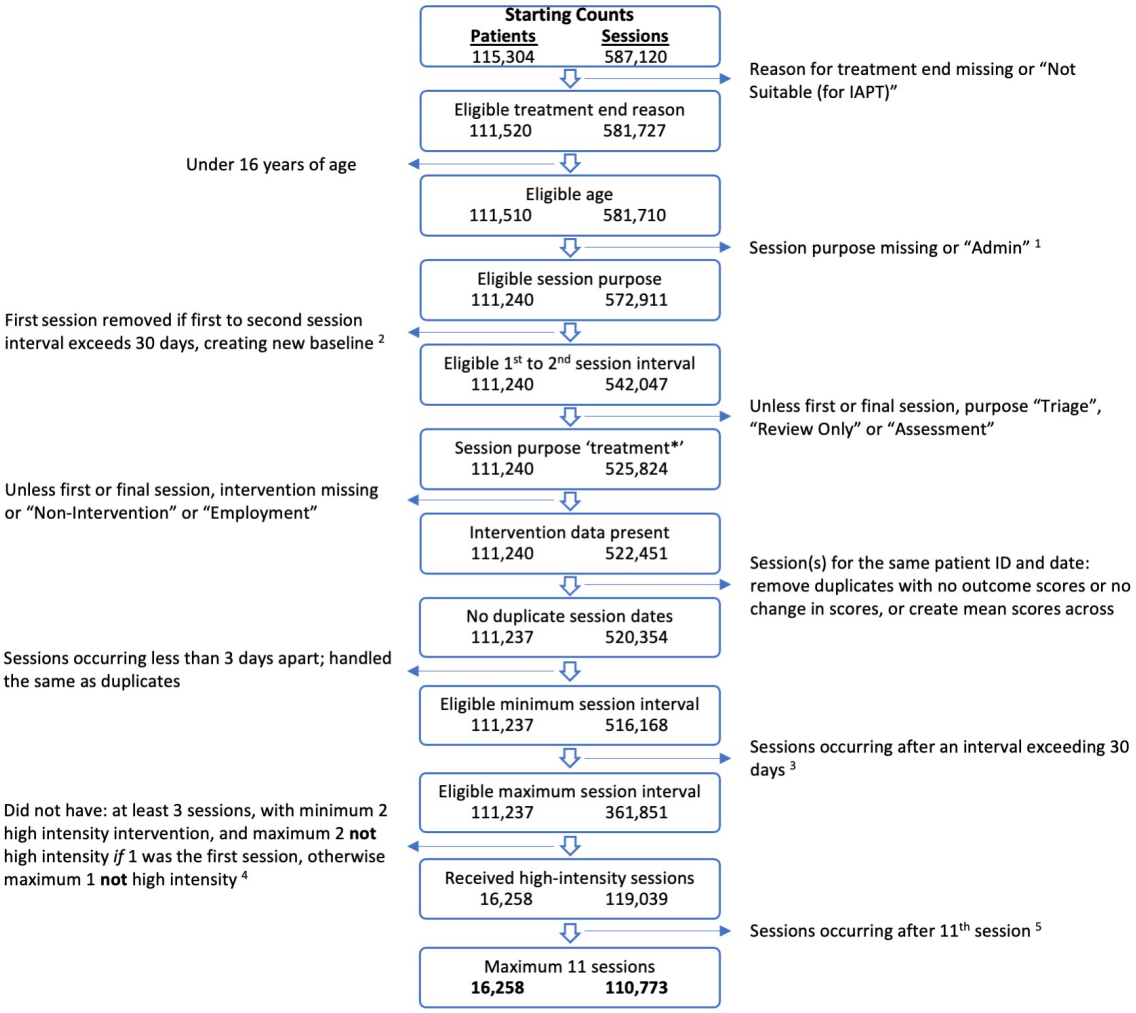


Figure 1. Flowchart detailing the exclusions applied to retain patient data suitable for analysis

[1] Record of administrative activities; not treatment or assessment sessions with outcomes. [2] The first session in IAPT is often assessment or triage, with treatment beginning formally at the second session. If the first to second session interval exceeded 30 days, the second session was used as the baseline. This was not repeated, as therapy with additional long intervals between the 'new' baseline and second session was considered too disrupted for stable conclusions.[3] Growth mixture modelling assumes invariant time intervals across patients. High-intensity treatment is usually weekly but the data's naturalistic nature introduced substantial variance. To limit this, sessions after a 30-day interval were excluded. Supplementary Table 2 presents session interval descriptives with and without this filter. [4] Limited to high-intensity therapies due to differences in the average number of sessions and level of structure between intensities. However, most IAPT patients receive at least one low-intensity session due to a stepped care format, and the baseline assessment/triage session may also be labelled low intensity. [5] Including baseline assessment i.e., 10 treatment sessions. Session number descriptives before and after are reported with Supplementary Table 1.

(see the IAPT manual for details of therapy protocols and techniques; NCCMH, 2021). We modelled up to 11 time points; a baseline assessment and 10 treatment sessions. This ensured sufficient complete data across time points for the analysis and avoided modelling more sessions than a substantial proportion of patients had data for (Supplementary Table 1), which could make optimal solutions invalid and unstable (Lutz et al., 2005). To determine support for generalisability of the results, patients included in the analytical sample were compared to excluded patients using suitable group difference tests. The large sample size could result in statistical significance of negligible differences, therefore we primarily considered effect sizes. To be representative of patients receiving treatment, we did not exclude patients scoring below clinical thresholds on the symptom scales.

4.3.2 Measures

Outcomes were self-report questionnaires administered at each session. Current symptoms of depression and anxiety were measured, respectively, by the Patient Health Questionnaire 9-item version (PHQ9; Kroenke et al., 2001; range 0-27; case threshold in IAPT ≥10) and the Generalized Anxiety Disorder 7-item scale (GAD7; Spitzer et al., 2006; range 0-21; IAPT case threshold ≥8). Reliable improvement in a trajectory class was defined as a decrease of six or more points on the PHQ9, and at least four on the GAD7 (Gyani et al., 2013; NCCMH, 2021). The selection of baseline patient characteristics to investigate for associations with trajectory class membership was informed by the available data and previously reported associations with treatment outcomes (Delgadillo, Dawson, et al., 2017; Delgadillo et al., 2016; Robinson, Kellett, et al., 2020; Saunders et al., 2019). These were: total scores from each of the PHQ9 and GAD7 (when not the outcome modelled), Work and Social Adjustment Scale (WSAS; functional impairment measure; range 0-40; Marks, 1986; Mundt et al., 2002), age (in 10 year blocks), gender (female, male), ethnicity (White, Black, Asian, Mixed, other), employment (employed, unemployed, non-worker (e.g., retired, student)), psychotropic medication (not prescribed, prescribed) and disability (no, yes). In IAPT, 'problem descriptors' are used to indicate the disorder that is the agreed focus of treatment, in line with NICE guidance (NICE, 2011). Problem descriptors are based on ICD-10 diagnostic codes, but as noted in the introduction (section 1.8), their use does not mean that a patient necessarily met all diagnostic criteria for the disorder or that they

would not meet criteria for other disorders. As such, comorbidity of disorders was unknown. Problem descriptor categories were depression, generalised anxiety disorder (GAD), PTSD, adjustment disorder, obsessive-compulsive disorder (OCD), mixed anxiety and depressive disorder, panic/phobia (panic disorder, agoraphobia, social phobia, specific phobia), and 'other' which included infrequent descriptors such as somatoform disorder. Despite showing strong associations in an existing study (Saunders et al., 2019), phobia scales were not included due to high, potentially non-random, missingness.

4.3.3 Statistical analyses

4.3.3.1 Trajectory class models

We estimated separate growth mixture models (GMMs) for symptoms of depression and anxiety. To first determine the single trajectory form which best fit the observed data overall we compared linear (Bone et al., 2021), quadratic (Saunders et al., 2019; Sunderland et al., 2012) and negative log-linear (base 10; Lutz et al., 2009; Lutz et al., 2014) trajectories ('latent growth curves'). We then incrementally modelled up to six latent trajectory classes using GMM, with each trajectory specified to the optimal form identified in the first step. This revealed whether the observed patient symptom trajectories were homogeneous and therefore well-represented by the single latent growth curve, or if differences between symptom trajectories were better captured by multiple latent classes. The upper limit of six classes was informed by past studies of trajectories during psychological therapy, which suggest between three and five classes (Lutz et al., 2009, 2014; Saunders et al., 2019). Thus, the unlikely six class model acted as a ceiling to test against. To determine the optimum number of classes, we compared each estimated model to a model with one fewer class. These comparisons were based on several considerations. We used a number of fit indices, with a preference for the Bayesian Information Criterion (BIC; Supplementary Information 2). As BIC does not always show a minimum value within a reasonable range of models, elbow plots were created to aid identification of the point of diminishing gains from the addition of a class. We also favoured models that were clinically interpretable, and reasonable in terms of theory and existing literature. This included favouring parsimony, for example, if a model differed by an additional trajectory class distinctive only by the intercept and not the change pattern,

the model with fewer classes was preferred. Class segregation is imperfect, therefore patients have a likelihood ('posterior probability') of belonging to each class. To be clinically meaningful and statistically stable, we opted for models where each class contained at least 1% of the sample, based on allocating patients to their most likely trajectory class (Jung & Wickrama, 2008). Finally, we required models to converge and for estimated values to be within the range of the outcome measure. We described patients in each class and the overlap of patients between depression and anxiety classes.

The most simplified version of GMM is called 'latent class growth analysis' (LCGA). LCGA assumes that all patients in a class follow exactly the same trajectory and as such the variance in the intercept and slope are zero. Although the assumption is often unrealistic, and can result in trajectory classes that differ only in terms of starting score, it is recommended to perform this as a useful and less computationally demanding starting point (Jung & Wickrama, 2008; B. Muthén & Muthén, 2000; van de Schoot et al., 2017). We compared this simpler model to a GMM where the variance in each intercept was freed such that it could have a non-zero value but was constrained to be equal between classes, whilst the slope variance was fixed at zero. This forced differences to be revealed in the patterns of change over time (slope of each class) rather than initial symptom levels (intercept of each class). This specification has successfully been used in previous symptom based GMMs (e.g., Lutz et al., 2014). Latent growth curve and LCGA results were similar between IAPT service-specific models (Supplementary Figure 1) therefore analyses were performed across all four IAPT services. We used Mplus version 8.3 (L. K. Muthén & Muthén, 2017), alongside R version 3.6.3 (R Core Team, 2021) and the MplusAutomation package (Hallquist & Wiley, 2018). For greater technical detail of the procedure and guidance on interpreting fit indices, see Supplementary Information 2.

4.3.3.2 Missing data

Missing data in the outcomes was handled using full information maximum likelihood with robust standard errors (Mplus option 'MLR') for non-normal distributions (Supplementary Figure 2). If a patient had no observed data for an outcome they were omitted from the model of that outcome.

4.3.3.3 Associations of baseline variables with trajectory class

Following identification of the optimal number of classes, multinomial regressions were performed to test for associations between variables measured at the start of treatment and trajectory class membership. Explorations of missing and complete data between baseline variables did now show any association patterns. To handle missing data in the baseline variables and maximise sample size, we performed multiple imputation using the 'mice' package in R (van Buuren & Groothuis-Oudshoorn, 2011). We specified 20 imputed datasets and 20 iterations. The multinomial regressions were performed with each multiply imputed dataset, and estimates were pooled. Patients' most likely class was the regression outcome. IAPT service was included as a covariate. We assessed statistical significance using a Bonferroni adjusted *p*-value threshold of $p < 0.025$ to account for the two independent models.

4.4 Results

4.4.1 Sample characteristics

Sample characteristics for the 16,258 patients are presented in Table 1. The majority had received at least one session of CBT (50%) or counselling (49%); a small proportion received other treatments (6%) such as interpersonal therapy. The average number of sessions received, after limiting to a maximum of 11 including the baseline assessment, was 6.8 (SD = 2.9); 12% received only the minimum 3 sessions and 18% received 11 (Supplementary Table 1). At each session, over 99% of the patients remaining in treatment had complete PHQ9 and GAD7 scores (Supplementary Table 3). Patients in the analytical sample were similar to excluded patients, besides differences that reflected inclusion criteria (e.g., number of sessions; Supplementary Table 4).

Table 1. Baseline characteristics of patients who received high-intensity psychological therapy for symptoms of depression and anxiety (N = 16,258)

Variable		Mean (SD; range) or Count (proportion %)
Age (years)		37.55 (13.36; 16 - 94)
Gender	Male	5262 (32%)
ref: Female	*Missing*	*17 (0.1%)*
Depression symptoms (PHQ9)		13.98 (6.39; 0 - 27)
	Missing	*153 (0.9%)*
Anxiety symptoms (GAD7)		12.53 (5.39; 0 - 21)
	Missing	*154 (1%)*
Case on PHQ9 and/or GAD7 [1]	Yes	13735 (84%)
	No	2368 (15%)
	Missing	*155 (1%)*
Functional impairment score (WSAS)		17.58 (9.31; 0 - 40)
	Missing	*5447 (34%)*
Problem descriptor [2]	Depression (ref)	6703 (41%)
	Other	1423 (9%)
	GAD	1393 (9%)
	Adjustment disorder	1320 (8%)
	PTSD	1132 (7%)
	MADD	1129 (7%)
	Panic/phobia	1003 (6%)
	OCD	550 (3%)
	Missing	*1605 (10%)*
Psychotropic medication	Prescribed	5545 (34%)
ref: Not prescribed	*Missing*	*736 (5%)*
Ethnicity	White (ref)	9789 (60%)
	Black	2964 (18%)
	Mixed	1111 (7%)
	Asian	961 (6%)
	Other	557 (3%)
	Missing	*876 (5%)*
Employment status	Employed (ref)	10033 (62%)
	Unemployed	3572 (22%)
	Non-worker [3]	2222 (14%)
	Missing	*431 (3%)*
Disability	Yes	1575 (10%)
ref: No	*Missing*	*0%* [4]

Number of sessions (including baseline assessment)		6.81 (2.91; 3 - 11)
Recovered [5]	Yes	7028 (43%)
	Missing	*2737 (17%)*
Reason for end of treatment	Discharge	13138 (81%)
	Dropout	2328 (14%)
	Referral to another service	792 (5%)
Service	0	7027 (43%)
	1	3402 (21%)
	2	3244 (20%)
	3	2585 (16%)

Baseline values are presented for variables measured at multiple time points. [1] Case thresholds: PHQ9 ≥10, GAD7 ≥8. [2] Indicates the disorder treated: GAD = generalised anxiety disorder; PTSD = post-traumatic stress disorder; MADD = mixed anxiety and depressive disorder; panic/phobia = e.g., panic disorder, social phobia; OCD = obsessive-compulsive disorder; Other = e.g., somatoform disorder. [3] Non-worker = e.g., retired. [4] No negative responses recorded; the absence of any value was counted as a negative response rather than missing. [5] Only calculated for patients above PHQ9/GAD7 case threshold at baseline and with observed scores for their final session, otherwise 'missing'. Represents recovery within the 10 treatment sessions; if patients received more sessions and then recovered, they would appear unrecovered here. For categorical variables included in the regressions, ref = reference category (proportion not shown if binary).

4.4.2 Growth mixture models

For both depression and anxiety symptoms, the best-fitting latent growth curve form was quadratic (Supplementary Information 3). On average, patients' symptoms were initially moderate (Kroenke et al., 2001; Spitzer et al., 2006) and improved steadily from baseline (session 0) over approximately the first five treatment sessions, showing reliable improvement by session eight. For latent class growth analysis, a four-class model was selected as the optimum solution for each outcome (Supplementary Information 4). The trajectories were largely uninformative as they primarily differed in terms of baseline severity (intercept). We therefore estimated growth mixture models (GMMs) with free but equal variance in the intercepts; these are our focus.

A four-class GMM was selected for depression symptoms, and also for anxiety symptoms. This decision was based on fit indices, with focus on the BIC and other considerations such as previous literature (see Table 2 and Figure 2 for fit indices, and Supplementary Information 5 - 8 for further model selection information, model descriptives and class-based sample descriptives).

Table 2. Fit indices for growth mixture models of depression symptoms (PHQ9; top) and anxiety symptoms (GAD7; bottom) during psychological therapy (N = 16,258)

Depression symptoms GMM	Parameters	AIC	BIC	Entropy	VLMR LRT *p*-value	Individuals per class (%)
Growth Curve	25	613935	614128	NA	NA	100
Two Class	29	610886	611109	0.538	< 0.001	79.1, 20.9
Three Class	33	608940	609194	0.593	< 0.001	52.3, 22.7, 24.9
Four Class	37	607920	608204	0.600	< 0.001	13.5, 52.5, 17.6, 16.4
Five Class	41	607573	607888	0.606	0.0017	10.1, 20.9, 16.4, 2.2, 50.5
Six Class	45	607186	607532	0.632	0.0001	2.9, 9.4, 44.4, 25.5, 2.5, 15.3
Anxiety symptoms GMM	**Parameters**	**AIC**	**BIC**	**Entropy**	**VLMR LRT *p*-value**	**Individuals per class (%)**
Growth Curve	25	595485	595677	NA	NA	100
Two Class	29	592459	592682	0.453	< 0.001	29.6, 70.4
Three Class	33	590113	590367	0.602	< 0.001	31.3, 40.5, 28.2
Four Class	37	588890	589175	0.591	< 0.001	24.8, 21.8, 11.9, 41.5
Five Class	41	588523	588838	0.635	< 0.001	11.7, 1.7, 40.7, 22.3, 23.6
Six Class	45	588228	588574	0.609	0.0002	34.0, 6.3, 17.5, 3.0, 19.2, 19.9

AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, VLMR LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. Proportions are based on a patient's most likely class. Lower values on AIC and BIC indicate better fit. Higher entropy represents better distinction. Significant VLMR LRT p-value indicates the model is a better fit than one with one fewer class.

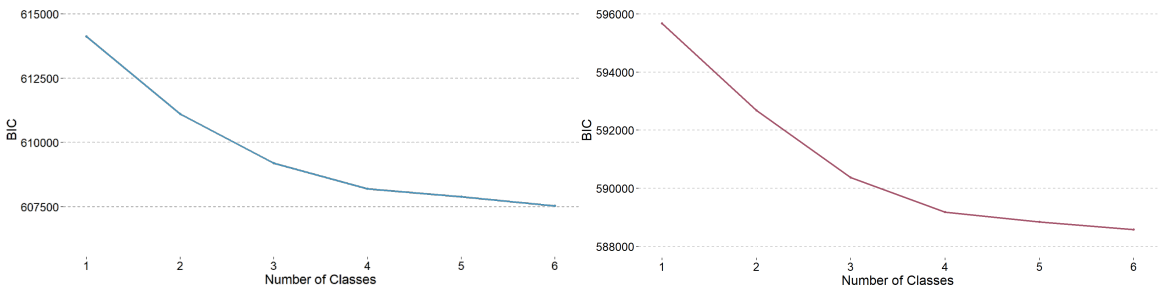


Figure 2. Elbow plot of Bayesian Information Criterion (BIC) values for growth mixture models of depression symptoms (left) and anxiety symptoms (right)

Figure 3 presents the mean estimated trajectories of the four classes for each outcome. The trajectory classes were very similar for each model, therefore we describe them with the same labels. Three classes demonstrated moderate-severe symptoms at baseline. One of these showed no change (moderate-severe plateau; grey diamonds), the second showed steady improvement (moderate-severe, gradual improvement; pink squares) and the third had fast improvement which plateaued after the sixth treatment session (moderate-severe, fast improvement; blue triangles). A fourth class showed mild symptoms and minimal improvement up to the fourth session (mild, small improvement; green circles).

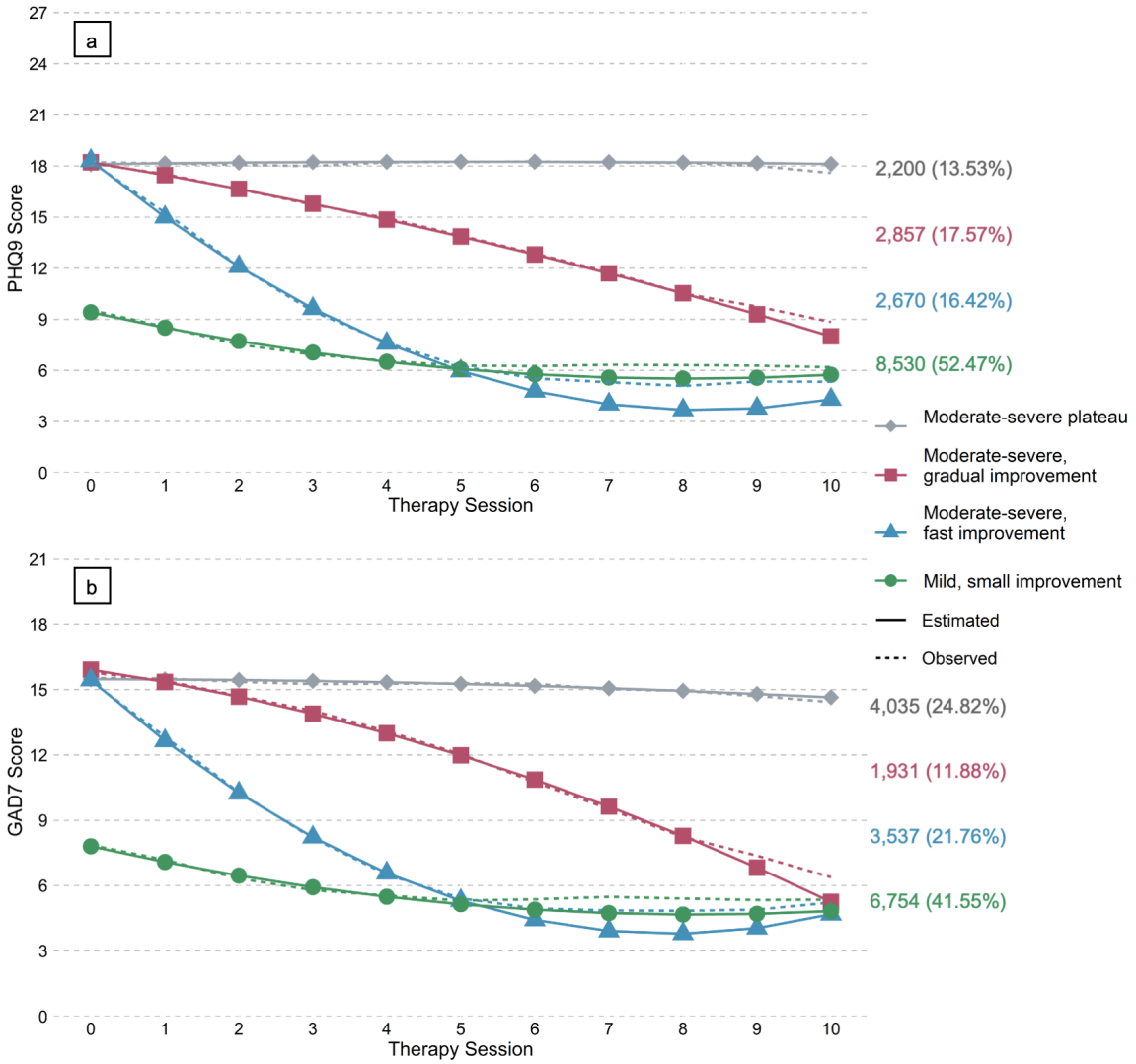


Figure 3. Four-class growth mixture model of a) depression symptoms (PHQ9) and b) anxiety symptoms (GAD7), during high-intensity psychological therapy (N = 16,258)

This figure shows the model estimated mean and observed mean trajectories of each class. Patients had a likelihood of belonging to each trajectory class; counts and proportions (%) are based on their 'most likely' class membership.

4.4.2.1 Depression symptoms growth mixture model

Approximately half the patients were assigned to the mild class and 14-18% in each of the other classes. Reliable improvement occurred on average by the seventh treatment session in the gradual class, and the second in the fast class, but was not observed in the plateau or mild class. At the final treatment session, almost no patients (< 1%) in the moderate-severe plateau class had 'recovered' to below PHQ9 and GAD7 case thresholds used in IAPT (< 10 and < 8, respectively). Recovery was also low in the gradual improvement class (19%), but higher in the fast improvement and mild classes (68% and 59%). Dropout was higher in the plateau and gradual improvement classes (19% and 20%) than the fast improvement and mild classes (15% and 11%). The plateau class had the highest proportion referred to other services (12%). Full descriptives are in Supplementary Information 6.

4.4.2.2 Anxiety symptoms growth mixture model

More patients in the anxiety GMM were assigned to the moderate-severe plateau and moderate-severe, fast improvement classes (25% and 22%, respectively), than in the depression model. Fewer were in the moderate-severe, gradual improvement and mild, small improvement classes (12% and 42%). Reliable improvement occurred on average by the sixth treatment session in the gradual class and the second session in the fast class but, as in the depression model, was not observed in the plateau or mild class. Similarly, most patients who followed a plateau or gradual improvement trajectory had not recovered by the tenth treatment session (1% and 35% recovered, respectively), whilst the majority of the fast improvement and mild classes had (69% and 64%). Again, the plateau class had the highest proportion of dropout (23%) and onward referrals (9%). Model and class-based sample descriptives are in Supplementary Information 8.

4.4.2.3 Overlap of trajectory class membership between the selected models

Table 3 shows the overlap of patients' trajectory classes between the depression and anxiety symptom models. For example, 12% of the overall sample were most likely to belong to the moderate-severe plateau class for both outcomes. This included 89% of patients in the depression plateau class who were in the anxiety plateau class, and 48% of patients from the anxiety plateau class who were in the depression plateau class.

Table 3. Overlap of patients' most likely class for four-class growth mixture models of depression (PHQ9) and anxiety (GAD7) symptoms

a) Overlap of class membership across outcomes; values are proportions of the total sample

	Moderate-severe plateau	Moderate-severe, gradual improvement	Moderate-severe, fast improvement	Mild, small improvement
Moderate-severe plateau	12%			
Moderate-severe, gradual improvement	8%	7%		
Moderate-severe, fast improvement	1%	4%	11%	
Mild, small improvement	6%	5%	12%	36%

b) Values are proportions of patients from the depression class (row) who were assigned to the anxiety class (column)

	Anxiety Class			
Depression Class	**Moderate-severe plateau**	**Moderate-severe, gradual improvement**	**Moderate-severe, fast improvement**	**Mild, small improvement**
Moderate-severe plateau	89%	4%	1%	6%
Moderate-severe, gradual improvement	40%	38%	11%	12%
Moderate-severe, fast improvement	7%	11%	66%	17%
Mild, small improvement	9%	6%	17%	68%

c) Values are proportions of patients from the anxiety class (row) who were assigned to the depression class (column)

	Depression Class			
Anxiety Class	**Moderate-severe plateau**	**Moderate-severe, gradual improvement**	**Moderate-severe, fast improvement**	**Mild, small improvement**
Moderate-severe plateau	48%	28%	4%	19%
Moderate-severe, gradual improvement	5%	56%	15%	25%
Moderate-severe, fast improvement	1%	9%	50%	41%
Mild, small improvement	2%	5%	7%	86%

For panel (b), reading along a row indicates the distribution of most likely anxiety classes for all patients who were most likely to belong to the depression class named on the row. Panel (c) presents the converse information.

4.4.3 Associations of baseline variables with trajectory class

Regression results are reported in terms of odds ratios (ORs), compared with the reference trajectory class moderate-severe plateau, which was the least 'favourable' trajectory class (see Figure 4, or for precise values use Supplementary Tables 5 and 6). Therefore, a baseline variable with an OR of 1 had no bearing on trajectory class membership. An OR between 0 and 1 indicated a negative relationship, meaning that an increase in the baseline variable was associated with lower odds of being in the specified trajectory class than the reference trajectory (i.e., more likely to be in the *least favourable*, moderate-severe plateau class). An OR greater than 1 indicated higher odds of being in the specified trajectory than the reference trajectory (i.e., more likely to be in one of the three *more favourable* classes; moderate-severe, gradual improvement; moderate-severe fast improvement; mild, small improvement). ORs of categorical variables are interpreted in comparison to a reference category. For example, for the depression model, a patient with a problem descriptor of OCD had 0.52 the odds of being in the fast improvement than the plateau class, *compared with a patient with depression*.

Across depression and anxiety symptom models, patients who were unemployed compared with employed had lower odds of being in any of the more favourable classes than the reference class, moderate-severe plateau. Furthermore, higher functional impairment scores and non-worker status (e.g., retired), were associated with lower odds of being in the moderate-severe, fast improvement or mild, small improvement trajectory classes. However, these variables were not associated with significantly different odds of the moderate-severe, gradual improvement class.

Reporting a disability showed lower odds of being in any of the more favourable classes in the depression model, and of the gradual or fast improvement classes in the anxiety model. Prescribed psychotropic medication was associated with lower odds of the fast improvement or the mild, small improvement class in the depression model, and of any favourable trajectory class in the anxiety model. Higher scores on the other symptom scale (e.g., GAD7 for the depression model) were associated with lower odds of the mild trajectory class for both models, and additionally with lower odds of the fast improvement

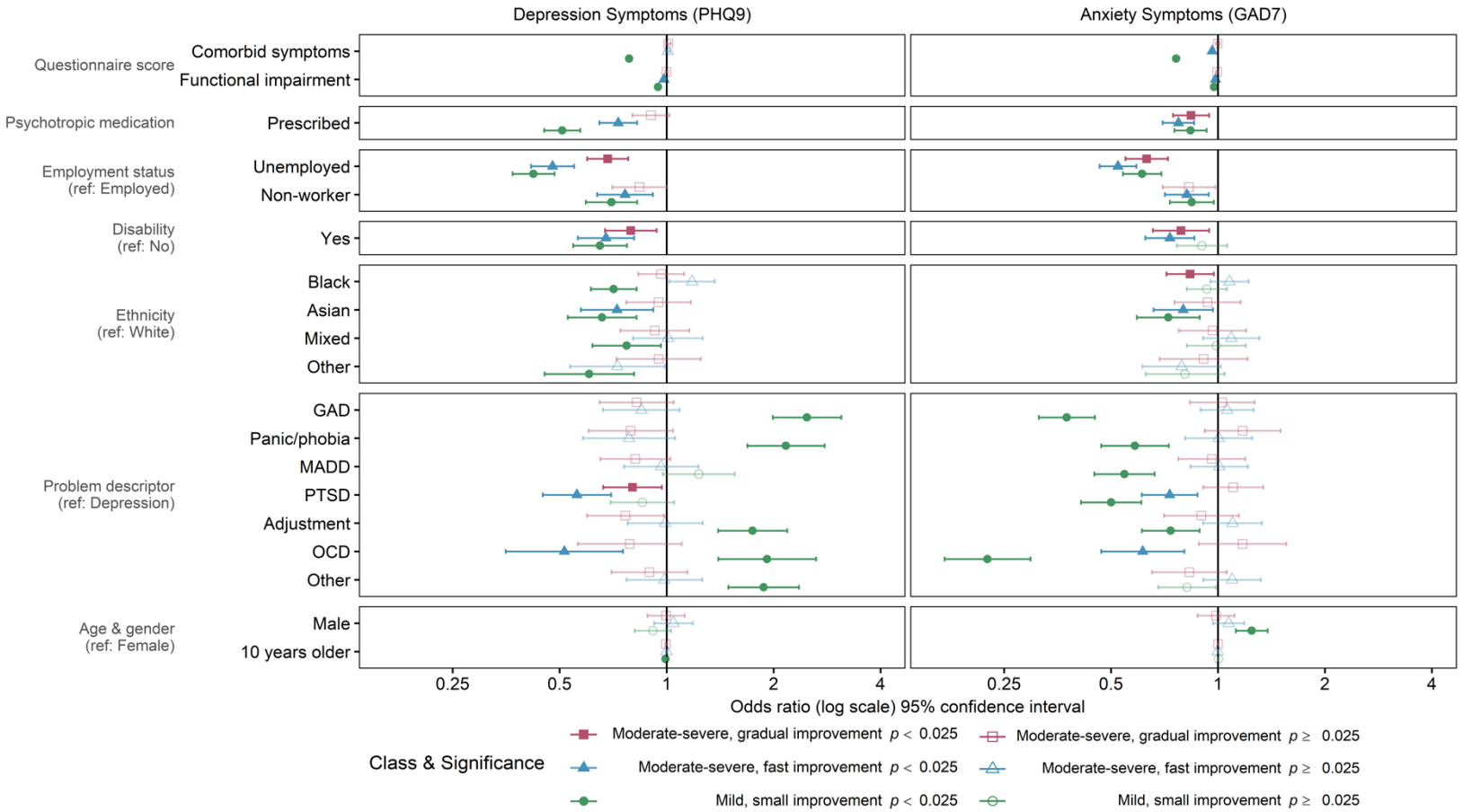


Figure 4. Multinomial regressions of baseline variables and trajectory class membership for a four-class growth mixture model of depression symptoms and a four-class growth mixture model of anxiety symptoms during high-intensity psychological therapy (N = 16,258)

Reference class = moderate-severe plateau. Service where the patient received treatment was a covariate with four categories. PHQ9 = depression symptoms, GAD7 = anxiety symptoms. Comorbid symptoms = GAD7 in the depression model, PHQ9 in the anxiety model. Functional impairment = WSAS. Non-worker = e.g., retired, student. GAD = generalised anxiety disorder; Panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; MADD = mixed anxiety and depressive disorder; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; Other = e.g., somatoform disorder.

class in the anxiety model. In the depression symptoms model, compared with White ethnicity patients, patients who identified as Black, Asian, multi-ethnic ('mixed'), or 'other' ethnicity were less likely to follow the mild trajectory than the moderate-severe plateau. Patients identifying as Asian also had lower odds of the fast improvement class. In the anxiety model, the only similarities were the associations observed with Asian ethnicity.

The main differences between models were in their associations with problem descriptor. In the depression model, compared with patients with a depression problem descriptor, patients with any other descriptor, besides PTSD or MADD, were more likely to be in the mild class. The opposite pattern was observed in the anxiety model, such that all descriptors besides 'other' were associated with lower odds of the mild class than the plateau class. However, in both models, OCD and PTSD were associated with lower odds of the fast improvement class. PTSD was additionally associated with lower odds of the depression model gradual improvement class. Age and gender had no notable associations, except in the anxiety model where patients reporting male gender were more likely to be in the mild class than females.

4.5 Discussion

4.5.1 Overview

We identified patterns of change in depression and anxiety symptoms during high-intensity psychological therapy among 16,258 patients. For both outcomes individual differences in the data were best explained by four classes of trajectories. Overall, some baseline variables were generally associated with greater likelihood of the no change plateau trajectory class than any more favourable class. These included being unemployed, reporting a disability or prescribed medication. Other baseline characteristics differentiated between the more favourable classes, including functional impairment, non-worker status, ethnicity and specific problem descriptors.

4.5.2 Growth mixture models

The depression model and the anxiety model each had three classes with moderate-severe symptom scores at baseline. Using only intake symptom severity would therefore not allow a clinician to distinguish between three very different trajectories. One of these trajectories showed no change, labelled moderate-severe plateau. The other two classes reliably improved: moderate-severe, gradual improvement changing more steadily than moderate-severe, fast improvement. The fourth class, mild, small improvement, had average baseline symptoms below clinical thresholds, and thus less capacity for large improvements. These trajectory classes were broadly consistent in number and shape with a previous IAPT study (Saunders et al., 2019), although the optimal anxiety model selected in that analysis had a fifth class representing another, less severe, plateau trajectory. These similar findings occurred despite several sample and methodological differences of our study, including a greater proportion of patients receiving counselling and fewer receiving CBT. This perhaps reflects meta-analytic evidence of equivalency between therapies (Cuijpers et al., 2021).

The gradual improvement class may be particularly clinically relevant as the apparent lack of early response could lead to premature alteration or termination of treatment. Symptom change was initially indistinguishable from the moderate-severe plateau trajectory and only showed reliable improvement after six or seven treatment sessions. Recovery by the final session was lower in the gradual than the fast improvement class, especially for the depression model. This was reflected in the class overlap; more patients in the depression gradual class did not experience improvement in anxiety symptoms, as shown by a higher proportion belonging to the anxiety plateau class. The gradual improvement class might represent patients who require a high number of treatment sessions to show recovery. Supporting this, a study of high-intensity CBT found that most patients who significantly improved did so within 14 sessions (Robinson, Kellett, et al., 2020). Furthermore, a systematic review of therapy dose recommended up to 26 sessions for patients who have improved yet not recovered before this point (Robinson, Delgadillo, et al., 2020). In the present sample, relatively few patients had received more than ten treatment sessions and therefore trajectories beyond this point were not modelled.

The moderate-severe plateau class may also represent patients who would benefit from more treatment, although some 'non-responders' would probably not recover even after a very high number of sessions (Howard et al., 1986). Patients likely to belong to this group might therefore be ideal candidates for clinical trials of novel therapies or additional support services. Notably, the frequency of onward referrals in this class suggested the presence of co-occurring problems that required specialist services. That said, despite limited symptom change, some patients may have experienced other, unmeasured, benefits such as prevention of deterioration or hospitalisation. More patients belonged to the plateau class in the anxiety as compared to the depression model. This may be driven by disorders such as OCD and PTSD, which on average require a greater dose of therapy to observe improvement than most depression and anxiety disorders (Robinson, Kellett, et al., 2020), and are better detected by the GAD7 than the PHQ9 (Kroenke et al., 2007). However, class overlap indicated that slightly over half of the anxiety plateau class showed mild or improving depression symptoms, whilst most patients in the depression plateau class showed no improvement on either measure.

Patients in the fast improvement class generally appeared to require fewer than ten treatment sessions. They were below case thresholds after an average of four treatment sessions and had the highest recovery rate, consistent with a meta-analysis reporting that reliable symptom improvement by the fourth session predicts recovery (Beard & Delgadillo, 2019). Although it is important to continue treatment for several sessions after remission to help prevent false positives or relapse (Robinson, Kellett, et al., 2020), this finding warrants further investigation as it could have implications for service efficiency.

The mean initial symptom score estimated by the model for the mild, small improvement trajectory class, for both depression and anxiety models, was below case threshold, yet all patients had received high-intensity therapy. However, it would be inappropriate to conclude that high-intensity treatment was unwarranted. Scores varied around these means and many patients met the threshold on the *other* symptom measure; the majority of mild class patients were an observed case on one or both measures. These patients may also have had additional symptoms not assessed by the PHQ9 or GAD7.

4.5.3 Associations of baseline variables with trajectory class

Several consistent associations with trajectory class membership were being unemployed, reporting a disability or prescribed medication, each of which was generally associated with lower odds of being in any of the more favourable classes compared with moderate-severe plateau. These associations are informative as they discriminated between individuals with similarly high baseline scores who improved or not. The prior IAPT trajectory study found no association with medication (Saunders et al., 2019) despite evidence of a relationship from studies of endpoint treatment outcomes (Robinson, Kellett, et al., 2020). The association with unemployment was consistent with findings from endpoint outcome studies (Delgadillo et al., 2016; Delgadillo, Huey, et al., 2017). Neither employment or disability were included in the prior trajectory study (Saunders et al., 2019), and the present findings highlight the importance of routinely recording them. These factors may negatively impact therapy response and acknowledging and supporting patients who are experiencing them might reduce the likelihood of no improvement. This supports the argument for psychological services working closely with employment advisors, as per the IAPT manual (NCCMH, 2021). Similarly, individuals with a PTSD or OCD problem descriptor may benefit from tailored treatment in the form of a greater number of sessions, as indicated by existing literature (Robinson, Kellett, et al., 2020). PTSD and OCD were the only problem descriptors associated with lower odds of fast improvement, and also, for PTSD, of gradual improvement in depression symptoms. Improvement may have been observed after a greater number of sessions than we were able to model or using measures more sensitive to symptoms of PTSD and OCD.

Another consistent finding was that higher functional impairment was associated with lower odds of the fast improvement or mild class than the plateau class, but not the gradual improvement class. This pattern of associations was also observed for non-worker status (e.g., retired) and, in the anxiety model, higher baseline depression symptoms. These variables may therefore be useful for discriminating between fast and gradual improvement for patients with similar baseline scores. For example, indicating that any improvement that does occur is likely to be gradual. The effect sizes for functional impairment were smaller than those for symptom scores, yet still notable given they

reflected the association for a one-unit difference on the WSAS, which has a wide scoring range.

Patients who identified as one of the minoritised ethnic groups rather than White ethnicity were less likely to follow the mild depression trajectory, indicating that they had more severe symptoms at intake. These findings are consistent with lower recovery rates in national IAPT data and indicate that some patients from minoritised ethnic groups may benefit from culturally-adapted treatment (Beck et al., 2019). The ethnic diversity of the sample allowed us to study ethnicity in more detail than the binary variable used in similar studies. This revealed that lower odds of the mild trajectory were also observed in the anxiety model for Asian patients, as well as lower odds of fast improvement. That notwithstanding, there was insufficient power to analyse associations with ethnicity at an even greater level of granularity and this will require further research efforts.

The Saunders et al. (2019) trajectory study similarly reported that higher symptom and functional impairment scores were associated with greater odds of the corresponding moderate-severe plateau class. Our results were also consistent with a finding that patients reporting higher baseline scores on the PHQ9, GAD7, WSAS, prescribed antidepressants, a disability or unemployment were less likely to improve with an increasing number of sessions, similar to our plateau class (Robinson, Kellett, et al., 2020).

4.5.4 Strengths and limitations

A large sample size produced more accurate estimates and permitted a broader range of ethnicity categories and baseline variables, such as employment and disability, compared with previous trajectory studies. Furthermore, modelling a greater number of sessions revealed trajectory classes that did not show improvement until later in treatment. Associations were identified with important and easily recorded factors, which could help to identify likely symptom trajectories during an initial assessment. However, the results may be confounded by the presence of additional variables indicative of treatment prognosis, such as chronicity (Kessler et al., 2017). Whilst routinely collected data may generalise to clinical settings more readily than clinical trial data, the provision of different

interventions, dependent on problem descriptor, makes IAPT data more complex to analyse. Investigations of treatment and descriptor-specific trajectories were beyond the scope of this study. The inclusion of patients with a depression descriptor in the anxiety symptoms model, and vice versa, may therefore have contributed to the high proportion in the mild trajectory classes. That notwithstanding, the assessment of transdiagnostic symptoms appears particularly important for treatment outcomes (O'Driscoll et al., 2021) so we chose to include all patients in both models. Furthermore, the problem descriptor only represents the condition being treated. This means that, although comorbidity is common between depression and anxiety, the extent of comorbidity was unknown in the current sample. The reliability and validity of IAPT problem descriptors has not been established. However, clinicians are trained to use ICD-10 to determine them, and in services with a higher frequency and accuracy of providing problem descriptors, patients show better outcomes (Clark et al., 2018; Saunders, Cape, et al., 2020). Another issue was dropout. Different reasons may underlie dropout, including dissatisfaction with treatment or treatment ending due to recovery before the clinically recommended number of sessions. Regardless, and in spite of the use of missing data methods, scores at later sessions were based on remaining individuals. Overall, our findings offer a broad representation of what we expect to see in routine clinical settings.

In terms of ethnicity, although generally representative of the local population that SLaM IAPT serves (ONS, 2012), the patient sample was more ethnically diverse than IAPT nationally. The two most frequently reported ethnicities were White (60%) and Black (18%), and in IAPT national data are White (80%) and Asian (6%) (NHS Digital, 2021). It is critical that mental health and treatment outcome research is applicable for all patients, and oversampling from minoritised groups can help to achieve this. In terms of ethnicity, this overrepresentation occurred naturally within our study; the diversity of this sample is a strength.

Finally, the GAD7 does not cover core symptoms for disorders such as PTSD and OCD. However, it does show reasonably high specificity and sensitivity for several disorders, including these (Kroenke et al., 2007), and therefore provides some indication of likely trajectories for them. Future studies should aim to model trajectories of disorder specific

symptoms, and also functional impairment, which is particularly important to patients and potentially improves later in treatment than symptoms (Howard et al., 1993; Zimmerman et al., 2006). This was not possible in the present sample due to the extent of missingness on these variables. Modelling a higher number of sessions than was available here would also be beneficial, to reveal more about patients who require a greater dose of therapy.

4.5.5 Implications and conclusions

A longitudinal person-centred analysis allowed us to identify substantial heterogeneity in depression and anxiety symptom change during psychological therapy, which could be explained by four classes of trajectories. The identified trajectory classes were differentially associated with baseline variables such as employment status and functional impairment. To be implemented clinically, these findings need to be replicated and validated to produce a predictive tool that outputs a patient's most likely symptom trajectory by combining potentially conflicting information across baseline variables. This could be used to inform treatment plans at the start of therapy and also monitor whether patients are 'on-track' according to their predicted trajectory, which can improve patient outcomes (Delgadillo et al., 2018). This would be especially relevant for patients in the gradual improvement class who on average did not show reliable improvement until session six or seven. This study is therefore a crucial first step towards clinical use of trajectory classes to ultimately improve patient outcomes and service efficiency.

4.6 References

Beard, J. I. L., & Delgadillo, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety*, *36*(9), 866–878.

Beck, A., Naz, S., Brooks, M., & Jankowska, M. (2019). *Improving Access to Psychological Therapies (IAPT) Black, Asian and minority ethnic service user positive practice guide*. BABCP. https://babcp.com/Portals/0/Files/About/BAME/IAPT-BAME-PPG-2019.pdf?ver=2020-06-16-004459-320

Bone, C., Delgadillo, J., & Barkham, M. (2021). A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology*, *68*(2), 219–231.

Community and Mental Health Team. (2014). *Announcement of methodological change: Improving Access to Psychological Therapies (IAPT) Reports*. Health and Social Care Information Centre.

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., … Furukawa, T. A. (2021). Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry: Official Journal of the World Psychiatric Association* , *20*(2), 283–293.

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

Delgadillo, J., Dawson, A., Gilbody, S., & Böhnke, J. R. (2017). Impact of long-term medical conditions on the outcomes of psychological therapy for depression and anxiety. *The British Journal of Psychiatry: The Journal of Mental Science*, *210*(1), 47–53.

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, *85*(9), 835–853.

Delgadillo, J., McMillan, D., Gilbody, S., de Jong, K., Lucock, M., Lutz, W., … Ali, S. (2021). Cost-effectiveness of feedback-informed psychological treatment: Evidence from the IAPT-FIT trial. *Behaviour Research and Therapy*, *142*, 103873.

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.

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy*, *73*, 1–7.

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, *41*(2), 159–164.

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: causal mediation of change. *Journal of Consulting and Clinical Psychology*, *61*(4), 678–685.

Jung, T., & Wickrama, K. A. S. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, *2*, 302–317.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., … Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*(1), 22–36.

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.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Lowe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–326.

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *17*(1), 1–14.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., … Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, *82*(2), 287–297.

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, *118*(1-3), 60–68.

Marks, I. M. (1986). *Behavioural Psychotherapy: Maudsley Pocket Book of Clinical Management*. Bristol: Wright/Institute of Psychiatry Publishing.

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.

Muthén, B., & Muthén, L. K. (2000). Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling With Latent Trajectory Classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882–891.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

National Collaborating Centre for Mental Health; NCCMH. (2021). *The Improving Access to Psychological Therapies Manual* (Version 5). National Collaborating Centre for Mental Health.

National Institute for Health and Care Excellence; NICE. (2011). *Common Mental Health Problems: Identification and Pathways to Care.* National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/CG123/chapter/1-Guidance#steps-2-and-3-treatment-and-referral-for-treatment.

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *30*(1), 79–96.

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.

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.

Saunders, R., Buckman, J. E. J., & Pilling, S. (2020). Latent variable mixture modelling and individual treatment prediction. *Behaviour Research and Therapy*, *124*, 103505.

Saunders, R., Cape, J., Leibowitz, J., Aguirre, E., Jena, R., Cirkovic, M., … Buckman, J. E. J. (2020). Improvement in IAPT outcomes over time: are they driven by changes in clinical practice? *Cognitive Behaviour Therapy*, *13*, e16.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Stewart, R., Soremekun, M., Perera, G., Broadbent, M., Callard, F., Denis, M., … Lovestone, S. (2009). The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry*, *9*, 51.

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, *75*(6), 864–874.

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy*, *50*(6), 374–380.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(3), 451–467.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*, *163*(1), 148–150.

Chapter 5: Trajectories of depression symptoms, anxiety symptoms and functional impairment during internet-enabled cognitive-behavioural therapy

A version of this chapter is available as a preprint on the PsyArXiv server and has received an invitation to revise and resubmit from *Behaviour Research and Therapy*. Supplementary materials for this chapter, as detailed in the text, are included in Appendix D.

Preprint citation: **Skelton, M.**, Catarino, A., Brown, S., Carr, E., Davies, M. R., Peel, A., … & Eley, T. C. (2022, February 20). Trajectories of depression symptoms, anxiety symptoms and functional impairment during internet-enabled cognitive-behavioural therapy. https://doi.org/10.31234/osf.io/qxz2k

Authors: **Megan Skelton** [1,2], Ana Catarino [3], Stephanie Brown [3], Ewan Carr [1], Molly R Davies [1], Alicia J Peel [1], Christopher Rayner [1], Gerome Breen [1,2], Thalia C Eley [1,2]

1. Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
2. National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK
3. Digital Futures Lab, ieso, Cowley Road, The Jeffreys Building, Milton, Cambridge, UK

5.1 Abstract

Differences between patient symptom trajectories during psychological therapy can be captured by underlying classes. This has not been explored in therapy delivered in real-time via the internet or for trajectories of functional impairment.

Patients experiencing common mental health problems received real-time cognitive-behavioural therapy sessions with a therapist using an online chat platform (N = 52,029). Trajectory classes of depression symptoms (PHQ9), anxiety symptoms (GAD7) and functional impairment (WSAS) were investigated using growth mixture modelling. Multinomial regressions tested for associations between baseline variables and trajectory class.

A four-class trajectory model was selected for each outcome and these optimal models were highly similar in form. Each showed three classes with initially moderate-severe symptoms or impairment: one demonstrated no change, one gradual improvement and one fast improvement. A fourth class had mild baseline scores and minimal improvement. In the moderate-severe baseline classes, patients in the two with improvement were more likely to be employed and not to have obsessive-compulsive disorder. Fast improvement was likelier than gradual or no change for older patients, patients without a disability, or with lower comorbid symptom or impairment scores. Associations with the functional impairment model were more similar to those with the depression model than the anxiety model. Results were largely consistent with findings from trajectory studies of in-person psychological therapy. This is an important step towards personalising therapy in terms of length, continuation and progress monitoring.

5.2 Background

Psychological and pharmacological treatment options are available for common mental health problems such as depression and anxiety. Both types of treatment can be effective, however, there is a general patient preference for psychological treatments (McHugh et al., 2013). Even so, not everyone experiencing a mental health problem seeks or receives treatment (McManus et al., 2016; Rayner et al., 2020). There are many potential reasons for this 'treatment gap', for example, a lack of availability of psychological therapy in one's accessible locale or perceived stigma of seeking help or receiving a diagnosis (Gulliver et al., 2010; Magaard et al., 2017).

One avenue to increase access to psychological therapy is to use computerised and internet-based approaches. Numerous programs and treatment delivery methods exist, with most using standardised text modules following principles of cognitive-behavioural therapy (CBT) via websites (Burger et al., 2020). These were described in more detail in the introduction (section 1.9). Evidence suggests that guided internet-delivered CBT with therapeutic support has similar efficacy to in-person CBT (Andersson et al., 2019; Andrews et al., 2018; D. Kessler et al., 2009). Arguably the closest approximation of traditional in-person CBT is guided, real-time therapy sessions involving synchronous therapist interaction via the internet, for example, using instant messaging or a video call. We will refer to this as 'internet-enabled CBT' to distinguish it from other forms, which are less likely to be suitable for more severe presentations of depression and anxiety (Karyotaki et al., 2021). In addition to helping decrease waiting lists, internet-enabled CBT has several advantages over in-person therapy that may help to reduce the treatment gap. Firstly, treatment can be attended wherever there is a suitable device and internet availability. This increases access for individuals who would have difficulty travelling to a site due to geographical location, physical disability, or mental health disorder such as agoraphobia. Further advantages are increased accessibility for individuals who benefit from sessions outside of normal office hours, and those who perceive stigma attached to visiting a therapist (Gega et al., 2004; Webb et al., 2017). During the COVID-19 pandemic, remote therapies became indispensable for treating new and existing patients.

Studies of in-person psychological treatment have identified substantial heterogeneity in outcomes (Kaiser et al., 2022) which can be partially explained by patient and treatment characteristics (Delgadillo et al., 2016; R. C. Kessler et al., 2017). These studies have largely focused on endpoint outcomes, such as symptom-based response and remission. More recently, researchers have extended this exploration to internet-enabled CBT and identified several correlates and predictors of positive endpoint outcomes including older age, absence of physical comorbidities, and genetic factors (Catarino et al., 2018; Karyotaki et al., 2018; Wallert et al., 2022). Differences between patients *during* a course of therapy have been comparatively overlooked, despite potential important implications. For example, patients with outcome trajectories showing improvement later in treatment could be at risk of inappropriate early discontinuation or alteration of therapy. Information about the likely trajectory of a patient's outcomes could inform expectations and decisions regarding the number of required sessions as well as the suitability and continuation of therapy. It could also be used to monitor whether a patient is deviating from their expected trajectory. Alerting clinicians to 'off-track' patients can improve patient outcomes (de Jong et al., 2021).

Symptom trajectories have been explored in two studies of patients receiving in-person psychological therapy for common mental health problems at National Health Service (NHS) England Improving Access to Psychological Therapies (IAPT) services (Saunders et al., 2019; Skelton et al., 2021 - **Chapter 4 in this thesis**). They analysed over 4,000 and 16,000 patients, respectively, and identified multiple latent classes (i.e., unobserved clusters) of patients who exhibited similar trajectories. Overall, similar trajectory classes emerged from each study and outcome, as described in Chapter 4. Baseline variables associated with being in a moderate-severe symptoms with improvement trajectory class, rather than the moderate-severe no change class, were: being employed (Chapter 4), not being prescribed psychotropic medication or reporting a disability (Chapter 4), and lower baseline scores on phobia items (Saunders et al., 2019). Additionally, patients showing fast improvement tended to report lower functional impairment (Chapter 4 and Saunders et al., 2019). Very few trajectory class analyses have been performed with data from internet-delivered therapy. We are aware of two studies that analysed samples of between 400 and 600 patients who had received either unguided or asynchronously guided internet-delivered

CBT (Lutz et al., 2017; Sunderland et al., 2012). These forms of internet-delivered treatment have minimal therapist involvement and little structure as patients can complete modules on their own schedule. Furthermore, few sessions were analysed, with measures from only two therapy sessions modelled in one, and five in the other. The importance of modelling later treatment sessions was highlighted by the identification of a gradual improvement trajectory class in Chapter 4.

In the present study, we aimed to determine whether findings from trajectory class studies of in-person psychological therapy extended to internet-enabled CBT. We investigated symptom change patterns and associations between baseline variables and trajectory class membership over 11 time points; an initial assessment and 10 treatment sessions. Existing evidence of similar efficacy between these delivery methods suggested that results would be similar. We additionally explored trajectories of functional impairment, an important yet often neglected outcome (McKnight & Kashdan, 2009; Zimmerman et al., 2008), which might not change synchronously with symptoms (Howard et al., 1993; McKnight & Kashdan, 2009).

5.3 Methods

5.3.1 Sample

The study sample consisted of data from NHS IAPT patients who had received internet-enabled CBT for a common mental health problem between January 2018 and June 2021. CBT was provided by a commercial platform from ieso (*Ieso*, 2022) that was developed for use in the NHS. At registration, patients agreed to the use of their anonymised data to support research purposes. The initial dataset contained records from 97,686 patients living in England and Scotland, who had self-referred or been referred by their general practitioner (further details in Supplementary Information 1). To produce a sample suitable for analysis, patients were excluded if they had not attended any therapy sessions (i.e., assessment only) and sessions were excluded if they occurred after a long interval or after 11 sessions including the assessment (Supplementary Information 2). Following this, 52,029 patients remained covering 310,808 sessions. To assess the generalisability of the results, we compared the included to the excluded patients using appropriate group

difference tests and focusing on effect sizes due to potential statistical significance of negligible differences. Patients who scored below clinical cut-offs on the core symptom scales were included in the analysis.

Patients were assigned to a CBT therapist accredited by the British Association for Behavioural and Cognitive Psychotherapies and attended hour-long weekly sessions via a confidential online written instant-messaging platform. In the initial assessment, anxiety symptoms, depression symptoms and functional impairment were recorded using questionnaires detailed in the measures section. The therapist then provided a 'problem descriptor', as described in the introduction (section 1.8). Here, we will refer to it as 'diagnosis'. Treatment course duration was determined by the therapist. Between appointments, patients and therapists could message each other using the platform, for example to arrange appointments or provide additional information on a topic discussed in therapy. Consistent with a CBT framework, patients also completed homework tasks.

5.3.2 Measures

Outcomes for trajectory modelling were self-report questionnaires routinely collected at each time point (session). Current depression symptoms were measured with the Patient Health Questionnaire 9-item version (PHQ9; Kroenke et al., 2001; total scores range 0-27; IAPT case threshold ≥10). Current anxiety symptoms were assessed using the Generalized Anxiety Disorder 7-item scale (GAD7; Spitzer et al., 2006; total scores 0-21; IAPT case threshold ≥8). The Work and Social Adjustment Scale (Marks, 1986; Mundt et al., 2002; WSAS) consists of five items, each scored from 0 to 8 to indicate functional impairment in a different life domain; relationships, work, home management, and private and social leisure activities. Patients not in employment (e.g., student, unemployed) did not answer the work item. Imputation for a total score could be unsuitable if observations are 'missing not at random' (Sterne et al., 2009). We therefore calculated complete case five-item and four-item total scores and ran analyses with each. Patients with all items complete for at least one session, and either all items complete or all items missing for other sessions, were included in the five-item model. If a patient was missing the work item at any session, they

were allocated to the four-item model. Unless stated, a reference to WSAS means the five-item scale; results from the reduced, unvalidated four-item score are in the supplementary.

To investigate relationships between patient characteristics and trajectory classes, we selected a number of self-reported baseline variables informed by the analysis in Chapter 4, as well as a previous study of trajectories during in-person therapy (Saunders et al., 2019). Total scores from the PHQ9 and GAD7, and individual scores from four WSAS items (i.e., excluding work) were included when not the modelled outcome. We also investigated single item measures of agoraphobia, social phobia and specific phobia, each scored from 0 to 8 to indicate the extent of avoidance of a situation or object. Other included variables were age (in 10-year blocks), gender (female, male), ethnicity (White, minoritised ethnic groups), employment (employed, unemployed, non-worker (e.g., retired, student)), disability (not reported, reported), psychotropic medication (not prescribed, prescribed), and diagnosis. Diagnosis included depression, generalised anxiety disorder (GAD), other anxiety (agoraphobia, panic disorder, social phobia, specific phobia, hypochondriacal disorder, unspecified anxiety), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), and 'other' (infrequent diagnoses e.g., adjustment disorder). These categories slightly differ to those used in Chapter 4 due to differences in observed frequencies. To ensure sufficient counts in each factor level for statistical power, variables underwent pre-processing in which levels were aggregated. We recognise that this means some variables, such as ethnicity, did not adequately reflect patient identity.

5.3.3 Statistical analysis

We used a person-centred longitudinal analysis called growth mixture modelling (GMM; Muthén & Muthén, 2000) to identify the presence of subgroups of patients with similar patterns of symptom change over therapy sessions. The analytical steps for the GMM, handling of missing data and associations between baseline variables and trajectory class membership were the same as described in Chapter 4. See Supplementary Figure 1 for outcome distributions and Supplementary Table 1 for the proportion of patients with outcomes in treatment per session. Statistical significance in the regression models was

assessed with a Bonferroni-adjusted threshold of $p < 0.016$ to account for three independent models.

5.4 Results

5.4.1 Sample descriptives

Sample descriptives for the 52,029 patients are presented in Table 1. Supplementary Table 2 presents the descriptives of the patients who were included in the WSAS model (N = 32,168). They were very similar to the overall sample besides a higher proportion reporting being employed, as expected (see measures section). The average number of time points (i.e., sessions) recorded, after limiting to maximum 11 including the initial assessment, was 6.0 (SD = 3.1); 5% had only 1 and 14% had 11 (Supplementary Information 2). The average interval between sessions was 11 days (SD = 6.0; Supplementary Information 2). Patients in the analysis were similar to excluded patients (N = 45,657) besides slightly milder symptoms (Cohen's *d* 0.15 - 0.29; Supplementary Table 3). The number of patients with at least one total score, for each scale, was PHQ9: 51,683 (99%), GAD7: 51,667 (99%), WSAS: 32,168 (62%). Pearson's correlations between the pairs of measures were all significant ($p < 0.05$: PHQ9-GAD7 = 0.64, PHQ9-WSAS = 0.61, GAD7-WSAS = 0.43).

Table 1. Sample descriptives of patients who received internet-enabled CBT (N = 52,029)

Variable		**Mean (SD); range or Count (proportion %)**
Age (years)	Mean (SD); Range	34.3 (12.3); 18 - 94
	Missing	*539 (1%)*
Gender ref: female	Male	13,572 (26%)
	Missing	*211 (0.4%)*
Depression symptoms (PHQ9)	Mean (SD); Range	12.6 (6.1); 0 - 27
	Missing	*537 (1%)*
Anxiety symptoms (GAD7)	Mean (SD); Range	12.2 (5.3); 0 - 21
	Missing	*556 (1%)*
Case on PHQ9 and/or GAD7[1]	Yes	42,984 (83%)
	Missing	*518 (1%)*
Functional impairment (WSAS)	Mean (SD); Range	16.1 (8.3); 0 - 40
	Missing	*12,545 (24%)*

Functional impairment (WSAS) Home management	Mean (SD); Range	2.9 (2.1); 0 - 8
	Missing	*800 (2%)*
Functional impairment (WSAS) Social leisure	Mean (SD); Range	3.8 (2.3); 0 - 8
	Missing	*800 (2%)*
Functional impairment (WSAS) Private leisure	Mean (SD); Range	2.8 (2.2); 0 - 8
	Missing	*800 (2%)*
Functional impairment (WSAS) Relationships	Mean (SD); Range	3.2 (2.2); 0 - 8
	Missing	*800 (2%)*
Functional impairment (WSAS) Work	Mean (SD); Range	3.3 (2.3); 0 - 8
	Missing	*12,530 (24%)*
Agoraphobia item	Mean (SD); Range	2.6 (2.5); 0 - 8
	Missing	*814 (2%)*
Social phobia item	Mean (SD); Range	3.3 (2.4); 0 - 8
	Missing	*814 (2%)*
Specific phobia item	Mean (SD); Range	2.5 (2.5); 0 - 8
	Missing	*814 (2%)*
Diagnosis [2]	Depression (ref)	21,296 (41%)
	GAD	14,203 (27%)
	Other anxiety	10,184 (20%)
	OCD	2,509 (5%)
	PTSD	2,107 (4%)
	Other	1,677 (3%)
	Missing	*53 (0.1%)*
Prescribed psychotropic medication ref: No	Yes	23,996 (46%)
	Missing	*1,564 (3%)*
Ethnicity ref: White	Minoritised ethnic groups	3,731 (7%)
	Missing	*13,459 (26%)*
Disability reported ref: No	Yes	4,731 (9%)
	Missing	*26,107 (50%)*
Employment status [3]	Employed (ref)	34,518 (66%)
	Non-worker	8,543 (16%)
	Unemployed	6,061 (12%)
	Missing	*2,907 (6%)*
Recovered (higher 'yes' if include more than the 10 treatment sessions modelled)	Yes	17,758 (34%)
	Missing (includes patients who were not a case at baseline)	*11,190 (22%)*

Baseline values are presented for variables measured at multiple timepoints. [1] Case thresholds: PHQ9 ≥10, GAD7 ≥8. [2] IAPT 'problem descriptor': GAD = generalised anxiety disorder; Other anxiety = e.g., panic disorder, social phobia; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; Other = e.g., adjustment disorders; further details in Measures. [3] 'Non-worker' = e.g., retired. For categorical variables included in the regressions, ref = reference category (proportion not shown if binary).

5.4.2 Growth mixture models

The optimum fitting latent growth curve shape for all three outcomes was quadratic (Supplementary Information 3). As was observed in Chapter 4, on average, patients had initially moderate depression and anxiety symptoms which decreased throughout treatment, with more marked improvement over the first five treatment sessions. The functional impairment latent growth curve showed average baseline scores indicating moderate functional disability (Mataix-Cols et al., 2005; Mundt et al., 2002) and small, consistent improvement to the final session. The initial modelling with variance constrained to zero resulted in rather uninformative classes (Supplementary Information 4) and therefore the focus of the results is the models with free but equal variance in the intercepts.

For each outcome, a four-class GMM was chosen as the optimal model (Table 2 and Figure 1, also see Supplementary Information 5-7 for further details of the model selection process including trajectory plots of all tested models). Average class trajectories went outside of the range of possible scores on the outcome measures in the five- and six-class symptom models, and the six-class model of functional impairment.

Table 2. Fit indices for growth mixture model of depression symptoms (PHQ9; top table; n = 51,683) anxiety symptoms (GAD7; middle table; n = 51,667) and functional impairment (WSAS; bottom table; n = 32,168)

Depression symptoms GMM	**Params**	**AIC**	**BIC**	**Entropy**	**VLMR LRT *p*-value**	**Class Proportions (%)**
Growth Curve	25	1657334	1657556	NA	NA	100
Two Class	29	1649626	1649883	0.514	< 0.001	75.4, 24.6
Three Class	33	1643957	1644249	0.602	< 0.001	58.8, 21.0, 20.1
Four Class	37	1640457	1640784	0.629	< 0.001	14.7, 14.8, 57.5, 13.1
Five Class	41	1638982	1639345	0.663	< 0.001	13.3, 56.8, 14.3, 13.5, 2.0
Six Class	45	1638057	1638455	0.654	< 0.001	14.2, 1.9, 14.1, 54.8, 10.9, 4.1

Anxiety symptoms GMM	Params	AIC	BIC	Entropy	VLMR LRT *p*-value	Class Proportions (%)
Growth Curve	25	1626515	1626737	NA	NA	100
Two Class	29	1617917	1618174	0.444	< 0.001	66.5, 33.5
Three Class	33	1611748	1612040	0.602	< 0.001	44.8, 29.0, 26.2
Four Class	37	1607426	1607753	0.609	< 0.001	19.3, 44.7, 18.9, 17.1
Five Class	41	1605757	1606120	0.652	< 0.001	17.3, 20.1, 44.4, 16.3, 2.0
Six Class	45	1604713	1605111	0.669	< 0.001	2.0, 19.7, 16.8, 42.9, 15.6, 3.0

Functional impairment GMM	Params	AIC	BIC	Entropy	VLMR LRT *p*-value	Class Proportions (%)
Growth Curve	25	1088331	1088541	NA	NA	100
Two Class	29	1082202	1082445	0.565	< 0.001	26.7, 73.3
Three Class	33	1078518	1078795	0.577	< 0.001	23.9, 62.2, 13.9
Four Class	37	1076928	1077238	0.619	0.0018	20.9, 60.5, 10.3, 8.4
Five Class	41	1076092	1076436	0.625	0.0038	17.8, 6.4, 7.0, 58.7, 10.1
Six Class	45	1075459	1075836	0.639	0.0011	57.7, 17.5, 8.1, 2.2, 4.6, 9.9

Params = parameters, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, VLMR LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. Proportions are based on a patient's most likely class. Lower values on AIC and BIC indicate better fit. Higher entropy represents better distinction. A significant VLMR LRT p-value suggests the current model is a better fit of the data than a model with one fewer class.

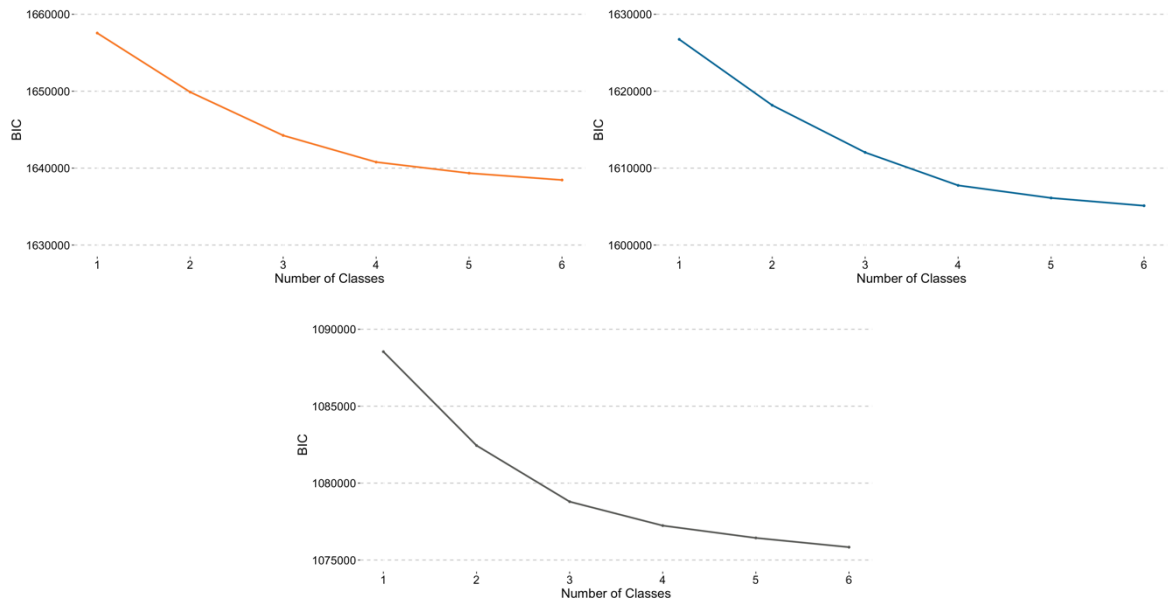


Figure 1. Elbow plot of Bayesian Information Criterion (BIC) values for growth mixture models of depression symptoms (PHQ9; upper left), anxiety symptoms (GAD7; upper right) and functional impairment (WSAS; bottom)

Figure 2 shows the model-estimated mean trajectory for each class from the selected four-class models, as well as observed mean patient trajectories per class. As observed in Chapter 4, the estimated trajectories were strikingly similar across the outcomes, and therefore the same labels were used. These labels were appropriate for the severity score thresholds of each measure (i.e., mild, moderate-severe; Kroenke et al., 2001; Mataix-Cols et al., 2005; Mundt et al., 2002; Spitzer et al., 2006). Three classes had initially moderate-severe scores, one of which showed no notable change (moderate-severe plateau; grey squares), the second improved steadily throughout (moderate-severe, gradual improvement; orange diamonds), and the third showed early improvement over the first five treatment sessions then little change (moderate-severe, fast improvement; blue triangles). In the symptom models, this latter trajectory showed slight deterioration in the last two sessions. A fourth class had mild baseline scores, with estimated mean intercept values below case thresholds in the symptom models (PHQ9 = 8.9, GAD7 = 7.8) and minimal change over treatment sessions (mild, small improvement; green circles). For brevity, these will be referred to as 'plateau', 'gradual', 'fast' and 'mild'. The estimated class trajectories closely represented the average observed trajectories although less so for the latter half of treatment, especially for the fast and mild classes.

5.4.2.1 Depression symptoms growth mixture model

Over half of the sample were most likely to belong to the mild class of the depression model (62%), and 11-13% to each of the other classes. The mean number of sessions received, including the assessment, ranged from 5.7 (SD = 2.5) for fast to 6.9 (SD = 3.3) for plateau. By the final session, the majority of patients in the fast and mild classes who had scored as a case on at least one of the PHQ9 and GAD7 at baseline now scored below on both ('recovery' 70% and 53% respectively; this does not reflect scores after the tenth treatment session in patients who received more sessions). Most patients in the gradual and plateau classes had not recovered (32% and 1.5% recovered). Full descriptives of patients in each class, for each outcome, are available in Supplementary Information 5-7.

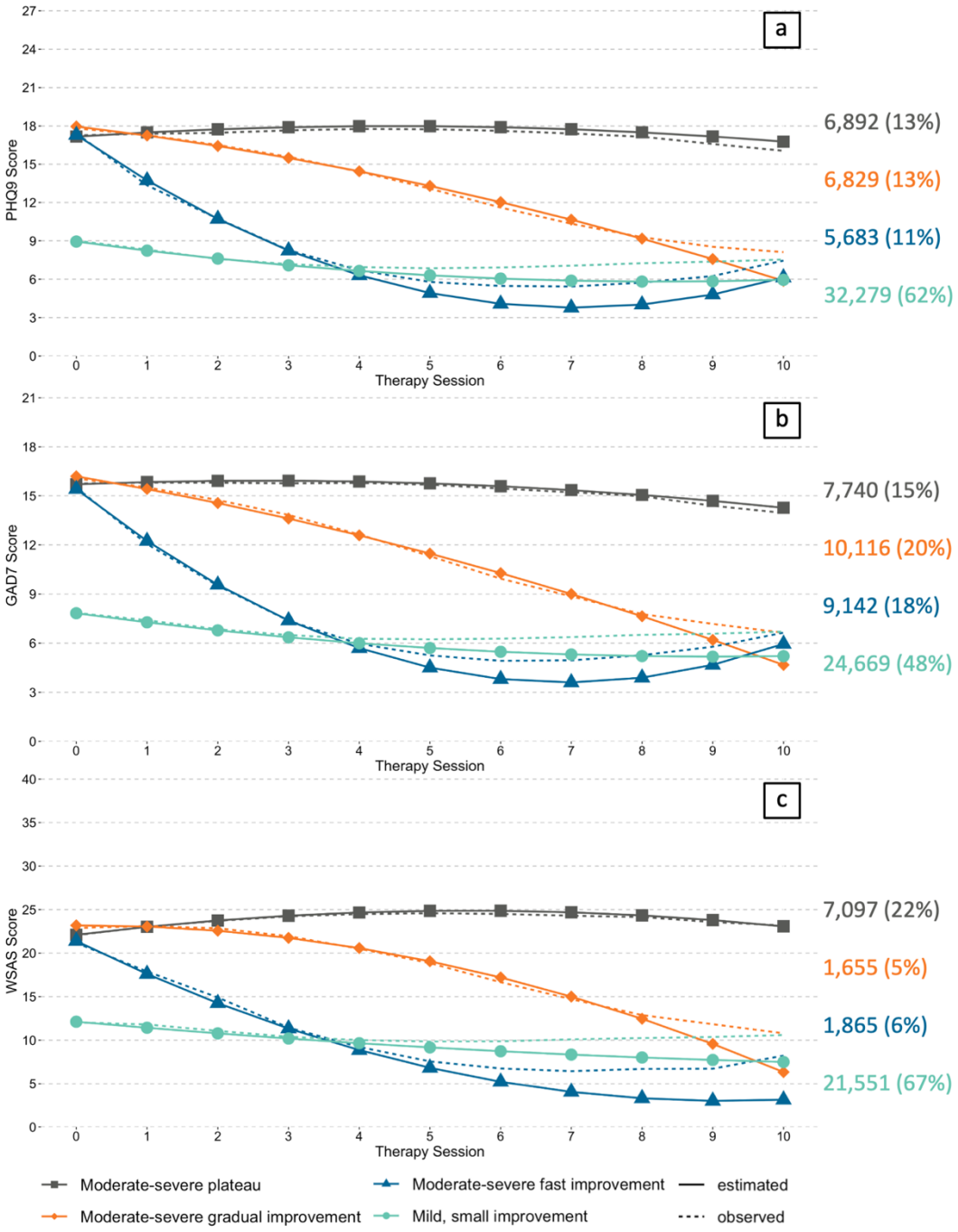


Figure 2. Four-class growth mixture models of patients during internet-enabled cognitive-behavioural therapy based on: a) depression symptoms (PHQ9; n = 51,683) b) anxiety symptoms (GAD7; n = 51,667) c) functional impairment (WSAS; n = 32,168)

Model-estimated and observed mean trajectories of each class. Therapy session 0 represents baseline pre-treatment assessment. Counts and proportions (%) are based on patients' 'most likely' trajectory membership. Total N = 52,029; patients with at least one recorded score for an outcome measure were included in that model.

5.4.2.2 Anxiety symptoms growth mixture model

Compared with the depression model, fewer patients in the anxiety model were most likely to belong to the mild class (48%), and more to the gradual or fast class (20% and 18%). The mean number of sessions received ranged from 5.6 for fast and mild (SD = 2.6 and 3.1, respectively) to 7.4 (SD = 3.1) for plateau. Similar proportions to the depression model recovered in each class, with the highest in the fast class (72%), followed by the mild (54%), gradual (34%) and plateau (1.5%) classes.

5.4.2.3 Functional impairment growth mixture model

The fast class of the functional impairment model continued to improve for longer than in the symptom models, however, this deviated from the observed scores. Similar to the depression model, the majority of the sample were most likely to belong to the mild class (67%). Nevertheless, there was a substantial proportion of patients in the plateau class (22%), with few assigned to either the gradual or fast improvement class (5% and 6%). The mild class had the lowest average number of sessions (5.6; SD = 3.1) and the gradual class the highest (8.4; SD = 2.0). There is no standard WSAS recovery threshold, however, most patients in the gradual, fast and mild classes showed symptom-based recovery (63%, 78%, 53% respectively), whilst few did in the plateau class (10%). Overlap of trajectory classes between the three models is presented in Supplementary Figure 2. Over a third of patients (37%) were in mild classes for all three outcomes, and 6% were in plateau classes for all.

5.4.3 Associations of baseline variables with trajectory class

As was described in Chapter 4, regression results are reported with odds ratios (ORs) representing the difference in odds of a specific trajectory class in comparison to the moderate-severe plateau class (see Figure 3 or Supplementary Tables 4 - 6). As an example, in the depression model, for every one point higher on the anxiety symptoms measure, the OR of being in the moderate-severe, gradual improvement class compared with plateau was 1.04. A patient with a score of 21 (the maximum) had 2.28 the odds of being in the gradual class than a patient with a score of 0 (the minimum). This demonstrates the importance of the unit and scale of the measure for interpretation. For each model, several results with OR values are included to assist interpretation of Figure 3.

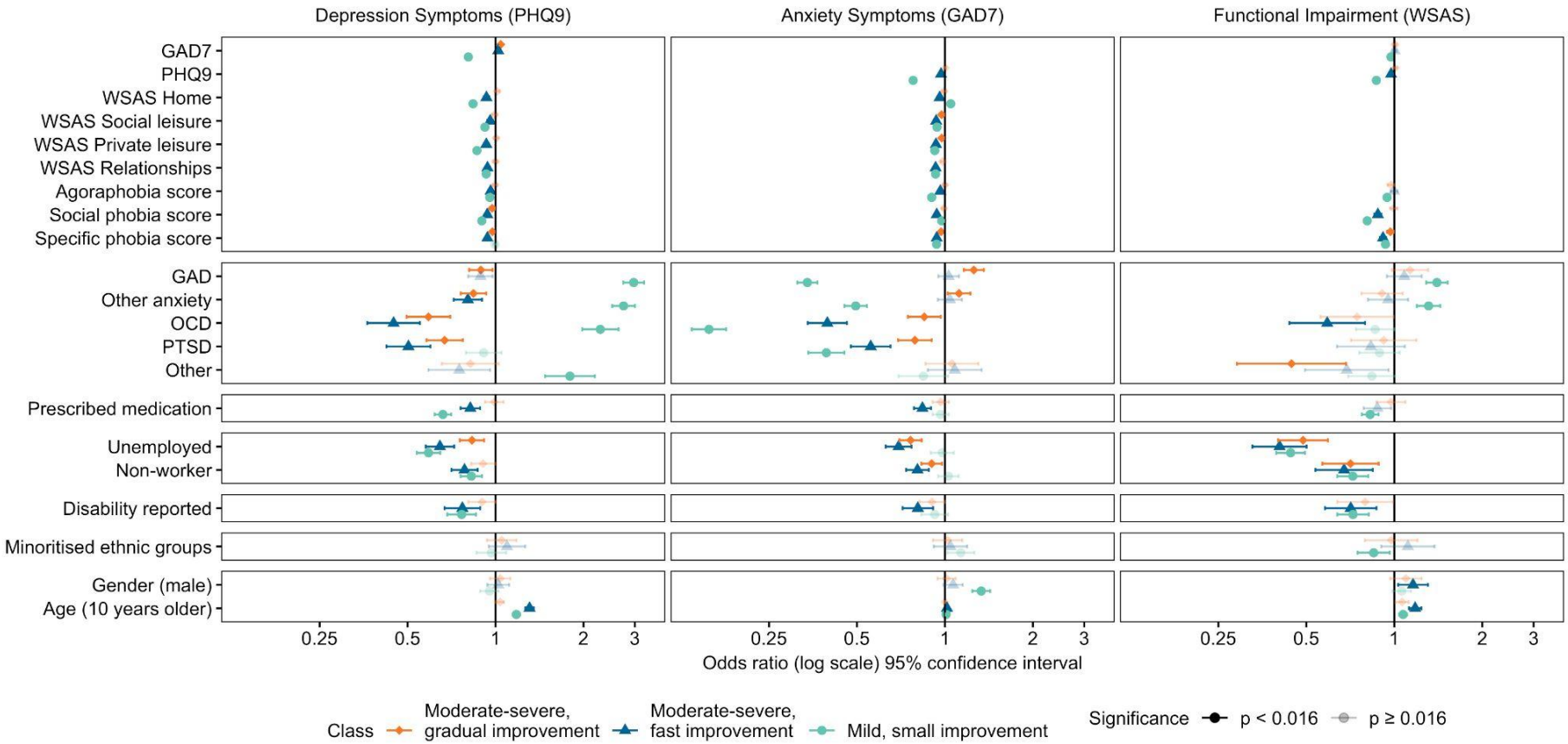


Figure 3. Odds ratios of patient trajectory class membership from multinomial regressions of baseline variables and four-class growth mixture models of depression symptoms (PHQ9; n = 51,683), anxiety symptoms (GAD7; n = 51,667) and functional impairment (WSAS; n = 32,168), during internet-enabled cognitive-behavioural therapy

Reference class = moderate-severe, plateau; points to the right of x = 1 indicate greater odds of the trajectory class indicated by colour. PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment. Non-worker = e.g., retired. GAD = generalised anxiety disorder; Other anxiety = e.g., panic disorder, social phobia; OCD = obsessive-compulsive disorder; PTSD = post-traumatic stress disorder; 'Other' = e.g., adjustment disorders. A Bonferroni-adjusted threshold of p < 0.016 was used to account for the three independent models.

5.4.3.1 Conditional growth mixture model of depression symptoms

Higher baseline anxiety symptoms were associated with lower odds of belonging to the mild depression model class (OR = 0.81) and negligibly higher odds of the gradual (OR = 1.04) or fast (OR = 1.02) class. For each of the four functional impairment items and three phobia items, a higher score was generally associated with lower odds of the fast or mild class. Social and specific phobia scores additionally differentiated the gradual class from plateau, but effects were small. Reporting prescribed psychotropic medication, a disability, being unemployed or a non-worker was associated with lower odds of the fast or mild class. Unemployment additionally showed lower odds of the gradual class. Compared with a depression diagnosis, patients with any other diagnosis besides PTSD were more likely to be in the mild class. Patients with 'Other anxiety', OCD or PTSD were less likely to belong to the gradual or fast class. No associations were found with the binary categories of ethnicity and gender. Older age was associated with higher odds of the fast or mild class.

5.4.3.2 Conditional growth mixture model of anxiety symptoms

A higher baseline score for depression symptoms, a functional impairment item or phobia item was associated with lower odds of the fast class than the reference class, plateau (ORs 0.93 - 0.97). Higher scores were also generally associated with lower odds of the mild class. Of these variables, only social and private leisure impairment and specific phobia differentiated the gradual class from plateau, albeit negligibly. Reporting prescribed medication or a disability was associated with lower odds of the fast class, as was being unemployed or a non-worker, which additionally showed lower odds of the gradual class. Unlike in the depression model, none of these factors were associated with the mild class. Compared with a depression diagnosis, all diagnoses besides Other were associated with lower odds of the mild class. OCD and PTSD were associated with lower odds of any more favourable class, especially mild. No significant association was found with ethnicity, or any notable association with age. Male gender, compared with female, was associated with higher odds of the mild class.

5.4.3.3 Conditional growth mixture model of functional impairment

Higher baseline symptom scores for anxiety, depression and phobia items were each associated with lower odds of the mild trajectory class, compared with the plateau class.

This was strongest for depression and social phobia (ORs = 0.87 and 0.81). Higher depression, social phobia and specific phobia symptoms were additionally associated with lower odds of the fast class, and specific phobia was associated with lower odds for any more favourable class. The pattern of associations with employment status and disability were the same as in the depression model, although point estimates for association with employment status were stronger in the present model. Consistent with the depression model, a diagnosis of GAD or 'Other anxiety' was associated with increased odds of the mild class, compared with a depression diagnosis and as observed in both symptom models, OCD was associated with lower odds of the fast class. There was some indication of lower odds of the mild class for patients from minoritised ethnic groups. As was observed for the symptom models, older age was associated with higher odds of the fast or mild class. Results for the GMM and regressions with the four-item WSAS, without the work item, were largely similar besides differences for diagnoses of OCD, PTSD and Other, and non-worker employment status (Supplementary Information 8).

5.5 Discussion

5.5.1 Overview

Depression symptoms, anxiety symptoms and functional impairment scores of 52,029 patients during internet-enabled CBT were each best modelled by four underlying classes of trajectories. These trajectory classes showed different associations with patient variables recorded at the start of treatment, such as employment status and comorbid symptoms. Strikingly, the trajectories identified in our analyses were highly similar to those found in analyses of patients who attended in-person psychological therapies services (Chapter 4, as well as Saunders et al., 2019). This study demonstrated for the first time that symptom trajectory classes from in-person therapy generalise to internet-enabled CBT, and additionally identified trajectory classes of functional impairment and their associations with baseline variables.

5.5.2 Growth mixture models

For each outcome, the heterogeneity in the observed data was better explained by four trajectory classes than a single average trajectory. All three selected models had three

trajectory classes which demonstrated initially moderate-severe symptoms or impairment followed by slopes showing: no change ('plateau'), gradual improvement, or fast improvement. A fourth class with lower average baseline values was called mild, small improvement. The similar initial scores of the moderate-severe classes indicate that, although a reliable predictor of endpoint outcome (Buckman et al. 2021), baseline severity is insufficient to indicate the trajectory that a patient may follow.

For each model, the majority of patients were most likely to belong to a mild trajectory, consistent with in-person therapy findings (Chapter 4 and Saunders et al., 2019). Around one-third were assigned to a mild trajectory class for all three outcomes. This may be attributable to some patients' primary severity being reflected in other symptoms such as phobias. Compared with symptom models, the functional impairment model had more patients in the mild, small improvement and plateau classes. This tentatively supports existing evidence that functional impairment takes longer to exhibit change than symptoms (Howard et al., 1993), highlighting the need to include impairment as a treatment outcome.

The plateau and gradual classes showed the lowest recovery rates. As discussed in Chapter 4, low recovery could be partially attributed to patients who required a higher number of sessions to improve than were available to model here (Robinson, Delgadillo, et al., 2020). This may be especially applicable for impairment if symptomatic improvement is required first (Howard et al., 1993). A trajectory study of in-person pharmacological and psychological depression therapies over a six-month period revealed late improvement in patients who had shown little symptom change at three months and suggested that short timeframes can overestimate nonresponse (Thibodeau et al., 2015).

5.5.3 Associations of baseline variables and class membership

Overall, the associations between baseline variables and symptom class membership were similar to those reported in two existing trajectory class studies of patients who received in-person therapy (Chapter 4 and Saunders et al., 2019). In the present study, there were several consistent patterns of association across the three outcomes. As also observed in Chapter 4, there was a fairly stable pattern of lower odds of any of the more favourable

classes for patients who were unemployed, and similar yet attenuated associations for non-workers. A higher score on a symptom scale, functional impairment or phobia item was generally associated with lower odds of fast improvement or mild classes, although effects were small. The social phobia item showed especially strong associations with the functional impairment fast and mild classes. This suggests that impairment is a particularly relevant outcome for patients with higher social phobia symptoms. Although some impairment and phobia items showed lower odds of the gradual improvement class, relationships with fast improvement were stronger and more consistent. These variables could potentially differentiate expected speed of change, as was observed for the total WSAS score in Chapter 4.

Compared with depression, all other diagnoses showed higher odds of a mild depression model trajectory class (except PTSD) and lower odds of a mild anxiety model trajectory class (besides 'Other'). This pattern was also observed in Chapter 4. Interestingly, associations between diagnosis and functional impairment trajectory class were more similar to associations with the depression model, with GAD and 'Other anxiety' more likely to show a mild trajectory. This did not appear to be driven by differences in diagnostic frequencies between the impairment model subset and overall sample (Supplementary Table 4). Alongside this, the relationship between higher symptom scores and lower odds of the mild class was stronger in the functional impairment model for depression symptoms than anxiety symptoms. This was also consistent with a stronger correlation between the baseline WSAS and PHQ9 scores than with the GAD7 score. These findings could be interpreted in the context of severe impairment being observed more frequently in depression than anxiety disorders (Rapaport et al., 2005). It might also be attributable, however, to greater conceptual overlap between the PHQ9 and WSAS items. Another notable association with diagnosis was that PTSD and OCD showed lower odds of depression or anxiety model gradual or fast classes. The association of each of these two diagnoses and the fast class was also observed in Chapter 4, suggesting that this is an especially robust and generalisable finding. In the present study, OCD was also associated with lower odds of the functional impairment model fast class, indicating that this diagnosis may be of particular clinical relevance. There is evidence that a greater number of therapy sessions may be required to observe symptom improvement in OCD than most depression

and anxiety disorders (Robinson, Kellett, et al., 2020). This suggests that OCD is not necessarily associated with nonresponse but requires a greater dose of therapy than was modelled in our analysis.

Positive associations between the *absence* of a disability and older age with fast and mild trajectories reflected findings from investigations of endpoint outcomes in internet-enabled CBT (Catarino et al., 2018; Karyotaki et al., 2018). The association with age may be specific to internet-delivered treatments, as it was not observed in either in-person therapy study. The trajectory-based approach additionally revealed that, although fast improvement was less likely than the plateau class for patients who were younger or reporting a disability, they had comparable odds of gradual improvement. Therefore, although these patients have higher odds of no change than fast improvement, that does not mean that they are not likely to improve, but it is likely to be gradual if they do.

Some previously reported associations were not found in this analysis, such as lower odds of the anxiety model mild trajectory class for unemployment and prescribed medication in Chapter 4. This might be due to the baseline variables included, and therefore adjusted for, which had not yet been simultaneously tested in an analysis of in-person therapy. Rather than replication, the aim was to explore a wide range of factors that could feasibly be clinically implemented.

5.5.4 Strengths and limitations

This study revealed novel insights into the potential trajectories of symptoms and functional impairment for patients receiving internet-enabled CBT for common mental health disorders. Increasing reliance on remote technologies to provide therapy makes these findings particularly timely. Furthermore, the large sample size provided statistical power to accurately estimate multiple significant associations between baseline variables and trajectory classes. The in-person data used in Chapter 4 was less standardised than the data analysed in the present study, for example, having greater variation in treatment type and session regularity. Despite this, results were highly similar, both in terms of trajectory shapes and associations with baseline variables. This is consistent with reports of similar

endpoint outcomes between these methods of treatment delivery. Our results additionally highlighted the importance of monitoring impairment alongside symptoms.

There were a number of limitations to this study. One issue was estimated trajectories in the five- and six-class models falling outside the possible range of the measure. This is more likely to occur in complex models with many time points and unconstrained variance and can indicate inappropriate models (van der Nest et al., 2020). Furthermore, the estimated mean trajectories slightly diverged from the observed mean trajectories, especially for later sessions and for the functional impairment model. This could be due to treatments ending (recovery, referral or dropout) earlier than the ten treatment sessions modelled, such that later time points relied on data from fewer remaining patients. Unfortunately, there was no reliable data on the end of treatment reason. Replication in a trial with a prespecified number of sessions across patients could help clarify how missing data influenced the models but this design would remain subject to dropout. However, four-class symptom models were converged that were strikingly similar to models from independent samples, supporting the utility and validity of the trajectories. Another issue with missing data was the frequent 'not applicable' response to the WSAS work item. Models estimated in a patient subset without valid responses were highly similar to those from the full measure in terms of trajectory shape and class membership allocation but there were some differences in the identified baseline associations. This might suggest that separate models would be required depending on an individual's employment status. Ideally, patients would be encouraged to respond to the item where possible, for example, students could report impairment in their ability to study.

Baseline variables were selected based on availability and previously reported in-person findings, however, there was no available data on factors such as disorder chronicity and comorbidity which are associated with endpoint outcomes (R. C. Kessler et al., 2017) and might also differentiate trajectory classes. Future explorations of trajectory associations could additionally be informed by data-driven approaches, such as data mining of therapy transcripts, which internet-enabled CBT readily allows for (Ewbank et al., 2020). Another limitation was that a binary ethnicity variable was used due to low numbers of patients identifying as being from specific minoritised ethnic groups. In Chapter 4, a more ethnically

diverse sample showed different relationships with symptom trajectories for patients of Asian, Black, multi-ethnic and 'other' ethnicity, but functional impairment was not modelled. Further investigation is required with more representative ethnic groups. Future studies of the trajectories of symptoms and functional impairment should also ideally include a higher number of sessions than was available to us here. Late improvement and greater recovery may then be observed in some patients who showed plateau or gradual trajectories. Additionally, given the symptomatic heterogeneity of depression and anxiety, and different associations between symptoms and impairment domains (Fried & Nesse, 2014), individual questionnaire items and class membership could be explored. The different associations seen with impairment in home management in the present study indicates the relevance of this. Finally, modelling trajectories of core symptoms for disorders such as phobias and PTSD would be useful; although the GAD7 has reasonably high sensitivity for these disorders, it does not fully capture them (Kroenke et al., 2007).

5.5.5 Implications and conclusions

This study showed that individual differences in trajectories of symptoms and functional impairment during internet-enabled CBT can be captured by four underlying classes. These trajectory classes were associated with baseline patient variables such as employment and comorbid symptoms. The plateau class might reflect patients who require support in other domains to benefit from treatment, for example employment or unmeasured factors such as parenting or finance. This requires further exploration. Identifying patients likely to show fast improvement could assist service efficacy, whilst recognising the need for some sessions after remission to guard against false positives and relapse (Robinson, Kellett, et al., 2020). Gradual improvement patients could be at risk of premature treatment termination, by themselves or their clinician, as the trajectory over the first few sessions is indistinguishable from a plateau. Information on the possibility of later improvement might help prevent therapy from inappropriately ending, which could be especially important for observing improvement in functional impairment. While this study provides a step towards personalised treatment approaches, further research is required to produce validated predictive models that can combine information across variables to estimate a patient's most likely trajectory during an initial assessment.

Internet-enabled psychological therapy shows promise as an alternative to in-person treatment, with similar efficacy and even potential advantages that could play an integral role in improving treatment access (Andrews et al., 2018). Our findings were highly similar to studies of in-person therapy, suggesting that conclusions can be generalised across these modalities. Using variables available at the start of treatment to predict the timing of improvement could inform treatment plans, ensuring that therapy is not concluded or altered prematurely, and thereby increase the chances of the best possible outcomes for patients.

5.6 References

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry: Official Journal of the World Psychiatric Association*, *18*(1), 20–28. https://doi.org/10.1002/wps.20610

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders*, *55*, 70–78. https://doi.org/10.1016/j.janxdis.2018.01.001

Buckman, J. E. J., Saunders, R., Cohen, Z. D., Barnett, P., Clarke, K., Ambler, G., … Pilling, S. (2021). The contribution of depressive "disorder characteristics" to determinations of prognosis for adults with depression: an individual patient data meta-analysis. Psychological Medicine, 51(7), 1068–1081. https://doi.org/10.1017/S0033291721001367

Burger, F., Neerincx, M. A., & Brinkman, W.-P. (2020). Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. *Journal of Medical Internet Research*, *22*(1), e12599. https://doi.org/10.2196/12599

Catarino, A., Bateup, S., Tablan, V., Innes, K., Freer, S., Richards, A., … Blackwell, A. D. (2018). Demographic and clinical predictors of response to internet-enabled cognitive-behavioural therapy for depression and anxiety. *BJPsych Open*, *4*(5), 411–418. https://doi.org/10.1192/bjo.2018.57

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

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

Ewbank, M. P., Cummins, R., Tablan, V., Catarino, A., Buchholz, S., & Blackwell, A. D. (2020). Understanding the relationship between patient language and outcomes in internet-

enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 1–13. https://doi.org/10.1080/10503307.2020.1788740

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

Gega, L., Marks, I., & Mataix-Cols, D. (2004). Computer-aided CBT self-help for anxiety and depressive disorders: experience of a London clinic and future directions. *Journal of Clinical Psychology*, *60*(2), 147–157. https://doi.org/10.1002/jclp.10241

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*, *10*, 113. https://doi.org/10.1186/1471-244X-10-113

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, *41*(2), 159–164. https://www.ncbi.nlm.nih.gov/pubmed/3516036

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: causal mediation of change. *Journal of Consulting and Clinical Psychology*, *61*(4), 678–685. https://doi.org/10.1037/0022-006x.61.4.678

ieso. (2022). https://www.iesogroup.com/

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice* (3), 294–303. https://doi.org/10.1037/cps0000079

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., Efthimiou, O., Miguel, C., Maas genannt Bermpohl, F., Furukawa, T. A., Cuijpers, P., ... & Forsell, Y. (2021). Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA psychiatry*, *78*(4), 361-371.

Kessler, D., Lewis, G., Kaur, S., Wiles, N., King, M., Weich, S., … Peters, T. J. (2009). Therapist-

delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. *The Lancet*, *374*(9690), 628–634. https://doi.org/10.1016/S0140-6736(09)61257-5

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., … Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26*(1), 22–36. https://doi.org/10.1017/S2045796016000020

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

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Lowe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–326. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Lutz, W., Arndt, A., Rubel, J., Berger, T., Schröder, J., Späth, C., … Moritz, S. (2017). Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression. *Journal of Medical Internet Research*, *19*(6), e206. https://doi.org/10.2196/jmir.7367

Magaard, J. L., Seeralan, T., Schulz, H., & Brütt, A. L. (2017). Factors associated with help-seeking behaviour among individuals with major depression: A systematic review. *PloS One*, *12*(5), e0176730. https://doi.org/10.1371/journal.pone.0176730

Marks, I. M. (1986). *Behavioural psychotherapy: Maudsley pocket book of clinical management*. Wright/Institute of Psychiatry Publishing.

Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., … Marks, I. M. (2005). Reliability and validity of the work and social adjustment scale in phobic disorders. *Comprehensive Psychiatry*, *46*(3), 223–228. https://doi.org/10.1016/j.comppsych.2004.08.007

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. *The Journal of Clinical Psychiatry*, *74*(6), 595–602. https://doi.org/10.4088/JCP.12r07757

McKnight, P. E., & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment

research. *Clinical Psychology Review*, *29*(3), 243–259. https://doi.org/10.1016/j.cpr.2009.01.005

McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. NHS Digital.

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

Muthén, B., & Muthén, L. K. (2000). Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling With Latent Trajectory Classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882–891.

Rapaport, M. H., Clary, C., Fayyad, R., & Endicott, J. (2005). Quality-of-Life Impairment in Depressive and Anxiety Disorders. *American Journal of Psychiatry*, *162*(6), 1171–1178. https://doi.org/10.1176/appi.ajp.162.6.1171

Rayner, C., Coleman, J. R. I., Purves, K. L., Carr, E., Cheesman, R., Davies, M., … al., E. (2020). *Sociodemographic inequalities in lifetime treatment seeking and receipt for common mental health problems*. https://doi.org/10.31234/osf.io/vsu4g

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *30*(1), 79–96. https://doi.org/10.1080/10503307.2019.1566676

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

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

Skelton, M., Carr, E., Buckman, J. E. J., Davies, M. R., Goldsmith, K. A., Hirsch, C., … al., E. (2021). *Trajectories of depression and anxiety symptoms during psychological therapy for common mental health problems*. https://doi.org/10.31234/osf.io/8scpx

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, 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

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, *338*, b2393. https://doi.org/10.1136/bmj.b2393

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy*, *50*(6), 374–380. https://doi.org/10.1016/j.brat.2012.03.005

Thibodeau, M. A., Quilty, L. C., De Fruyt, F., De Bolle, M., Rouillon, F., & Bagby, R. M. (2015). Latent classes of nonresponders, rapid responders, and gradual responders in depressed outpatients receiving antidepressant medication and psychotherapy. *Depression and Anxiety*, *32*(3), 213–220. https://doi.org/10.1002/da.22293

van der Nest, G., Lima Passos, V., Candel, M. J. J. M., & van Breukelen, G. J. P. (2020). An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, *43*, 100323. https://doi.org/10.1016/j.alcr.2019.100323

Wallert, J., Boberg, J., Kaldo, V., Mataix-Cols, D., Flygare, O., Crowley, J. J., … Rück, C. (2022). Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Translational Psychiatry, 12*(1), 1-10. https://doi.org/10.1038/s41398-022-02133-3

Webb, C. A., Rosso, I. M., & Rauch, S. L. (2017). Internet-Based Cognitive-Behavioral Therapy for Depression: Current Progress and Future Directions. *Harvard Review of Psychiatry*, *25*(3), 114–122. https://doi.org/10.1097/HRP.0000000000000139

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Boerescu, D., & Attiullah, N. (2008). Remission in depressed outpatients: more than just symptom resolution? *Journal of Psychiatric Research*, *42*(10), 797–801.

Chapter 6: General Discussion

6.1 General overview of findings

The overarching aim of this thesis was to better understand treatment-related phenotypes in depression and anxiety. I assessed methods of measuring depression and anxiety for genetic studies, specifically, reported pharmacological treatment and scales of symptom severity and functional impairment (Chapters 2 and 3). I also investigated longitudinal patterns of the same scales during psychological therapy (Chapters 4 and 5). This final chapter presents a discussion of the findings from the four empirical studies in relation to one another. I will first provide an overview of the key findings from each study and then contextualise these findings in terms of the existing literature. I will then discuss the strengths and limitations of the statistical methods, samples, and measures, before outlining directions for future research.

In the first study, Chapter 2, I explored whether self-reported medication use is a feasible method to ascertain depression and anxiety cases, and identify additional cases, when clinical information is unavailable or sparse. Genetic correlations were high (r_g = 0.6 - 0.7) between use of antidepressant or anxiolytic medication in the UK Biobank and existing case-control phenotypes derived from diagnostic questionnaires or interviews. Polygenic risk scores (PRS) from the diagnostic measures also indicated genetic overlap, significantly predicting variance in medication use. A minority of participants reported antidepressant or anxiolytic medication but were not identified as cases by other methods available in the UK Biobank, such as hospital records or reported help-seeking. Lower genetic overlap between this 'novel' group and the existing diagnostic phenotypes indicated noisier genetic signal and unlikely true case-status. The results suggested that this brief phenotype would be useful as the sole method of case ascertainment only where no better validated phenotypes are available, rather than identifying *additional* cases to supplement existing measures.

In Chapter 3 I investigated other depression and anxiety treatment-related phenotypes in a sample with lifetime experience of depression and/or anxiety from the Genetic Links to

Anxiety and Depression (GLAD) Study. The three phenotypes were measures of depression symptoms (PHQ9), anxiety symptoms (GAD7) and functional impairment (WSAS). A heritable component was found for each, and the genetics underlying symptom severity were largely the same as those for functional impairment (r_g = 0.8 - 0.9). This suggested that, although impairment is important to measure from a diagnostic and patient perspective, genetic studies using symptom severity as a brief phenotype will capture most of the same variants. The results therefore added to the existing body of evidence for analysing symptom measures when full diagnostic criteria are unavailable. Moreover, the results suggested that genetic information from genome-wide association studies (GWAS) of symptom severity is sufficient for prognostic prediction models, as genetic data on functional impairment would be unlikely to explain substantial additional variance. Across both Chapters 2 and 3, high genetic correlations offered new additional evidence for brief phenotypes to identify depression and anxiety in genetic studies.

Chapters 4 and 5 used growth mixture modelling to investigate the existence of underlying subgroups of patients with similar treatment outcome trajectories during psychological therapy for depression or anxiety. Both studies used data from Improving Access to Psychological Therapies (IAPT) services, and the same measures were explored as in Chapter 3. The first study was performed with electronic treatment records of over 16,000 patients who received high-intensity psychological therapy from physical services in South London. The second study investigated trajectories in a sample of more than 50,000 patients who had received cognitive-behavioural therapy (CBT) delivered in real-time via the internet. In both chapters, the trajectories of depression and anxiety symptoms were investigated. In Chapter 5, functional impairment was also modelled. Across all outcomes in both studies, individual variation in trajectories could be captured by four latent classes. The shape of the trajectories was highly similar across the measures: three moderate-severe trajectories showing no change (‘plateau’), fast improvement, and gradual improvement, respectively, and a mild trajectory with minimal change. Trajectory classes had different associations with routinely recorded patient variables, which again were broadly consistent across the studies.

6.2 Discussion of findings in the context of core themes and previous literature

6.2.1 Brief phenotyping with treatment-related measures can be useful when detailed information is scarce or unavailable

Depression and anxiety are common disorders arising from the complex interplay of a variety of aetiological factors. They are extremely costly and burdensome both to the individual and society. Depression and anxiety are moderately heritable, and the genetic factors that underlie them appear to be largely shared. This similarity can be leveraged by studying them as a single 'internalising' phenotype. Yet, progress in identifying genetic variants associated with depression and anxiety has been hindered by several factors. Foremost, they are highly polygenic, such that they are influenced by a large number of genetic variants each of very small effect size. Two approaches to increasing statistical power for detecting variants have been highlighted: decreasing sample heterogeneity and increasing sample size. The latter can be achieved through resource-saving phenotyping methods that are less detailed than diagnostic interviews or questionnaires, called 'brief phenotyping'.

Recent successes in variant discovery have largely been due to the use of brief phenotyping (e.g., Levey et al., 2020). Support for the validity of brief phenotypes primarily comes from high genetic correlations with cohorts using more diagnostic methods of case ascertainment as well as across measures of differing levels of detail within cohorts (e.g., Howard et al., 2019). In Chapter 2, most medication use cases in the UK Biobank sample (> 85%) were already captured by the item of help-seeking for internalising symptoms. This resulted in comparatively few 'novel' medication use cases, who conferred noise rather than relevant genetic signal. Help-seeking is a prerequisite of obtaining prescription medication and would also capture those receiving other treatments. This highlights how the potential of a brief phenotype needs to be considered in terms of the specific study and other available measures. For example, although arguably more useful than the medication item in the UK Biobank, help-seeking phenotypes, including self-report of a diagnosis from a medical professional, will fail to capture a proportion of true cases due to the 'treatment gap'. An optimal brief phenotype may therefore be experience of lifetime core symptoms, which a recent study found accounted for most of the heritability in a

diagnostic depression questionnaire (Jermy et al., 2021). On the other hand, lifetime symptom measures are not always available and therefore alternate methods of defining cases will continue to be useful, whether in isolation or for triangulation with other less clinically precise methods to improve reliability (Glanville et al., 2021).

Current symptom score measures have also been successfully used as brief phenotypes of depression and anxiety in genetic studies (Direk et al., 2017; Levey et al., 2021; Purves et al., 2020). Compared with lifetime diagnostic measures, these may be less influenced by recall bias, quicker to complete and, when used as a quantitative trait, provide greater statistical power (Yang et al., 2010). Symptom scores are also not reliant on accurate prevalence values for conversion to liability scale estimates. Furthermore, a symptom-based approach is less limited by the validity of diagnoses (Insel et al., 2010; Kendler, 2016). The use of brief phenotypes is based on the theory that the genetic liability to mental health disorders lies on a normal distribution, and thus many of the same genetic variants influence subclinical levels. There is mounting evidence for this, to which Chapter 3 contributed by showing high genetic correlations between symptom severity and functional impairment, given that impairment is a key "missing" element of diagnoses when compared to symptom severity. However, there was possible indication of modest unique genetic variance as each genetic correlation slightly, but significantly, differed from unity. This warrants further investigation. The results also suggested that functional impairment scores could similarly be used to phenotype depression and anxiety if other measures, including symptom scales, were unavailable. This is likely to capture a broader phenotype, consistent with previous suggestions that functional impairment may represent a transdiagnostic method of identifying psychiatric cases to maximise sample size for genetic studies (McGrath et al., 2013). The brief phenotypes of Chapters 2 and 3 could feasibly complement each other by identifying slightly different groups. Supplementing scales of current symptoms or impairment with questions about current treatment use might capture individuals currently below clinical ranges due to the effectiveness of treatment as well as those who have not sought or received treatment but who present severely on the scales.

The future of psychiatric genetics lies in both brief phenotyping to increase sample sizes, and detailed diagnostic techniques to improve specificity. These approaches are complementary, with brief phenotyping offering a logical, practical approach while detailed clinical information is collected. Brief phenotypes have revealed significant loci with plausible relationships to brain regions, biological processes, and other phenotypes (Howard et al., 2019; Levey et al., 2020). Detailed diagnostic phenotypes can then be used for further investigations of possible variants, restricted to smaller, less heterogeneous, subsets of patients that may be especially informative. This is a central part of the Psychiatric Genomic Consortium's current work (Sullivan et al., 2018). Brief phenotyping is not without its limitations, but appears to be a practical, resource-saving method of identifying likely psychiatric cases and increasing power for polygenic prediction.

6.2.2 There is meaningful individual variation in trajectories of treatment outcomes during in-person and internet-enabled psychological therapy

There appear to be subgroups during psychological therapy such that patients do not all exhibit a similar longitudinal response pattern. Prior to the work in this thesis, most existing studies were arguably underpowered due to small sample sizes and had not explored trajectories beyond the first few sessions (e.g., Lutz et al., 2014). Furthermore, to my knowledge existing studies had not investigated trajectories in an internet-delivered format consisting of real-time CBT sessions with a therapist (internet-enabled CBT). Chapters 4 and 5 differed from one another not only in the modality of therapy delivery but in the number of minimum sessions received and the therapies offered, with the in-person sample receiving a variety of high-intensity treatments as opposed to being restricted to CBT. In line with evidence of equivalence across therapy types and delivery methods (e.g., Cuijpers et al., 2019, 2021) results were largely consistent across the two studies in terms of number and shape of trajectories, as well as associations with baseline variables. The optimal models selected for symptoms were also highly similar to those from analysis in two North London IAPT services that had approximately a third of the sample size and several methodological differences (Saunders et al., 2019).

Relationships between patient variables and trajectory classes provided information that could potentially guide decision-making, improve service efficiency and enable timely adjustments. This information might be especially useful to highlight where a patient may benefit from continued treatment despite apparent non-response. Some of the identified associations confirmed observations from studies of endpoint outcomes, for instance, not being in employment is a consistent correlate of both non-recovery and the moderate-severe plateau trajectory classes. Other associations with trajectory classes were more informative as they distinguished between different rates of improvement. A group of particular clinical importance was the gradual improvement class, who could be at risk of inappropriate early alteration or termination of treatment. Compared with odds of the moderate-severe plateau trajectory class, higher baseline functional impairment showed slightly lower odds of being in the fast or mild symptom classes, but no difference in the odds for the gradual class. Splitting the functional impairment scale into items in Chapter 5 revealed slight nuances to this relationship for the anxiety trajectory classes, indicating the value of item-based analyses. Furthermore, higher depression symptoms showed lower odds of fast improvement in anxiety or impairment scores, but not of gradual improvement. Therefore, unlike characteristics that are associated with lower likelihood of any of the favourable trajectories (e.g., unemployment), these do not necessarily reduce the possibility of improvement. However, where improvement occurs, it will likely be gradual. The similar findings across these studies hints at the possibility that, generally, conclusions regarding treatment outcomes from in-person therapy samples can be generalised to internet-enabled therapy.

6.2.3 Functional impairment

Results from Chapter 3 indicated that, despite only moderate phenotypic overlap, symptom severity is highly genetically similar to functional impairment. This may have value in prediction models of disorder risk and treatment response, which could use genetic information based on symptoms for both symptom and impairment outcomes. Phenotypic correlations with functional impairment were stronger for depression than anxiety symptoms in both Chapters 3 and 5. Furthermore, associations with baseline variables for impairment trajectories in Chapter 5 were more similar to those for depression symptom trajectories. As discussed in Chapter 3, this might indicate that the

WSAS is more sensitive to depression symptoms, or specifically, the PHQ9 in comparison with the GAD7. This should be considered when working with these measures, as impairment may be especially impacted by GAD symptoms of impaired concentration or sleep problems (Fried & Nesse, 2014), which are not assessed by the GAD7.

Chapter 5 also showed that functional impairment follows similar trajectories to symptoms during psychological therapy. The modelling technique did not allow us to determine temporal patterns between the three outcomes for the same individuals and therefore whether change occurred at a similar rate. However, most patients (roughly 90%) were allocated to the functional impairment model moderate-severe plateau or mild, minimal improvement trajectory class. This higher proportion than observed in the symptom models suggested that, within the sessions modelled, improvement in functional impairment was less likely than in symptoms. This is broadly consistent with reports of a lag in improvement of impairment following symptom improvement.

6.3 Strengths and limitations

The analyses in this thesis were performed with samples from four different sources and using multiple methods. General advantages and limitations to each of these will be discussed here.

6.3.1 Methods

6.3.1.1 Genome-wide association studies do not capture all genetic influences

Two of the chapters included genome-wide association studies (GWAS; Chapters 2 and 3). Although the main purpose of GWAS is typically to identify significantly associated variants, here it was performed primarily to enable secondary analyses with the resulting summary statistics, such as genetic correlations. GWAS is a robust, hypothesis-free method that has revealed many replicable findings for depression and anxiety, and subsequent analyses have provided biological insights about their aetiology. However, GWAS has several limitations that will have impacted secondary analyses, and these limitations also apply to estimates of heritability using individual-level data (GCTA-GREML). First, if causal variants

are neither genotyped directly nor tagged by other genotyped variants, their effects will be missed. Furthermore, the methods used here were limited to common variants and will therefore not have captured rare variants that, under a selection model, may have larger effects. Rare copy number variants have been implicated to have a role in depression, highlighting the possible loss of information here (Kendall et al., 2019). Although genetic correlations are independent of heritability, such that correlations can be high for phenotypes with low, non-zero heritability and vice versa, heritability does reflect the power to detect a significant correlation. Therefore, underestimates of heritability due to failure to capture causal variants can produce inaccurate genetic correlations. Second, interactions were not modelled. Current understanding is that interactions between variants contribute little to genetic variance (Hill et al., 2008; Polderman et al., 2015). However, our lack of understanding of interactions between genes and the environment is more problematic. This is essential for devising interventions and preventions using environmental modifications, including psychological therapy. Significant interactions have proven difficult to identify, with an even greater multiple testing burden and a lack of well-defined and measured environments (Tam et al., 2019). Polygenic risk scores provide an approach to assessing gene-environment interactions (e.g., Colodro-Conde et al., 2018) but more finely characterised, longitudinal cohorts with genetic data may be required to investigate further. Gene-environment correlations further complicate the relationship between genotype and phenotype. Family-based designs can help to disentangle these influences and avoid spurious inflation of genetic influence (Morris et al., 2020; Young et al., 2018). Finally, like all GWAS, GWAS of self-reported medication use or symptom scales for variant discovery would require follow-up analyses, such as fine-mapping and functional annotation, to elucidate causal variants and associated biological mechanisms.

6.3.1.2 Growth mixture modelling as a useful heuristic of a complex reality

Growth mixture modelling can reveal hidden subgroups in longitudinal data, informing a person-centred approach to mental health treatment. However, there are criticisms of this method. The grouping of patients into classes seemingly contradicts the narrative of the dimensionality of mental health that has featured throughout this thesis. Several considerations are important here: growth mixture models output a probability of belonging to each trajectory class for each individual, and there is variance within classes,

allowing patients to deviate from one another to some extent. Ideally, I would have weighted our regression models with baseline characteristics by the class probabilities such as in the 'three step' method implemented in Mplus software. However, this did not allow for simultaneous multiple imputation of missing data in the baseline variables. With unlimited time, I would have liked to run the 'three step' method, or weighted regressions in R, as a sensitivity analysis. The large sample size and similar results across samples does perhaps suggest that findings would remain largely the same. Furthermore, as much as grouping can be considered crude, decisions cannot be made on the basis of dimensions. Even in the presence of confidence ratings for interpretation of a statistical recommendation, thresholds have to be imposed to determine how to treat a patient.

Growth mixture modelling requires a-priori expectations of the number and shape of trajectories, ideally from previous literature, as there is no single index of an optimal model. Model selection was guided by existing studies showing three to five trajectory classes and considerations including fit indices. This degree of subjectivity in model selection means it is possible that others would argue for a different one. Some of the trajectory classes may reflect regression to the mean or spontaneous remission, nonetheless, this does not lessen their potential to provide guidance of outcome change timings, whether directly due to treatment or not. Moreover, the finding of three groups with similar intake scores but different slopes renders regression to the mean a less likely explanation.

It is an assumption that the classes from growth mixture models represent real entities found in clinical practice rather than spurious creations of the model specification. Whether the classes 'make sense' is a key, yet highly subjective, consideration (Ram & Grimm, 2009). It seems fair to say that not all patients do respond at the same rate, but whether they belong to meaningful, qualitatively different subgroups cannot be known. This assumption can, however, be supported by further evidence in the form of replication in independent data, and exploration of expected relationships with other variables (Ram & Grimm, 2009). Such congruence was largely seen between the findings from Chapters 4 and 5, which were also both similar to a prior existing study (Saunders et al., 2019). Furthermore, observations such as lower odds of a mild anxiety trajectory for an anxiety problem descriptor provide additional face validity for the groups. In summary, growth

mixture models are a heuristic that represent a far more complicated reality. However, they draw attention to the risks of considering patients as a single group with similar expectations as to how they will respond to treatment and also provide information that could be used in clinical decision-making and monitoring of patient progress.

6.3.2 Samples

6.3.2.1 Ancestral diversity in volunteer-based cohorts

A critical priority in genetic research is ensuring that findings are applicable across diverse ancestries. The patient sample from South London and Maudsley (SLaM) IAPT services illustrated that minoritised ethnic groups in the UK are using mental healthcare services, yet study samples meant to inform our understanding of treatment do not reflect this (although note that ancestry and ethnicity are not the same; Peterson et al., 2019). UK Biobank and the GLAD Study are excellent "big data" resources for mental health research but do not represent the UK population; both cohorts are heavily skewed to White, highly educated women of European ancestry. This is a general problem throughout genetics research. As of January 2019, 78% of individuals in GWAS studies were of European ancestry, and only 2% African (Sirugo et al., 2019). Trait-associated variants are often the same across ancestries, but allele frequencies differ. Variants that are highly relevant in a population may therefore be low frequency or completely absent in European populations, thereby conferring little or no risk. Furthermore, population differences in linkage disequilibrium mean that a variant that captures the effects of the causal variant in one population may not do so in another. A focus on European ancestry therefore has the potential to exacerbate health disparities. PRS derived from European GWAS perform poorly in populations of other ancestries which could lead to over or underestimates of risk (A. R. Martin et al., 2019). Efforts to oversample from specific populations are essential to tackle this problem. Some such studies already exist and show impressive progress, for example, Genes & Health which has genotyped over 50,000 Bangladeshi and Pakistani individuals in England (Finer et al., 2020). Recruitment of diverse samples will be insufficient; we need ancestry-specific sequenced reference genomes and large ancestry-specific genotyping arrays, although whole genome sequencing will be useful here. Finally, cross-ancestry analysis can increase sample size, magnify variant associations, and improve

PRS prediction and accuracy of effect size estimates (Peterson et al., 2019). Several of these advantages were demonstrated in a recent multi-ancestry depression GWAS of almost one million participants (Meng et al., 2022).

6.3.2.2 Routine treatment records from IAPT; a rich yet 'noisy' research resource

Two studies in this thesis capitalised on the wealth of routinely recorded data available from IAPT electronic treatment records. As well as large sample sizes, IAPT samples provide greater representation and generalisability than clinical trials of depression and anxiety treatments. Unlike IAPT services, clinical trials commonly exclude patients experiencing substance use disorders, symptoms considered too mild or severe, psychotic symptoms, or high suicidal risk (Lorenzo-Luaces et al., 2018). A study within IAPT specifically recommended against exclusion on the basis of alcohol use, which, whilst high, was not associated with recovery (Buckman et al., 2018). One review reported that more than 75% of psychotherapy or antidepressant trials for depression excluded patients due to a substance use disorder (Lorenzo-Luaces et al., 2018), despite high comorbidity with depression. On the other hand, the needs of complex cases in IAPT may be unmet by the services (C. Martin et al., 2022). Finally, the availability of remote technologies to deliver IAPT treatments offers research opportunities to investigate these in a routine setting. Existing evidence of prognostic predictors for internet-delivered therapy is inconclusive for variables such as symptom severity, which may be due a reliance on data from clinical trials of patients with milder presentations. Further analysis of routine treatment data is required as the use of internet-delivered treatment is likely to continue to increase to meet patient demand.

The greater external validity of routine data introduces greater potential for confounders. Some potential confounders can be controlled for in statistical analysis, such as service-level differences which are associated with patient outcomes (D. M. Clark et al., 2018). The specific IAPT service did not appear to exert large influence on the trajectory models in Chapter 4, however, it may be important at a broader geographical level. Stratification or clustering by service might be more informative than statistical control. This applies to other variables, such as use of medication that might influence outcomes in combination with psychotherapy (Cuijpers et al., 2020). Generally, the absence of a control group limits

the attribution of outcome changes to therapy. The benefits of a control group are discussed in the future directions section.

Pre-processing of the IAPT data used in Chapter 4 resulted in a substantial proportion of the work involved in this thesis: investigation of variables, discussions with IAPT professionals (including both clinicians and analysts) and higher-level decision making. As an example, many of the responses were in the format of free text as opposed to category selection. The IAPT technical output specification identifies 12 high-intensity and 10 low-intensity treatments, but the data I received had over 100 distinct treatment values. The 'step' variable that should indicate intensity did not clearly map onto these and was deemed unreliable by IAPT professionals. Therefore, to restrict to a sample who had received high-intensity treatment, I mapped treatments to intensity levels. I consulted IAPT professionals and where they were unsure, I explored data on variables such as average number of sessions or type of therapist. Generally, there were issues determining the reliability of variables. One such variable was the reason for the end of treatment; within the Chapter 5 data, it was deemed too unreliable for use at all. Another key decision that applied to both IAPT samples was the extent of interval variation permitted between sessions. This was considered important to satisfy model assumptions and as response rates can differ depending on session frequency (Bruijniks et al., 2020; Saunders et al., 2020). However, highly similar results were found in the previous trajectory study with no such filtering (Saunders et al., 2019), suggesting this may be unnecessary in future analyses.

Differences in the number of sessions received by patients complicates modelling and may reflect different processes (Barkham et al., 1996; Bone et al., 2021). NICE recommends around 12-20 sessions for high-intensity treatment but the average in the present samples was roughly 7, including assessment. I subsequently limited the models to up to 10 treatment sessions, as comparatively few patients had received more than this. For greater confidence in the validity and reliability of the trajectories at later sessions, more data for those timepoints is required. This issue is observed in other IAPT samples (Barkham & Saxon, 2018; Robinson et al., 2020), as well as to some extent in clinical trials.

6.3.3 Measures: PHQ9, GAD7 and WSAS

Three self-report score-based measures played a vital role in this thesis: the PHQ9 measuring depression symptoms, the GAD7 measuring anxiety symptoms and the WSAS measuring functional impairment. The brevity of these makes them ideal for repeated assessments and they have good psychometric properties. The total score from each measure was used, rather than creating categories with arbitrary thresholds. Indeed, binary definitions of endpoint treatment outcome require greater standardisation as inconsistencies result in substantial discrepancies between studies (Loerinc et al., 2015). Furthermore, as previously described, quantitative measures are not reliant on the validity of diagnoses and have been shown to better represent psychopathology (Haslam et al., 2012; Markon et al., 2011). It is likely that treatment outcomes will continue to be defined by binary measures as thresholds are required to make clinical decisions. However, the results here highlight the benefit of complementing binary definitions with quantitative scales that do not rely on pre-specified thresholds, and monitoring these *during* therapy.

Missing data on the WSAS work item was a recurring problem. Interestingly, this goes without mention in the majority of studies using the WSAS, with no indication as to how it is handled. In IAPT, individual mean imputation is performed using the other four items, but this may be inappropriate if missingness is related to the true values. Sensitivity analyses were performed in Chapters 3 and 5 of WSAS four-item scores. Although trajectory models were highly similar, the associations with baseline variables slightly differed. The sample without the work item in Chapter 3 was insufficient for a thorough investigation but analysis of a four-item score in the whole sample provided some support. Ideally, another measure of functional impairment would have been available for comparison. An objective measure of impairment would also be informative, as negative cognitive biases seen in depression and anxiety (Mathews & MacLeod, 2005) may lead to lower subjective ratings of impairment on the WSAS.

The high comorbidity between depression and anxiety, as well as commonalities including genetic profile, effective treatments and symptom overlap, render it sensible to study them together. A strength of this thesis is the inclusion of patients with anxiety disorders. Historically, adult anxiety disorders have been understudied compared with depression

(Kroenke et al., 2007). This imbalance has continued into much of the current personalised treatment research (Cohen & DeRubeis, 2018) and studies of internet therapy (Massoudi et al., 2019). Anxiety disorders are also under-recognised by General Practitioners (GPs) (Buszewicz & Chew-Graham, 2011; Wittchen et al., 2002); a greater research focus could increase clinical attention. As well as maximising sample size, the combined study of depression and anxiety is consistent with the concept of internalising disorders. This broader category is less susceptible to diagnostic changes and decisions (Kendler, 2016) and has useful transdiagnostic implications (e.g., O'Driscoll et al., 2021). However, other 'fear-based' anxiety disorders may be more phenotypically and genetically distinct from depression and GAD (L. A. Clark & Watson, 2006; Morneau-Vaillancourt et al., 2020), and are not as well assessed by the GAD7. Future research would likely benefit from making this distinction. IAPT guidelines specify use of an anxiety-disorder specific measure in place of the GAD7 where appropriate, yet I found the completion of these to be very low. Assessment of anxiety disorder subtypes within IAPT and other data sources would permit research into exactly how these differ in terms of treatment outcomes both during and at the end of therapy.

6.4 Future directions

6.4.1 Genetics of depression and anxiety - what's next in the field

As whole genome sequencing (as opposed to SNP array-based genotyping) becomes more accurate and affordable it will become more commonly used. This will improve the identification of rare variants as well as variant associations in samples of non-European ancestries. Meanwhile, large, fully sequenced reference cohorts will permit accurate imputation of a greater number of rare variants for GWAS using SNP arrays. Continued efforts to increase sample size are recommended for both genotyping and sequencing. GWAS with increasingly large samples will be valuable until the discovery of associated common variants plateaus (Wray et al., 2018) and existing whole-genome sequencing studies suggest effect sizes will require substantial statistical power (Kendall et al., 2021). Brief phenotyping therefore continues to have a place at the psychiatric genomics table. Linkage of genotyped cohorts with electronic health records offers one approach to this and has already led to the most successful depression GWAS in terms of variant discovery

to date (Levey et al., 2021). Furthermore, information can be maximised from existing cohorts by identifying alternate phenotypes, as per Chapter 2. If participants have already been phenotyped, genotype data can be collected retrospectively as the effects of DNA are largely stable from birth. As already described, efforts to increase sample size will need to be complemented by detailed phenotyping of smaller, clinically informative samples in which genetic variants identified by brief phenotypes can be dissected and depression and anxiety-specific variants found (Schwabe et al., 2019).

6.4.2 Symptom severity and functional impairment

Our interpretation of the phenotypic and genetic relationship between symptom severity and functional impairment could be further informed using other variables. For example, stressors like financial problems or clinical factors such as chronicity could feasibly impact coping resources and thus lower the symptom threshold at which an individual reports greater impairment. Meanwhile, variables such as social support might increase the threshold. These interactions would result in incomplete phenotypic overlap. Indeed, social isolation, socioeconomic status, and coping behaviours are phenotypically associated with functional impairment (Brown et al., 2007; Chow et al., 2022). Cognitive ability (Smagula et al., 2015) and personality factors are also (Verboom et al., 2011), as well as being genetically associated with symptom severity. As described in Chapter 3, various tools are available to investigate the mechanism underlying the genetic correlation between symptom severity and functional impairment (e.g., mtCOJO; Zhu et al., 2018).

6.4.3 Other factors to explore in studies of psychological treatment

There are several other factors that likely influence individual treatment outcomes but have yet to be studied in trajectory-based analyses. Non-specific or 'process' treatment factors are arguably more amenable to change to improve patient prognosis than other variables. Therapist effects are thought to account for 5% of differences in treatment outcome (Johns et al., 2019) and personalised allocation to a therapist can increase likelihood of a positive outcome (Delgadillo et al., 2020). Other variables that could impact trajectory class include chronicity, comorbidities, social support and outcome expectancy.

Time-varying covariates that might exert influence on outcome trajectories could also be incorporated into models, for example, medication changes, stressful life events, and CBT homework compliance (Callan et al., 2019). Finally, treatment is not an isolated event. Each patient has their own mental health history, potentially including past treatment. The point of access could in itself confer benefits; relief at getting help, the process of talking and being listened to, receiving a label that 'explains' one's problems. We might observe individual differences prior to treatment formally commencing, as well as during the period leading up to seeking treatment. Longitudinal cohorts would be especially revealing about the whole process of treatment seeking, access and response.

To understand individuals who seek treatment but do not attend, or only attend some of, their session, greater research attention is required for dropout and non-attendance. Of the 1,456,446 referrals to IAPT in 2020-2021, only 44% completed treatment (IAPT Team, NHS Digital, 2021). In a sample of 363 GP referrals to IAPT, 50% never attended and these individuals were more likely to have higher risk scores and suicidal ideation, indicating a high treatment need (Di Bona et al., 2014). Qualitative interviews of patients who did not attend IAPT treatment revealed themes of waiting time, lack of individualisation, difficulty attending rigid appointment times during the working day, and concerns about transport and safety accessing later evening appointments (Marshall et al., 2016). Internet-enabled therapy will hopefully help to close the treatment gap by offering an effective alternative to in-person therapy that removes some of these barriers. More reliable recording of the reason for the end of treatment would permit larger studies of characteristics associated with dropout from IAPT, and thus alert therapists to patients at risk. Dropout after some treatment might also be identifiable from specific trajectory classes. A potential avenue to decrease dropout is more frequent treatment, over the same number of sessions (Bruijniks et al., 2020). However, resources may not be available for this in IAPT so additional approaches need to be investigated.

Functional impairment was completed at a lower rate than symptom measures in the IAPT samples and is not included in IAPT outcomes such as recovery or reliable improvement. This is surprising given the original rationale of IAPT services was to lessen the impact of common mental health disorders on the economy. Arguably, if a person has sought

treatment, they have been motivated to do so due to distress or impairment and thus fulfil this diagnostic criterion for depression and most anxiety disorders. In fact, the inclusion of functional impairment as a diagnostic criterion is a source of great debate, with critics arguing that it is a consequence rather than a feature of the disorder, is poorly defined and could lead to prodromal stages being ignored (L. A. Clark et al., 2017). However, functional impairment provides an indicator for treatment prioritisation, avoids pathologising normal variation of emotions, and is of great importance to patients (L. A. Clark et al., 2017; Zimmerman et al., 2006). Therefore, broadening treatment outcomes to include functional impairment would offer a more informative reflection of the effectiveness of treatment and possibly help to individualise treatment by highlighting greatest domains of impairment. It would also be worthwhile to explore which therapeutic approaches or adjunctive support (e.g., financial advice) are most effective at improving functioning.

6.4.4 Item-level analyses provide a richer source of information than total sum scores

Item-level genetic and trajectory-based analysis could be especially informative due to the heterogeneity of symptom profiles between patients with depression and anxiety. This granular approach is not at odds with a broader transdiagnostic focus on internalising disorders, rather, analyses can be performed of depression and anxiety symptoms. There is existing support for item-specific response to treatment, for example, improvement appears to be greater in depressed mood than anhedonia (Dunn et al., 2019). Future analyses of trajectories during therapy and endpoint outcomes could therefore include questionnaire items as both explanatory and outcome variables. This could inform therapists about specific domains that require focus and also highlight where psychological treatment approaches are falling short, as in the case of improving positive affect.

There is also evidence that particular symptoms are associated with greater impairment (Fried & Nesse, 2014; Olatunji et al., 2007; Sacchetti et al., 2015). In a sample of around 3,000 patients with depression, sad mood and concentration problems explained the most variance in functional impairment, and hypersomnia the least (Fried & Nesse, 2014). Differences were also observed at the level of functional domain, for example, fatigue had strong associations with home management impairment but not close relationships. In

Chapter 5, items of the WSAS were included as separate baseline variables and revealed potential differences in relationships with symptom trajectory classes. For example, the home management item appeared to have an opposite direction of association for the mild class between the depression and anxiety models, indicating the value of this approach.

Finally, genetic studies have indicated that variant associations differ at the item level such that two patients with completely distinct symptom profiles could carry different risk variants (Cai et al., 2020; Nagel et al., 2018; Thorp et al., 2020). Functional impairment might similarly have specific domains under different genetic influences and specific patterns of associations with symptom items. Using total scores in this thesis may therefore have obscured potential differences in item-level phenotypic and genetic relationships.

6.4.5 Genetic prediction of outcome trajectories during psychological therapy

On the basis of the research presented in this thesis, a logical next step is to incorporate genetic factors in trajectory class models of treatment outcomes. This could be achieved by medical record linkage of individuals with genotype data who have accessed psychological treatment, for example, within GLAD Study participants (a combined dataset is being sought but not yet available). Developmental studies of depression symptoms across childhood and adolescence have already shown associations between trajectory classes and PRS (Lussier et al., 2020; Rice et al., 2018). Moreover, there is some evidence for genetic influences on psychological treatment response ('therapygenetics'; Lester & Eley, 2013). The maximum amount of variance that genetics can prognostically explain will be limited by the heritability of psychological treatment response. To date, analyses have been hampered by small samples and no significant SNP heritability estimate has yet been identified (Rayner et al., 2019). However, PRS for depression and intelligence have shown predictive value in models of endpoint treatment outcomes (Wallert et al., 2022), and an association has been reported between PRS for autism spectrum disorder and rate of symptom change (Andersson et al., 2019). PRS therefore might help distinguish subgroups who improve early in therapy compared with more gradually, as well as patients who do not show any reliable change. The only known attempt at this used a score of five candidate genes (Kelley et al., 2018) but the significant finding was likely spurious based on our

knowledge of the expected effect sizes. Ultimately, genetic and non-genetic predictors could be implemented into a clinical decision tool that would combine information from multiple variables to provide likelihood estimates of outcome trajectories.

Ideally, future studies will also explore genetic associations with treatment-specific response, to help inform personalised treatment allocation. Identifying prescriptive predictors of treatment outcome requires trials with a comparison treatment or control group (Cohen & DeRubeis, 2018). Clinical trials are thus able to make inferences about causality, however, they are often underpowered in comparison with the large samples available from routine treatment. Clinical trials informed by observational research may therefore be the most fruitful approach to detect treatment-specific effects and stronger evidence of causal associations. As an illustration of this, observational analyses of IAPT records informed a randomised trial which showed that assigning patients to treatment intensity using a machine-learning algorithm improved outcomes compared with stepped care (Delgadillo et al., 2022). In this thesis, a parallel can be drawn between GWAS of brief treatment-related phenotypes and trajectory-based analyses of electronic treatment records. Both maximise scale over 'depth' of information to identify statistically significant effects and require complementary smaller studies of greater precision.

It may sound implausible to suggest that genetic data could be used to predict treatment outcomes for patients with depression and anxiety. It is worth bearing in mind that a patient only needs to be genotyped once to assess their genetic risk profile for multiple traits, and these calculations can be done at any time after genotyping. PRS will likely only be useful for differentiating individuals at the very extremes of response. Despite this, any variable that explains even a small proportion of variance in treatment outcomes boosts the cumulative power of a multivariate model containing other clinically relevant variable. This therefore improves the chance of helping patients with these highly distressing, burdensome disorders. Whole genome sequencing is already offered in the NHS to identify markers of rare disorders and cancer. However, the clinical use of PRS for mental health disorders will not be possible until they are more representative across ancestries, and there are frameworks in place for communicating results, to avoid possible misinterpretation and stigmatisation (Palk et al., 2019). The need for greater diversity also

applies to treatment outcome research to ensure that results apply across patients from different ethnicities, as well as different intersectional identities. Any clinical tool would need to be validated within therapy services to determine whether information about predictions of patient trajectories is accurate and ultimately improves outcomes for patients.

6.5 Conclusions

Brief phenotypes of self-reported medication use and symptom scores can be used for genetic studies of depression and anxiety when other more suitable measures are scarce or unavailable. Information resulting from these studies could be used to calculate polygenic risk scores, which can be incorporated into prognostic prediction models. Although functional impairment is an important diagnostic consideration, it appears to share largely the same genetic variants that are associated with symptom severity. Patients show individual differences in trajectories of both symptom severity and functional impairment throughout psychological therapy. These differences can be captured by four classes which show associations with baseline characteristics, and this is observed for treatment delivered both in-person and via the internet. Information about variables associated with specific outcome trajectories could inform the development of clinical decision tools. Research is beginning to move towards combining electronic records of treatment outcomes with genetic information, providing large samples for analysis which will further inform our understanding of what factors influence different rates of recovery.

6.6 References

Andersson, E., Crowley, J. J., Lindefors, N., Ljótsson, B., Hedman-Lagerlöf, E., Boberg, J., … Rück, C. (2019). Genetics of response to cognitive behavior therapy in adults with major depression: a preliminary report. *Molecular Psychiatry*, *24*(4), 484–490.

Barkham, M., Rees, A., Stiles, W. B., Shapiro, D. A., Hardy, G. E., & Reynolds, S. (1996). Dose--effect relations in time-limited psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, *64*(5), 927.

Barkham, M., & Saxon, D. (2018). The effectiveness of high-intensity CBT and counselling alone and following low-intensity CBT: a reanalysis of the 2nd UK National Audit of Psychological Therapies data. *BMC Psychiatry*, *18*(1), 321.

Bone, C., Delgadillo, J., & Barkham, M. (2021). A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology*, *68*(2), 219–231.

Brown, C., Battista, D. R., Sereika, S. M., Bruehlman, R. D., Dunbar-Jacob, J., & Thase, M. E. (2007). Primary care patients' personal illness models for depression: relationship to coping behavior and functional disability. *General Hospital Psychiatry*, *29*(6), 492–500.

Bruijniks, S. J. E., Lemmens, L. H. J. M., Hollon, S. D., Peeters, F. P. M. L., Cuijpers, P., Arntz, A., … Huibers, M. J. H. (2020). The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *The British Journal of Psychiatry: The Journal of Mental Science*, 1–9.

Buckman, J. E. J., Naismith, I., Saunders, R., Morrison, T., Linke, S., Leibowitz, J., & Pilling, S. (2018). The Impact of Alcohol Use on Drop-out and Psychological Treatment Outcomes in Improving Access to Psychological Therapies Services: an Audit. *Behavioural and Cognitive Psychotherapy*, *46*(5), 513–527.

Buszewicz, M. J., & Chew-Graham, C. (2011). Improving the detection and management of anxiety disorders in primary care. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, *61*(589), 489–490.

Cai, N., Choi, K. W., & Fried, E. I. (2020). Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. *Human Molecular Genetics*, *29*(R1), R10–R18.

Callan, J. A., Kazantzis, N., Park, S. Y., Moore, C. G., Thase, M. E., Minhajuddin, A., … Siegle,

G. J. (2019). A Propensity Score Analysis of Homework Adherence-Outcome Relations in Cognitive Behavioral Therapy for Depression. *Behavior Therapy*, *50*(2), 285–299.

Chow, T. K., Bowie, C. R., Morton, M., Lalovic, A., McInerney, S. J., & Rizvi, S. J. (2022). Contributors of Functional Impairment in Major Depressive Disorder: a Biopsychosocial Approach. *Current Behavioral Neuroscience Reports*, *9*(2), 59–72.

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.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *18*(2), 72–145.

Clark, L. A., & Watson, D. (2006). Distress and fear disorders: an alternative empirically based taxonomy of the "mood" empirically based taxonomy of the "mood" and "anxiety" disorders. *British Journal of Psychiatry*, *189*, 481–483.

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review of Clinical Psychology*, *14*, 209–236.

Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., … Martin, N. G. (2018). A direct test of the diathesis-stress model for depression. *Molecular Psychiatry*, *23*(7), 1590–1596.

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and Acceptability of Cognitive Behavior Therapy Delivery Formats in Adults With Depression: A Network Meta-analysis. *JAMA Psychiatry*, *76*(7), 700–707.

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry: Official Journal of the World Psychiatric Association*, *19*(1), 92–107.

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., … Furukawa, T. A. (2021). Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry: Official Journal of the World Psychiatric Association*, *20*(2), 283–293.

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., … Barkham, M. (2022). Stratified Care vs Stepped Care for Depression: A Cluster Randomized Clinical Trial. *JAMA Psychiatry* , *79*(2), 101–108.

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, *85*(9), 835–853.

Delgadillo, J., Rubel, J., & Barkham, M. (2020). Towards personalized allocation of patients to therapists. *Journal of Consulting and Clinical Psychology*. https://doi.org/10.1037/ccp0000507

Di Bona, L., Saxon, D., Barkham, M., Dent-Brown, K., & Parry, G. (2014). Predictors of patient non-attendance at Improving Access to Psychological Therapy services demonstration sites. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2014.08.005

Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., … Sullivan, P. F. (2017). An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biological Psychiatry*, *82*(5), 322–329.

Dunn, B. D., German, R. E., Khazanov, G., Xu, C., Hollon, S. D., & DeRubeis, R. J. (2019). Changes in Positive and Negative Affect During Pharmacological Treatment and Cognitive Therapy for Major Depressive Disorder: A Secondary Analysis of Two Randomized Controlled Trials. *Clinical Psychological Science*, 2167702619863427.

Finer, S., Martin, H. C., Khan, A., Hunt, K. A., MacLaughlin, B., Ahmed, Z., … van Heel, D. A. (2020). Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *International Journal of Epidemiology*, *49*(1), 20–21i.

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PloS One*, *9*(2), e90311.

Glanville, K. P., Coleman, J. R. I., Howard, D. M., Pain, O., Hanscombe, K. B., Jermy, B., … Lewis, C. M. (2021). Multiple measures of depression to enhance validity of major depressive disorder in the UK Biobank. *BJPsych Open*, *7*(2), e44.

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*(5), 903–920.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, *4*(2), e1000008.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352.

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

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *Am J Psychiatry* (7), 748–751.

Jermy, B. S., Glanville, K. P., Coleman, J. R. I., Lewis, C. M., & Vassos, E. (2021). Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Molecular Psychiatry*, *26*(12), 7337–7345.

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

Kelley, M. E., Dunlop, B. W., Nemeroff, C. B., Lori, A., Carrillo-Roa, T., Binder, E. B., … Mayberg, H. S. (2018). Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse. *Depression and Anxiety*, *35*(10), 992–1000.

Kendall, K. M., Rees, E., Bracher-Smith, M., Legge, S., Riglin, L., Zammit, S., … Walters, J. T. R. (2019). Association of Rare Copy Number Variants With Risk of Depression. *JAMA Psychiatry*, *76*(8), 818–825.

Kendall, K. M., Van Assche, E., Andlauer, T. F. M., Choi, K. W., Luykx, J. J., Schulte, E. C., & Lu, Y. (2021). The genetic basis of major depression. *Psychological Medicine*, *51*(13), 2217–2230.

Kendler, K. S. (2016). The nature of psychiatric disorders. *World Psychiatry: Official Journal of the World Psychiatric Association*, *15*(1), 5–12.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Lowe, B. (2007). Anxiety

disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–326.

Lester, K. J., & Eley, T. C. (2013). Therapygenetics: Using genetic markers to predict response to psychological treatment for mood and anxiety disorders. *Biology of Mood & Anxiety Disorders*, *3*(1), 4.

Levey, D. F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., … Stein, M. B. (2020). Reproducible Genetic Risk Loci for Anxiety: Results From ∼200,000 Participants in the Million Veteran Program. *The American Journal of Psychiatry*, *177*(3), 223–232.

Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., … Gelernter, J. (2021). Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nature Neuroscience*, *24*(7), 954–963.

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review*, *42*, 72–82.

Lorenzo-Luaces, L., Zimmerman, M., & Cuijpers, P. (2018). Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants? *Journal of Affective Disorders*, *234*, 8–13.

Lussier, A. A., Hawrilenko, M., Wang, M.-J., Choi, K. W., Cerutti, J., Zhu, Y., … Dunn. (2020). Genetic susceptibility for major depressive disorder associates with trajectories of depressive symptoms across childhood and adolescence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. https://doi.org/10.1111/jcpp.13342

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., … Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology*, *82*(2), 287–297.

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychological Bulletin*, *137*(5), 856–879.

Marshall, D., Quinn, C., Child, S., Shenton, D., Pooler, J., Forber, S., & Byng, R. (2016). What IAPT services can learn from those who do not attend. *Journal of Mental Health* . https://doi.org/10.3109/09638237.2015.1101057

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591.

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 / the British Psychological Society*, *61*(1), 157–174.

Massoudi, B., Holvast, F., Bockting, C. L. H., Burger, H., & Blanker, M. H. (2019). The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. *Journal of Affective Disorders*, *245*, 728–743.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, *1*, 167–195.

McGrath, L. M., Cornelis, M. C., Lee, P. H., Robinson, E. B., Duncan, L. E., Barnett, J. H., … Smoller, J. W. (2013). Genetic predictors of risk and resilience in psychiatric disorders: a cross-disorder genome-wide association study of functional impairment in major depressive disorder, bipolar disorder, and schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, *162B*(8), 779–788.

Meng, X., Navoly, G., Giannakopoulou, O., Levey, D., Koller, D., Pathak, G., … Kuchenbaecker, K. (2022). Multi-ancestry GWAS of major depression aids locus discovery, fine-mapping, gene prioritisation, and causal inference. In *bioRxiv* (p. 2022.07.20.500802). https://doi.org/10.1101/2022.07.20.500802

Morneau-Vaillancourt, G., Coleman, J. R. I., Purves, K. L., Cheesman, R., Rayner, C., Breen, G., & Eley, T. C. (2020). The genetic and environmental hierarchical structure of anxiety and depression in the UK Biobank. *Depression and Anxiety*, *37*(6), 512–520.

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*(16), eaay0328.

Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., & van der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications*, *9*(1), 905.

O'Driscoll, C., Buckman, J. E. J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., … 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.

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

Palk, A. C., Dalvie, S., de Vries, J., Martin, A. R., & Stein, D. J. (2019). Potential use of clinical polygenic risk scores in psychiatry - ethical implications and communicating high polygenic risk. *Philosophy, Ethics, and Humanities in Medicine: PEHM*, *14*(1), 4.

Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C.-Y., Popejoy, A. B., Periyasamy, S., … Duncan, L. E. (2019). Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*, *179*(3), 589–603.

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709.

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A Major Role for Common Genetic Variation in Anxiety Disorders. *Molecular Psychiatry*, *25*(12), 3292–3303.

Ram, N., & Grimm, K. J. (2009). Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *International Journal of Behavioral Development*, *33*(6), 565–576.

Rayner, C., Coleman, J. R. I., Purves, K. L., Hodsoll, J., Goldsmith, K., Alpers, G. W., … Eley, T. C. (2019). A genome-wide association meta-analysis of prognostic outcomes following cognitive behavioural therapy in individuals with anxiety and depressive disorders. *Translational Psychiatry*, *9*(1), 150.

Rice, F., Riglin, L., Thapar, A. K., Heron, J., Anney, R., O'Donovan, M. C., & Thapar, A. (2018). Characterizing Developmental Trajectories and the Role of Neuropsychiatric Genetic Risk Variants in Early-Onset Depression. *JAMA Psychiatry* . https://doi.org/10.1001/jamapsychiatry.2018.3338

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.

Sacchetti, E., Frank, E., Siracusano, A., Racagni, G., Vita, A., & Turrina, C. (2015). Functional impairment in patients with major depression in clinical remission: results from the VIVAL-D-Rem, a nationwide, naturalistic, cross-sectional survey. *International Clinical Psychopharmacology*, *30*(3), 129–141.

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.

Saunders, R., Cape, J., Leibowitz, J., Aguirre, E., Jena, R., Cirkovic, M., … Buckman, J. E. J. (2020). Improvement in IAPT outcomes over time: are they driven by changes in clinical practice? *Cognitive Behaviour Therapy*, *13*, e16.

Schwabe, I., Milaneschi, Y., Gerring, Z., Sullivan, P. F., Schulte, E., Suppli, N. P., … Middeldorp, C. M. (2019). Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies. *Psychological Medicine*, *49*(16), 2646–2656.

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, *177*(1), 26–31.

Smagula, S. F., Butters, M. A., Anderson, S. J., Lenze, E. J., Dew, M. A., Mulsant, B. H., … Reynolds, C. F. (2015). Antidepressant Response Trajectories and Associated Clinical Prognostic Factors Among Older Adults. *JAMA Psychiatry*, *72*(10), 1021–1028.

Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børglum, A. D., Breen, G., … Psychiatric Genomics Consortium. (2018). Psychiatric Genomics: An Update and an Agenda. *The American Journal of Psychiatry*, *175*(1), 15–27.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews. Genetics*, *20*(8), 467–484.

Thorp, J. G., Marees, A. T., Ong, J.-S., An, J., MacGregor, S., & Derks, E. M. (2020). Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9. *Psychological Medicine*, *50*(14), 2385–2396.

Verboom, C. E., Sentse, M., Sijtsema, J. J., Nolen, N. A., Ormel, J., & Penninx, B. W. J. H. (2011). Explaining heterogeneity in disability with major depressive disorder: Effects of personal and environmental characteristics. *Journal of Affective Disorders*, *132*, 71–81.

Wallert, J., Boberg, J., Kaldo, V., Mataix-Cols, D., Flygare, O., Crowley, J. J., … Rück, C. (2022). Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Translational Psychiatry, 12*(1), 1-10. https://doi.org/10.1038/s41398-022-02133-3

Wittchen, H.-U., Kessler, R. C., Beesdo, K., Krause, P., Höfler, M., & Hoyer, J. (2002). Generalized anxiety and depression in primary care: prevalence, recognition, and management. *The Journal of Clinical Psychiatry*, *63 Suppl 8*, 24–34.

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell*, *173*(7), 1573–1580.

Yang, J., Wray, N. R., & Visscher, P. M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology*, *34*(3), 254–257.

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., … Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*, *50*(9), 1304–1310.

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., … Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*, *9*(1), 224.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*, *163*(1), 148–150.

Appendix A - Supplementary materials for Chapter 2

Supplementary materials for "Self-Reported Medication Use as an Alternate Phenotyping Method for Anxiety and Depression in the UK Biobank"

1. S.Info 1: Medication Code Classification Mappings
2. S.Info 2: Diagnostic Criteria for Indicators of Anxiety and Depression in the UK Biobank
3. S.Table 1: Crosstabs of Complete Data for Indicators of Anxiety and Depression in the UK Biobank
4. S.Table 2: Crosstabs of Cases for Indicators of Anxiety and Depression in the UK Biobank
5. S.Info 3: Physical Conditions Excluded in the Sensitivity Analysis
6. S.Table 3: Sample Sizes for Medication Phenotypes in Main and Sensitivity Analysis
7. S.Info 4: GWAS Results and Heritability Estimates for Medication Status and Medication Only
8. S.Fig 1: Manhattan and QQ Plot for GWAS of Medication Status
9. S.Fig 2: Manhattan and QQ Plot for GWAS of Medication Only
10. S,Table 4: Descriptives of Phenotypes Derived from the UK Biobank; Medication Status, Medication Only and Lifetime Internalising
11. S.Fig 3: Variance Explained by Polygenic Scores of Anxiety and Depression for Medication Status and Medication Only
12. S.Info 5: Sensitivity Analyses Results
13. S.Fig 4: Manhattan and QQ Plot for Sensitivity GWAS
14. S.Table 5: Genetic Correlations for Sensitivity Analysis
15. S.Table 6: Polygenic Score Results for Sensitivity Analysis

Supplementary Information 1. Medication Code Classification Mappings

Mappings were performed by Christopher Hübel, Héléna A. Gaspar and Katrina Davis (Davis et al., 2019), as detailed below.

Antidepressants		***Anxiolytics***	
Drug ID	**Drug Name**	**Drug ID**	**Drug Name**
1140879616	amitriptyline	1140863144	zopiclone
1140921600	citalopram	1140863152	diazepam
1140879540	fluoxetine	1140863202	temazepam
1140867878	sertraline	1140865016	zolpidem
1140916282	venlafaxine	1140863182	nitrazepam
1140909806	dosulepin	1140863302	lorazepam
1140867888	paroxetine	1140883656	hydroxyzine
1141152732	mirtazapine	1140928004	zimovane
1141180212	escitalopram	1140862810	phenergan
1140879634	trazodone	1140882082	promethazine
1140867876	prozac	1140879730	buspirone
1140882236	seroxat	1140863286	atarax
1141190158	cipralex	1140863442	oxazepam
1141200564	duloxetine	1140863120	loprazolam
1140867726	lofepramine	1140863328	chlordiazepoxide
1140879620	clomipramine	1140863176	lormetazepam
1140867818	nortriptyline	1140863292	ucerax
1140879630	imipramine	1140864916	stilnoct
1140879628	dothiepin	1141157496	diazepam
1141151946	cipramil	1140863454	buspar
1140867948	amitriptyline	1140863308	alprazolam
1140867624	prothiaden	1140863350	librium
1140867756	trimipramine	1140863310	xanax
1140867884	lustral	1140863440	meprate
1141151978	reboxetine	1140863112	dalmane
1141152736	zispin	1140909798	clomethiazole
1141201834	cymbalta	1140863378	meprobamate
1140867690	anafranil	1140863028	welldorm
1140867640	doxepin	1140867938	amitriptyline+chlor diazepoxide
1140867920	moclobemide	1140863110	flurazepam

1140867850	phenelzine	1140863036	heminevrin
1140879544	fluvoxamine	1140863372	medazepam
1141200570	yentreve	1140867136	neulactil
1140867934	triptafen	1140882312	sinequan
1140867758	surmontil	1140855870	almazine
1140867914	tranylcypromine	1140855832	atensine
1140867820	allegron	1140875434	carisoma
1141151982	edronax	1140863410	chloractil
1140882244	molipaxin	1140863016	chloral
1140879556	mianserin	1140855824	dichloralphenazone
1140867852	nardil	1140855890	dormonoct
1140867860	faverin	1140856040	methyprylone
1140917460	nefazodone	1140863194	mogadon
1140867938	amitriptyline+chlordiazepoxide	1140863106	rohypnol
1140867856	isocarboxazid	1140867668	tryptizol
1140867922	manerix		
1140910820	maoi		
1140882312	sinequan		
1140867944	tranylcypromine+trifluoperazine		
1140867784	ludiomil		
1140867812	norval		
1140867668	tryptizol		
1140867940	fluphenazine hydrochloride+nortriptyline		

Antipsychotics

Drug ID	**Drug Name**
1140868170	prochlorperazine
1140928916	olanzapine
1141152848	quetiapine
1140867444	risperidone
1140879658	chlorpromazine
1140868120	trifluoperazine
1141153490	amisulpride
1140867304	sulpiride

Mood Stabilisers

Drug ID	**Drug Name**
1140867490	lithium product
1140867494	camcolit 250 tablet
1140867498	liskonum 450mg m/r tablet
1140867504	priadel 200mg m/r tablet
1140872290	lamotrigine
1140872198	sodium valproate
1140872214	valproic acid
1141200004	pregabalin

1141152860	seroquel	1140863268	clobazam
1140867168	haloperidol		
1141195974	aripiprazole		
1140867244	stelazine		
1140867152	depixol		
1140909800	flupentixol		
1140867420	clozapine		
1140879746	promazine		
1141177762	risperdal		
1140867456	modecate		
1140867952	fluanxol		
1140867150	flupenthixol		
1141167976	zyprexa		
1140882100	zuclopenthixol		
1140867342	clopixol		
1140863416	largactil		
1141202024	abilify		
1140882098	fluphenazine		
1140867184	haldol		
1140867092	serenace		
1140882320	clozaril		
1140910358	cpz		
1140867208	perphenazine		
1140909802	levomepromazine		
1140867134	pericyazine		
1140867306	dolmatil		
1140867210	fentazin		
1140867398	fluphenazine		
1140867078	benperidol		
1140867218	pimozide		
1141201792	zaponex		
1141200458	denzapine		
1140867136	neulactil		
1140879750	thioridazine		
1140867180	dozic		
1140867546	fluspirilene		
1140928260	panadeine		
1140927956	sertindole		

Supplementary Information 2. Diagnostic Criteria for Indicators of Anxiety and Depression in the UK Biobank

Case and control status were based solely on the information in the item in question. For example, an individual without any ICD-10 mental health codes would be considered a control on that indicator, regardless of an interview item indicating depression. UK Biobank Field Codes are indicated by [f.numbers]. The Mental Health Questionnaire (MHQ) components of current anxiety and depression symptoms and self-reported diagnoses are not included, as the analytical sample used for the medication phenotypes excluded MHQ-responders. The MHQ lifetime symptom questionnaires are described here as these were used to create the comparison phenotype 'Lifetime Internalising'.

1) Lifetime Internalising (Component of MHQ)

Participants were considered as a Lifetime Internalising case if they met criteria for the Anxiety **and/or** the Depression module of the Composite International Diagnostic Interview Short Form (CIDI-SF); symptom-based questionnaires which are detailed below. The first two items for each module were 'screening items'. If individuals had responded to the MHQ but neither of the screening items were answered for a module, the individual was recorded as missing for that module. Controls responded to screening item(s) but did not fulfil case criteria for either module.

Anxiety

I. Ever had a time in their lives where they have felt worried, tense, or anxious most of the time for at least a month [f.20421; screening item]
II. Ever worried more than most people would in a similar situation [f.20425; screening item] **OR**
Ever had stronger worry than most people during their worst period of anxiety [f.20542]
III. Longest period spent worried or anxious was at least 6 months [f.20420]
IV. Worried most days during their worst period of anxiety [f.20538]
V. Had multiple worries during their worst period of anxiety [f.20540] **OR**
Worried about more than one thing during their worst period of anxiety [f.20543]
VI. Found it difficult to stop worrying during their worst period of anxiety [f.20541] **OR**
Often unable to stop worrying during their worst period of anxiety [f.20539] **OR**
Often had difficulty controlling their worry during their worst period of anxiety [f.20537]
VII. Somewhat or a lot of impact on their normal roles during their worst period of anxiety [f.20418]
VIII. Experienced a total of 3 or more of the following somatic symptoms during their worst period of anxiety:
Restlessness [f.20426] **OR** Keyed up or on edge [f.20423]

Easily tired [f.20429]
Trouble falling or staying asleep [f.20427]
Increased irritability [f.20422]
Tense, sore or aching muscles [f.20417]
Difficulty concentrating [f.20419]

Depression

I. Prolonged loss of interest in normal activities [f.20441] **OR**
Prolonged feelings of sadness or depression [f.20446]
II. Most or all of the day being affected during their worst episode of depression [f.20436]
III. Depressed days occurring almost every day or every day during their worst episode of depression [f.20439]
IV. Somewhat or a lot of impact on their normal roles during their worst period of depression [f.20440]
V. Experienced a total of 5 or more of the following symptoms:
Ever experienced prolonged feelings of sadness or depression [f.20446]
Ever experienced prolonged loss of interest in normal activities [f.20441]
Feelings of tiredness during their worst period of depression [f.20449]
Weight change during their worst period of depression [f.20536]
Sleep change during their worst period of depression [f.20532]
Difficulty concentrating during their worst period of depression [f.20435]
Feelings of worthlessness during their worst period of depression [f.20450]
Thoughts of death during their worst period of depression [f.20437]

2) Probable Depression

These items were presented to a subset of individuals at the baseline assessment session (91,847 in the present analytical sample). Cases met criteria for single episode depression, probable recurrent moderate depression or probable recurrent severe depression in the derived 'Bipolar and major depression status' [f.20126] item, created as per (Smith et al., 2013) and detailed below.

Depression

I. Single episode:
Ever felt depressed for a whole week **OR** Ever disinterested or unenthusiastic for a whole week
One depressive episode lasting at least one week in duration
Longest episode of depression at least two weeks in duration
Ever seen a GP for nerves, anxiety, tension or depression **OR** Ever seen a Psychiatrist for nerves, anxiety, tension or depression
II. Probable recurrent moderate:

Ever felt depressed for a whole week **OR** Ever disinterested or unenthusiastic for a whole week Two or more depressive episodes lasting at least one week in duration
Longest episode of depression at least two weeks in duration
Ever seen a GP for nerves, anxiety, tension or depression

III. Probable recurrent severe:
Ever felt depressed for a whole week **OR** Ever disinterested or unenthusiastic for a whole week Two or more depressive episodes lasting at least one week in duration
Longest episode of depression at least two weeks in duration
Ever seen a Psychiatrist for nerves, anxiety, tension or depression

Controls did not meet criteria on this item for depression or bipolar depression and were indistinguishable from those with missing data.

3) Help-Seeking

Cases: Reported yes to "Have you ever seen a General Practitioner (GP) for nerves, anxiety, tension or depression?" [f.2090] **OR**
Reported yes to "Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?" [f.2100]

Controls: Reported no to both of the above items

Missing: Responded "Do not know" or "Prefer not to answer" to both items

4) Hospital Records (ICD-10 Primary and Secondary)

Cases had any of the following codes:
***Anxiety* -** F400, F401, F402, F408, F409, F410, F411, F412, F413, F418, F419
***Depression* -** F320, F321, F322, F323, F328, F329, F330, F331, F332, F333, F334, F338, F339, F340, F341

Exclusions for cases and controls were:
Bipolar - F300, F308, F311, F314, F317, F380, F301, F309, F312, F315, F318, F348, F381, F302, F310, F313, F316, F319, F349, F388; Schizophrenia - F200, F201, F202, F203, F204, F205, F206, F208, F209, F220, F228, F229, F230, F231, F232, F233, F238, F239, F250, F251, F252, F258, F259; Autism Spectrum Disorder - F840, F841, F842, F843, F844, F845, F846, F847, F848, F849; Eating Disorder - F500, F501, F502, F503, F504, F505, F508, F509; ADHD: F900, F901, F908, F909

Controls had no record of any mental health “F” codes and were indistinguishable from those with missing data. Participants from Scotland did not have ICD codes for mental health.

5) Interview non-cancer illness code, self-reported

Cases had any of the following codes:

Anxiety **-** 1287

Depression **-** 1286

Exclusions for cases and controls were:

Mania/bipolar disorder/manic depression: 1291; Schizophrenia: 1289; Anorexia/bulimia/other eating disorder: 1470

Controls had none of the above codes, or the following codes, and were indistinguishable from those with missing data: Obsessive compulsive disorder: 1615; PTSD: 1469; Postnatal depression: 1531; Alcohol dependency: 1408; Opioid dependency: 1409; Other substance abuse/dependency: 1410; Stress: 1614; Nervous breakdown: 1288; Deliberate self-harm/suicide attempt: 1290

Supplementary Table 1. Cross-tabulation of individuals with complete (non-missing) data on each pair of anxiety and depression diagnostic indicators in UK Biobank (European ancestry, excluding Mental Health Questionnaire responders).

Diagnostic indicators	Medication Status	Medication Only	Hospital Code (ICD-10)	Interview Self-Report	Probable Depression	Help Seeking
Medication Status	**N = 191,177**	109,672 57.3%	181,558 95.0%	189,831 99.3%	41,635 21.8%	189,494 99.1%
Medication Only	109,672 100.0%	**N = 109,672**	109,672 100.0%	109,672 100.0%	22,744 20.7%	108,252 98.7%
Hospital Code	181,558 49.0%	109,672 29.6%	**N = 370,358**	368,064 99.4%	88,523 23.9%	367,758 99.3%
Interview Self-Report	189,831 49.7%	109,672 28.7%	368,064 96.3%	**N = 382,179**	90,990 23.8%	379,413 99.3%
Probable Depression	41,635 45.3%	22,744 24.8%	88,523 96.4%	90,990 99.1%	**N = 91,847**	91,532 99.7%
Help Seeking	189,494 49.5%	108,252 28.3%	367,758 96.1%	379,413 99.1%	91,532 23.9%	**N = 382,858**

Percentages are the proportion of the indicator named on the row who also held non-missing data for the indicator named on the column, thus both triangles of the cross-tabulation are displayed to include the proportion of each of the pair. Descriptions of these indicators are available in Supplementary Information 2. Medication Only cases and controls either did not meet case criteria or had incomplete data for all other diagnostic indicators. Hospital codes and interview self-report did not contain values to indicate missingness, therefore absence of a diagnostic code was used to define controls. As such, it appears as though there is no missing data for these variables.

Supplementary Table 2. Cross-tabulation of anxiety and depression cases, as classified by each pair of diagnostic indicators in UK Biobank (European ancestry, excluding Mental Health Questionnaire responders).

Diagnostic indicators	Medication Status	Medication Only	Hospital Code (ICD 10)	Interview Self-Report	Probable Depression	Help Seeking
Medication Status	**N = 22,218**	2,643 11.9%	4,894 22.0%	12,457 56.1%	3,179 14.3%	19,000 85.5%
Medication Only	2,643 100.0%	**N = 2,643**	0 0.0%	0 0.0%	0 0.0%	0 0.0%
Hospital Code	4,894 34.2%	0 0.0%	**N = 14,320**	5,932 41.4%	2,277 15.9%	11,954 83.5%
Interview Self-Report	12,457 45.6%	0 0.0%	5,932 21.7%	**N = 27,335**	4,948 18.1%	26,756 97.9%
Probable Depression	3,179 12.9%	0 0.0%	2,277 9.2%	4,948 20.1%	**N = 24,639**	24,639 100.0%
Help Seeking	19,000 14.0%	0 0.0%	11,954 8.8%	26,756 19.7%	24,639 18.1%	**N = 135,997**

Percentages are the proportion of individuals identified as cases by the indicator named on the row who were also cases as defined by the indicator named on the column. Thus, both triangles of the cross-tabulation are displayed to include the proportion of each of a pair. Descriptions of these indicators are available in Supplementary Information 2.

Supplementary Information 3. Physical conditions excluded in the sensitivity analyses

In the UK, antidepressants are prescribed for chronic pain (National Institute for Health and Care Excellence, 2017) and some anxiolytics for pain and epilepsy (Joint Formulary Committee, n.d.). Individuals with any of the following codes were excluded from the sensitivity analysis:

ICD-10 (Primary and Secondary) Codes

Epilepsy: G400, G401, G402, G403, G404, G405, G406, G407, G408, G409

Pain: R520, R521, R522, R529

Fibromyalgia: M7970, M7971, M7972, M7973, M7974, M7975, M7976, M7977, M7978, M7979

Interview Non-Cancer Illness Codes

Epilepsy: 1264

Participants with insomnia were not excluded due to sleep disturbance being a common symptom of depression.

Supplementary Table 3. Sample sizes for Medication Status and Medication Only in the UK Biobank (European ancestry, excluding Mental Health Questionnaire responders), in the main analysis and the sensitivity analysis excluding individuals with indications of physical conditions often treated with antidepressants or anxiolytics.

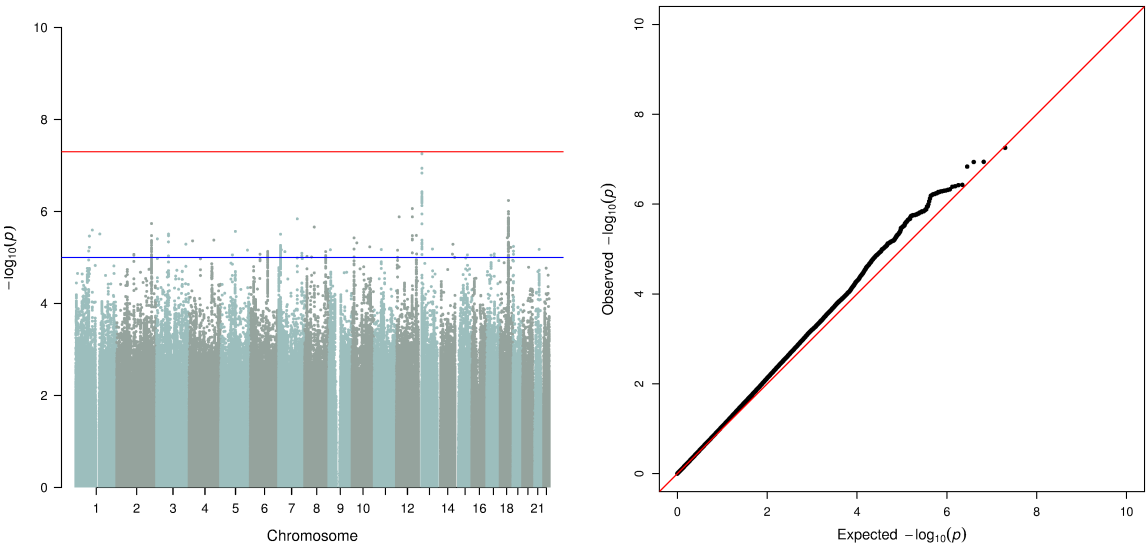
Phenotype	Cases	Controls
Medication Status	22,218	168,959
Medication Only	2,643	107,029
Medication Status sensitivity analysis	21,340	165,846
Medication Only sensitivity analysis	2,550	105,484

Medication Status was defined and screened solely using information on self-reported current medication use, with cases reporting antidepressant or anxiolytic use and controls reporting non-psychotropic medications. Medication Only was a subset of Medication Status, created by incorporating information from other indicators of mental health from UK Biobank (e.g., hospital record codes). In addition to the Medication Status definition, Medication Only cases and controls did not meet case criteria for any other mental health indicator. The conditions excluded in the sensitivity analysis included pain and epilepsy; details are available in Supplementary Information 3.

Supplementary Information 4. Genome-Wide Association Study Results and Heritability Estimates of Medication Status and Medication Only

Prior to the genome-wide association study (GWAS), phenotypes were adjusted for the effects of age, sex, genotyping batch, assessment centre and the first six genetic principal components, using logistic regression in R version 3.6.0 (R Core Team, 2019). For each phenotype, residuals from the logistic regression were tested for associations with genome-wide SNPs in 9,912,453 linear regressions using BGENIE software (version 1.2; Bycroft et al., 2018).

The GWAS of Medication Status did not identify any loci that were statistically significant at the genome-wide level. See Manhattan and quantile-quantile (QQ) plots in Supplementary Figure 1.



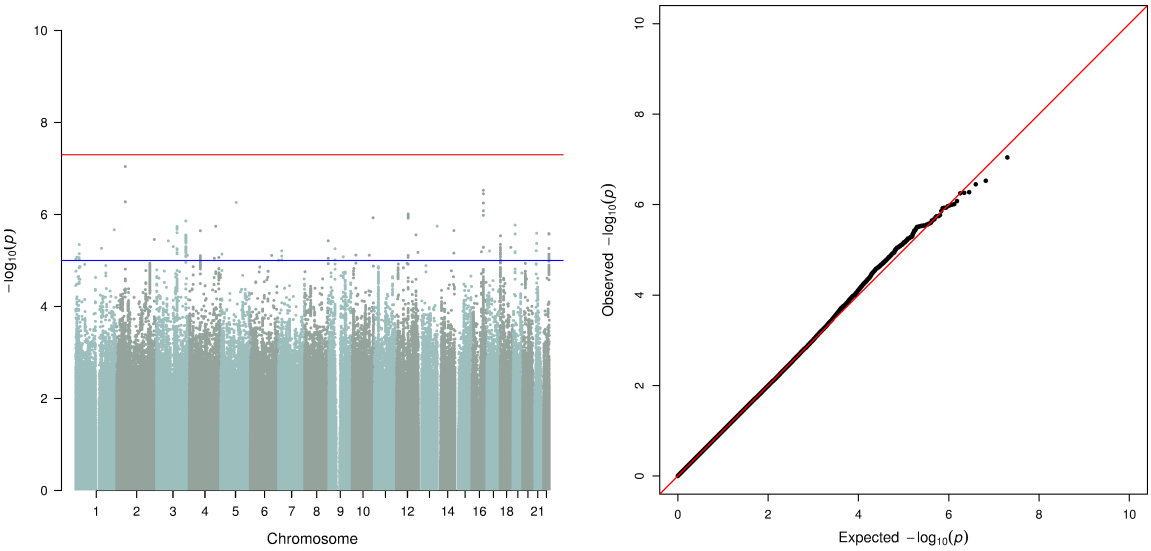
Supplementary Figure 1. Manhattan and QQ plots of the genome-wide association analysis of Medication Status in the UK Biobank (European ancestry, excluding Mental Health Questionnaire responders)

The Medication Status phenotype consists of 22,218 cases who self-reported current antidepressant or anxiolytic medication use, and 168,959 controls who did not report any psychotropic medication use. The red line on the Manhattan plot indicates the genome-wide significance threshold, corrected for multiple testing ($p < 5x10^{-8}$) and points above this line are variants significantly associated with case status. The blue line indicates the threshold for suggestive significance ($p < 1x10^{-5}$). The QQ plot presents the deviation from the null hypothesis (red dotted line) of no significantly associated variants.

The linkage disequilibrium score regression (LDSC; Bulik-Sullivan, Finucane, et al., 2015; Bulik-Sullivan, Loh, et al., 2015) estimate of SNP-based heritability of Medication Status was 0.074 (SE = 0.004) on the liability scale, when assuming that the prevalence in this UK Biobank sample (12%) was representative of the population prevalence. When population prevalence was assumed to be -/+10% sample prevalence, SNP heritability was 0.042 (SE =

0.001) and 0.090 (SE = 0.006) respectively. The LDSC intercept value of 0.991 (SE = 0.007) suggested that there was no genomic inflation due to population stratification.

The GWAS of Medication Only did not identify any loci that were statistically significant at the genome-wide level. Results are displayed via Manhattan and QQ plots in Supplementary Figure 2.



Supplementary Figure 2. Manhattan and QQ plots of the genome-wide association study of Medication Only in the UK Biobank (European ancestry, excluding Mental Health Questionnaire responders)

The Medication Only phenotype consists of 2,643 individuals who self-reported current antidepressant or anxiolytic medication use and 107,029 controls who did not report any psychotropic medication use, and neither cases nor controls met case criteria for other mental health indicators in UK Biobank. The red line on the Manhattan plot indicates the genome-wide significance threshold, corrected for multiple testing ($p < 5x10^{-8}$) and points above this line are variants significantly associated with case status. The blue line indicates the threshold for suggestive significance ($p < 1x10^{-5}$). The QQ plot presents the deviation from the null hypothesis (red dotted line) of no significantly associated variants.

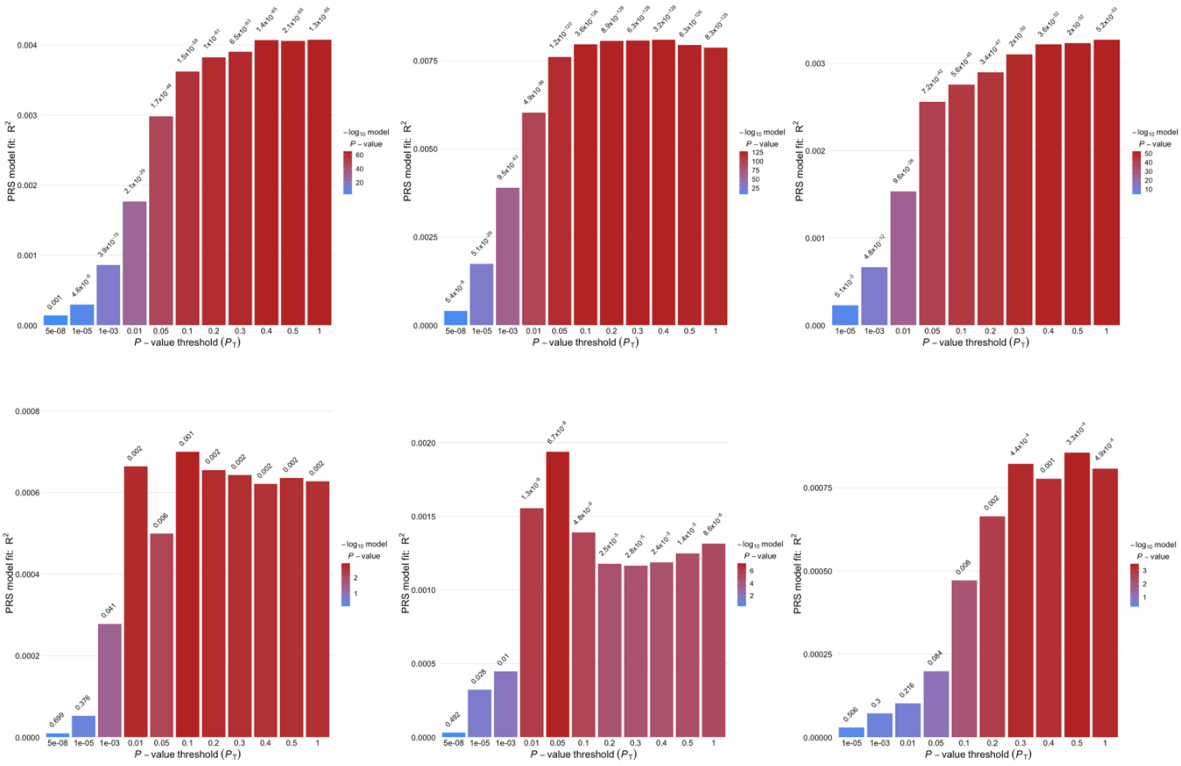
The LDSC estimate of SNP-based heritability of Medication Only was 0.053 (SE = 0.010) on the liability scale, when assuming the sample prevalence (2%) represented population prevalence. If the population prevalence was -1%/+10% of the sample prevalence, SNP heritability was 0.047 (SE = 0.008) and 0.087 (SE = 0.029) respectively. The LDSC intercept value of 0.994 (SE = 0.006) suggested that there was no genomic inflation due to population stratification.

Supplementary Table 4. Descriptives of self-reported current antidepressant or anxiolytic medication use and Lifetime Internalising phenotypes derived from the UK Biobank (European ancestry).

Descriptive cases controls	**Medication Status Excluding MHQ Responders**	**Medication Only Excluding MHQ Responders**	**Lifetime Internalising MHQ Responders**	**Medication Status and Lifetime Internalising**	**Medication Only and Lifetime Internalising**
n	22,218 168,959	2,643 107,029	32,160 91,732	-	-
Age	57.22 (7.83) 58.08 (7.97)	59.90 (7.25) 58.42 (7.99)	54.27 (7.54) 56.61 (7.67)	*d* = 0.39* *d* = 0.19*	*d* = 0.75* *d* = 0.23*
Sex (% Female)	67.70 53.71	66.89 49.44	68.30 51.81	*V* = 0.01 *V* = 0.02*	*V* = 0.01 *V* = 0.02*
Neighbourhood Deprivation	-0.67 (3.36) [0.14] -1.34 (3.07) [0.11]	-1.08 (3.24) [0.11] -1.56 (2.94) [0.10]	-1.44 (2.97) [0.16] -1.88 (2.73) [0.11]	*d* = 0.24* *d* = 0.18*	*d* = 0.12* *d* = 0.11*
Education (% University-Level)	21.33 [1.25] 25.04 [1.31]	18.99 [2.12] 25.26 [1.38]	46.78 [0.31] 46.45 [0.30]	*V* = 0.36* *V* = 0.27*	*V* = 0.28* *V* = 0.28*
Ever Smoked (% Yes)	54.44 [0.66] 48.31 [0.53]	47.18 [1.78] 45.21 [0.55]	46.91 [0.18] 40.67 [0.20]	*V* = 0.08* *V* = 0.08*	*V* = 0.01 *V* = 0.05*

Medication Status was defined solely using self-reported current use of antidepressant or anxiolytic medication in UK Biobank participants who did not respond to the Mental Health Questionnaire (MHQ). Medication Only were a subset who additionally did not meet case criteria for any other indicator of mental health disorder in UK Biobank. Lifetime Internalising was derived from anxiety and depression lifetime symptom-based questionnaires from the MHQ. Neighbourhood deprivation was measured by the Townsend Deprivation Index, where lower scores indicate less deprivation. Within each variable, the top row indicates the cases, and the bottom is the controls. For continuous variables, values are means with standard deviation of the mean in brackets. For categorical variables, values are proportions in percentages. If data are missing, the percentage is presented in square brackets. d = Cohen's d measure of effect size of the difference from an independent t-test of continuous variables. V = Cramer's V measure of effect size of the difference from a Pearson's chi-square test for categorical variables. * Indicates significance at $p < 0.001$.

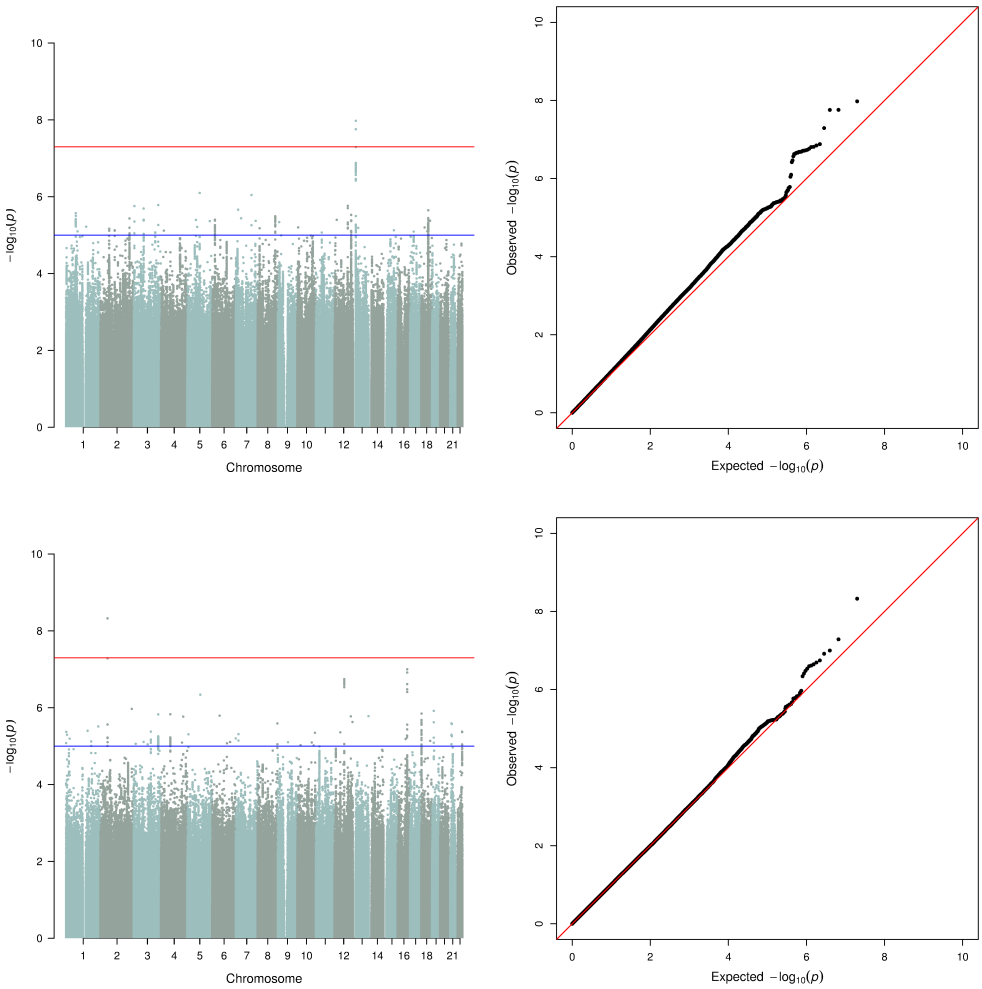
Supplementary Figure 3. Plots of polygenic scores created from genome-wide association studies of UKB-Anxiety (left column; Purves et al., 2020), PGC-Depression (middle column; Wray et al., 2018) and Lifetime Internalising (right column; performed in UK Biobank Mental Health Questionnaire responders), in Medication Status (top row) and Medication Only (bottom row) in the UK Biobank (European ancestry, excluding Mental Health Questionnaire responders)



The Medication Status phenotype consisted of 21,340 individuals who self-reported current antidepressant or anxiolytic medication use and 165,846 controls who did not report any psychotropic medication use. The Medication Only phenotype consisted of 2,550 individuals who self-reported current antidepressant or anxiolytic medication use but did not meet criteria for other mental health indicators in UK Biobank, and 105,484 controls who did not report psychotropic medication use or meet case criteria for any other indicators of mental health. The x-axis for each plot indicates the *p*-value threshold of association between genetic variants and the phenotype in the discovery sample, such that an increasing number of variants is included at higher thresholds. The y-axis is Nagelkerke's R2 (proportion of variance explained) converted to the liability scale, assuming that the phenotype sample prevalence is equal to the population prevalence.

Supplementary Information 5. Sensitivity Analyses Results

The sensitivity analysis GWAS of Medication Status and Medication Only, excluding individuals with evidence of specific physical disorders (pain and epilepsy), did not identify any loci that were statistically significant at the genome-wide level. Results are displayed via Manhattan and quantile-quantile (QQ) plots in Supplementary Figure 4.



Supplementary Figure 4. Manhattan and QQ plots of the genome-wide association test for the sensitivity analysis of Medication Status (top row) and Medication Only (bottom row) in the UK Biobank (European ancestry, excluding Mental Health Questionnaire responders)

The Medication Status phenotype consists of 21,340 individuals who self-reported current antidepressant or anxiolytic medication use and 165,846 controls who did not report any psychotropic medication use. The Medication Only phenotype consists of 2,550 individuals who self-reported current antidepressant or anxiolytic medication use but did not meet case criteria for other mental health indicators in UK Biobank, and 105,484 controls who did not report psychotropic medication use or meet case criteria for any other indicators of mental health. The red line on the Manhattan plot indicates the genome-wide significance threshold, corrected for multiple testing ($p < 5x10^{-8}$) and points above this line are variants significantly associated with case status. The blue line indicates the threshold for suggestive significance ($p < 1x10^{-5}$). The

QQ plot presents the deviation from the null hypothesis (red dotted line) of no significantly associated variants.

In the sensitivity analysis of Medication Status, the LDSC estimate of SNP-based heritability was 0.076 (SE = 0.005) on the liability scale, when assuming that the prevalence in this UK Biobank sample (11%) was representative of the population prevalence. When population prevalence was assumed to be -/+10% sample prevalence, SNP heritability was 0.042 (SE = 0.001) and 0.090 (SE = 0.007) respectively. The LDSC intercept value of 0.99 (SE = 0.007) suggested that there was no genomic inflation due to population stratification. For Medication Only, SNP-based heritability was estimated as 0.060 (SE = 0.010), when assuming that the prevalence in this UK Biobank sample (2%) was representative of the population prevalence. When population prevalence was assumed to be -1%/+10% sample prevalence, SNP heritability was 0.052 (SE = 0.007) and 0.099 (SE = 0.029) respectively. The LDSC intercept value of 0.993 (SE = 0.006) suggested that there was no genomic inflation due to population stratification.

The results of the genetic correlations and polygenic scoring in the sensitivity analysis were similar to those from the main analysis; see Supplementary Tables 5 and 6.

Supplementary Table 5. Genetic correlations for the sensitivity analyses of Medication Status and Medication Only from the UK Biobank (European ancestry, excluding Mental Health Questionnaire (MHQ) responders) and genome-wide association studies (GWAS) selected for comparison

Comparison GWAS	Medication Status excluding selected physical conditions	Medication Only excluding selected physical conditions
UKB-Anxiety (Purves et al. 2020, UK Biobank MHQ responders)	0.66 (0.06)	0.13 (0.12)
PGC-Depression (Wray et al. 2018, excluding UK Biobank)	0.72 (0.05)	0.49 (0.14)
Lifetime Internalising (Lifetime symptom-based questionnaires in UK Biobank MHQ responders)	0.58 (0.07)	0.17 (0.16)

The Medication Status phenotype was defined using self-reported current use of antidepressant or anxiolytic medication in individuals who did not complete the MHQ (N = 187,186, cases = 21,340). Medication Only were a subset who additionally did not meet case criteria for any other indicators of anxiety or depression in the sample (N = 108,034, cases = 2,550). Individuals were excluded if they had evidence of physical conditions associated with prescription of these medications (pain and epilepsy). Standard errors are shown in brackets.

Supplementary Table 6. Proportion of variance (Nagelkerke's R^2) in self-reported current antidepressant or anxiolytic medication use from the UK Biobank (European ancestry, excluding Mental Health Questionnaire (MHQ) responders) explained by polygenic scores of anxiety and depression

Discovery Sample	Medication Status excluding selected physical conditions	Medication Only excluding selected physical conditions
UKB-Anxiety (Purves et al. 2020, UK Biobank MHQ responders)	0.40% (p = 9.99 x 10^{-5}, P_T = 1; 200,935 SNPs)	0.08% (p = 4.6 x 10^{-3}, P_T = 0.1; 35,404 SNPs)
PGC-Depression (Wray et al., 2018, excluding UK Biobank	0.79% (p = 9.99 x 10^{-5}, P_T = 0.2; 37,430 SNPs)	0.13% (p = 9.99 x 10^{-5}, P_T = 0.05; 14,288 SNPs)
Lifetime Internalising (Lifetime symptom-based questionnaires in UK Biobank MHQ responders)	0.32% (p = 9.99 x 10^{-5}, P_T = 1; 201,839 SNPs)	0.07% (p = 3.8 x 10^{-3}, P_T = 0.3; 87,816)

Proportion of variance is presented on the liability scale (Lee, Goddard, Wray, & Visscher, 2012), assuming population prevalence is equal to sample prevalence. The Medication Status phenotype was defined using self-reported current use of antidepressant or anxiolytic medication in individuals who did not complete the MHQ (to provide independent discovery and target samples) (N = 187,186, cases = 21,340). Medication Only additionally did not meet case criteria for any other indicators of mental health condition in UK Biobank (N = 108,034, cases = 2,550). Individuals were excluded if they had evidence of physical conditions associated with prescription of these medications (pain and epilepsy).

References

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P., … Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47(11), 1236–1241.

Bulik-Sullivan, B., Loh, P., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3), 291–295.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209.

Davis, K. A. S., Cullen, B., Adams, M., Brailean, A., Breen, G., Coleman, J. R. I., … Hotopf, M. (2019). Indicators of mental disorders in UK Biobank-A comparison of approaches. *International Journal of Methods in Psychiatric Research*, e1796.

Joint Formulary Committee. *British National Formulary (online)*. London: BMJ Group and Pharmaceutical Press. Retrieved 23 January 2020 from https://www.medicinescomplete.com

Lee, S. H., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2012). A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*, 36(3), 214–224.

National Institute for Health and Care Excellence. (2017). *Medicines Optimisation in Chronic Pain (Key therapeutic topic)*. Retrieved from https://www.nice.org.uk/advice/ktt21/resources/medicines-optimisation-in-chronic-pain-pdf-58758008162245

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry*, 25(12), 3292-3303.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., … Pell, J. P. (2013). Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PloS One*, 8(11), e75362.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668–681.

Appendix B - Supplementary materials for Chapter 3

Supplementary materials for "The genetic overlap between depression and anxiety symptom severity and functional impairment"

Supplementary Table 1. Description of existing phenotypes for genetic comparison

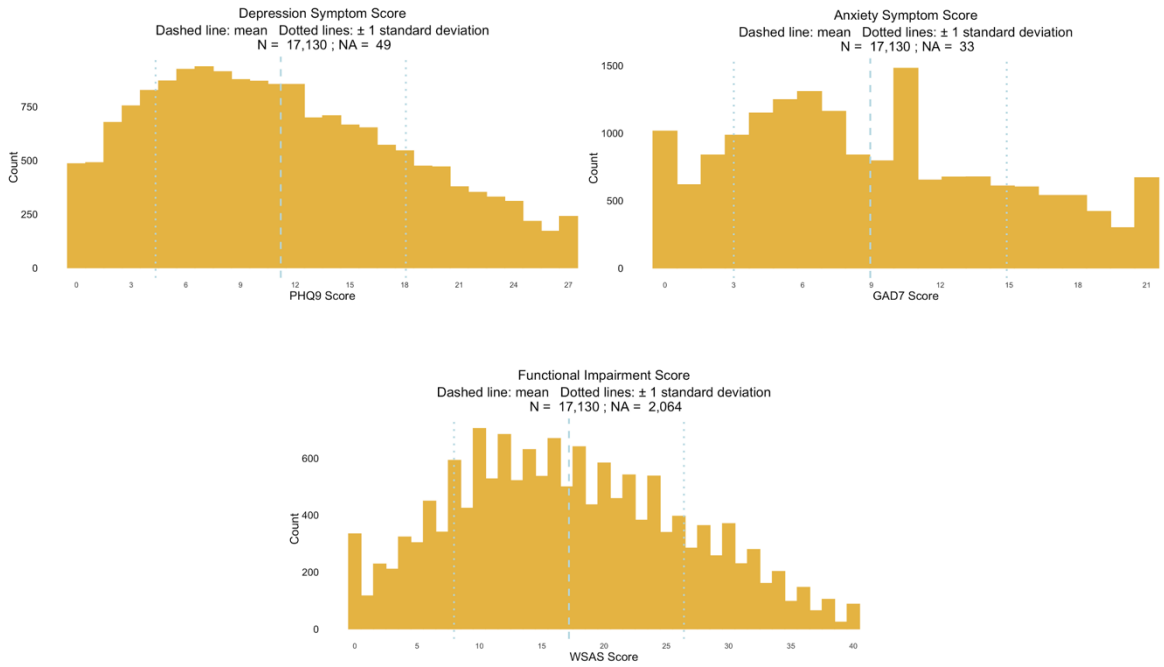
Phenotype	Author & PMID	N (n cases)	SNP h^2	SNP h^2 SE	z-score
Depression (Mega-analysis, excluding 23andme & UK Biobank: Case-control status derived from structured diagnostic interviews & inpatient medical records)	Wray et al. (2018) 29700475	143,265 (45,591)	0.069	0.005	14.766
Anxiety (UK Biobank: Case-control status derived from algorithm of CIDI-SF questionnaire & self-report diagnosis)	Purves et al. (2020) 31748690	83,566 (25,453)	0.100	0.007	14.286
Schizophrenia (Meta-analysis: Case-control status derived from clinical diagnosis)	Pardiñas et al. (2018) 29483656	105,318 (40,675)	0.411	0.017	24.587
Years of education (Meta-analysis: Derived from survey measures of highest educational qualification)	Lee et al. (2018) 30038396	766,345	0.108	0.003	35.967
Self-rated health (UKBiobank: Score from a single item)	Harris et al. (2017) 27864402	111,483	0.089	0.006	14.883

For all summary statistics the mean chi-square of the test statistic was > 1.0, intercepts were within the 0.9-1.1 range and *z*-scores > 4 (as per recommendations for genetic correlations; (Zheng et al., 2017). Values are taken from the present analysis rather than the original analyses and therefore might slightly differ due to calculations using only those variants that overlap with our phenotypes'.

Supplementary Table 2. Descriptives table of analysis sample from the Genetic Links to Anxiety and Depression (GLAD) Study

Variable	N = 17,130*
Age	39.5 (14.6); 16 - 93
Sex	
Female	13,365 (78.0%)
Male	3,765 (22.0%)
Employment status	
In paid employment or self-employed	10,309 (60.3%)
Full or part-time student	2,061 (12.1%)
Unable to work because of sickness or disability	1,862 (10.9%)
Retired	1,231 (7.2%)
Other (looking after home and/or family, doing unpaid or voluntary work)	871 (5.1%)
Unemployed	612 (3.6%)
None of the above	143 (0.8%)
Unknown	41
PHQ9	11.2 (6.9); 0 - 27
Unknown	49
GAD7	8.9 (5.9); 0 - 21
Unknown	33
WSAS	17.2 (9.2); 0 - 40
Unknown	2,064**
WSAS-4*	13.4 (7.5); 0 - 32
Unknown	49

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment, WSAS-4 = 4-item score of functional impairment. * Values are mean (SD); range *or* n (%). ** Majority unknown responded 'not applicable' to the work item. *** A WSAS 4-item total score without the work item was created; see Supplementary Information 1 for further detail.



Supplementary Figure 1. Histograms showing the distribution of each phenotype.

PHQ9 = depression symptoms (top left), GAD7 = anxiety symptoms (top right), WSAS = functional impairment (bottom)

Supplementary Information 1. Investigating the WSAS

Phenotypic

Of the 17,130 participants in the sample, 15,066 had a 5-item, complete, WSAS score. There were 17,107 participants with a 4-item score. Of the 2,041 missing any one item almost all (2,013) had responded 'not applicable' to the work item besides 2 who did not respond; the remaining were missing a different WSAS item. Several months after GLAD launched, an additional item regarding impairment in ability to study was added to the sign-up questionnaire for individuals who selected 'not applicable' to the work item and endorsed student status. This study impairment item was combined with the work item.

Cronbach's alpha for the 5-item WSAS was 0.85 and 4-item score without the work item was 0.83, indicating that internal reliability is marginally decreased without this item. In participants with all five items complete, the phenotypic correlation between the 4-item score and the work item score was .66 (the correlation between the 5-item and 4-item score is reported with Supplementary Table 2 but as this only includes participants with data for both, it is not informative).

Comparisons between participants with the full 5-item score and with a 4-item score due to missing the work item are shown in Table *I*. The mean 4-item score in individuals missing the work item was significantly lower, as were PHQ9 and GAD7 scores. Participants with

the work item missing were on average older than those with complete 5-item scores, and more often reported being retired or looking after the home or family than being employed.

Table *I*. Comparison of sample with complete data on the Work and Social Adjustment Scale (WSAS) and sample missing the work item of the WSAS

Variable	**Full 5-item WSAS** N = 15,066	**Only 4-item score; work missing** N = 2,015	***p****	**Effect size****
Age	37.7 (13.0)	52.8 (18.6)	< 0.001	-0.944
Gender (Female)	11,793 (78.3%)	1,531 (76.0%)	0.019	0.018
Employment			< 0.001	0.668
In paid employment or self-employed	10,183 (67.7%)	102 (5.1%)		
Full or part-time student	1,813 (12.1%)	245 (12.2%)		
Unable to work because of sickness or disability	1,618 (10.8%)	237 (11.8%)		
Retired	213 (1.4%)	1,013 (50.5%)		
Unemployed	528 (3.5%)	81 (4.0%)		
Looking after home and/or family	322 (2.1%)	230 (11.5%)		
Doing unpaid or voluntary work	253 (1.7%)	62 (3.1%)		
None of the above	106 (0.7%)	36 (1.8%)		
PHQ9	11.4 (6.9)	10.0 (6.9)	< 0.001	0.203
N missing	31	17		
GAD7	9.1 (5.9)	7.6 (6.0)	< 0.001	0.259
N missing	23	7		
WSAS 4-item	13.7 (7.4)	11.3 (7.4)	< 0.001	0.322
CIDI-SF Depression (Case)	12,997 (88.6%)	1,662 (87.5%)	0.14	0.011
CIDI-SF Anxiety (Case)	8,288 (62.3%)	946 (57.3%)	< 0.001	0.032
CIDI-SF Depression/Anxiety (Case)	13,746 (94.5%)	1,756 (93.5%)	0.086	0.013

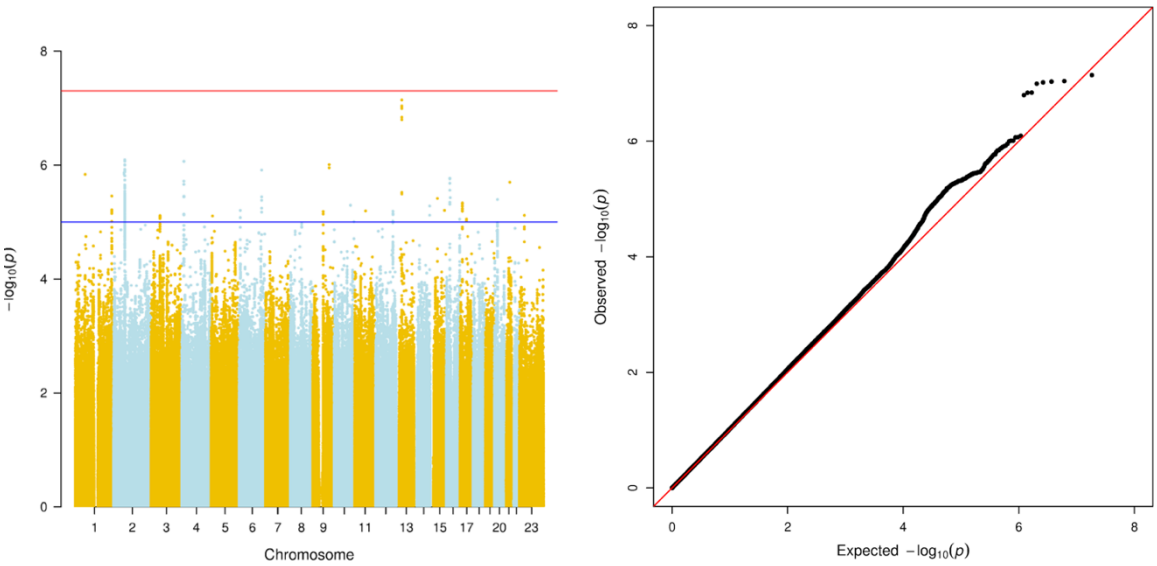
Values are mean (SD) *or* n (%). * Welch's two-sample t-test, or Pearson's Chi-squared test, or Fisher's exact test. ** Cohen's D or Cramer's V.

Genetic

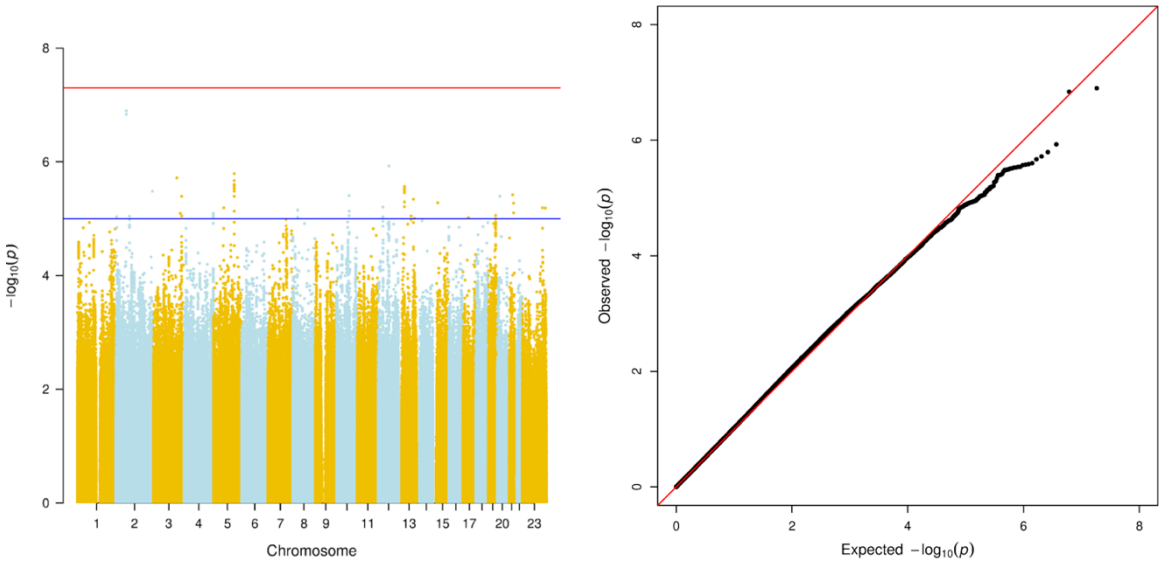
We were unable to perform genetic comparisons with the 4-item WSAS restricted to participants missing the work item (to assess the feasibility of imputation from a genetic point of view) as this group was too small (N = 2,013) such that SNP heritability estimates were not significantly different from zero. The bivariate-GREML estimated genetic correlation between the 4-item WSAS without the work item (N = 17,080; SNP h2 0.11 (SE = 0.03)) and the work-item (N = 15,094; SNP h2 0.07 (SE = 0.04)) was r_g = 0.82 (SE = 0.10). When restricted to participants with data on all five items (N = 15,065), r_g = 0.81 (SE = 0.10). This genetic correlation is unlikely to be significantly different from r_g = 1 and therefore, despite the phenotypic differences described above, we made the decision to use an individual mean imputed 5-item WSAS (N = 17,106) for the main analysis (sample is larger than 4-item WSAS as includes participants missing *any* one item, not just the work item).

Supplementary Figure 2. Quantile-quantile (QQ) and Manhattan plots

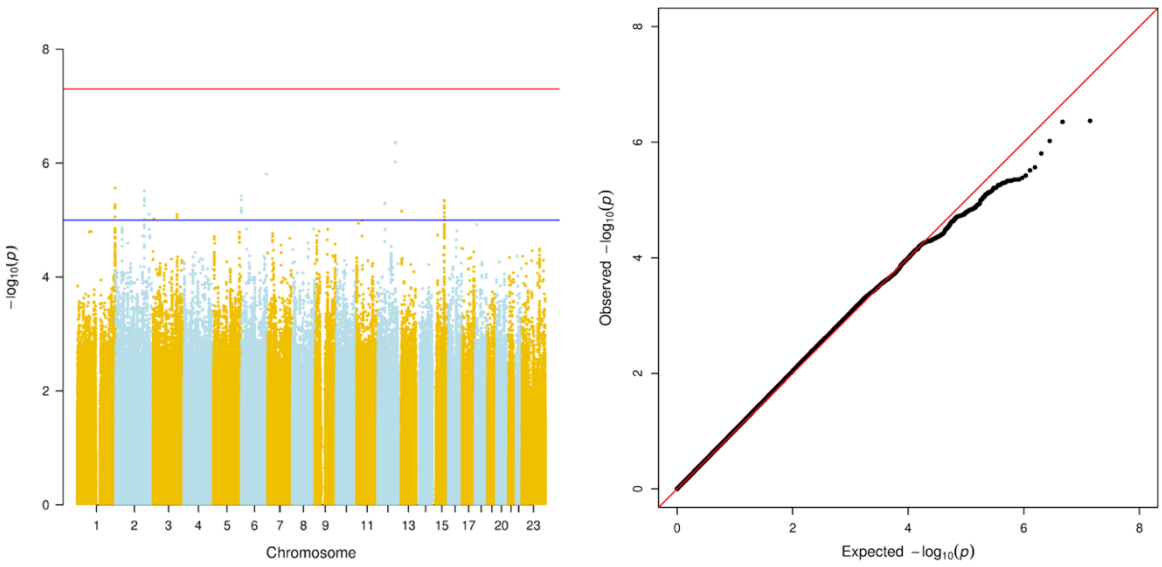
PHQ9



GAD7



WSAS



Supplementary Figure 2. QQ and Manhattan plots from GWAS of each phenotype

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment. QQ plot red line = null hypothesis of no significantly associated variants. A tail of points above this line in the top right would represent deviations from the null indicating true genetic signal. Manhattan plot red line = genome-wide significance threshold, corrected for multiple testing ($p < 5x10^{-8}$); points above this line represent variants significantly associated with the phenotype. Blue line = threshold for suggestive significance ($p < 1x10^{-5}$).

Supplementary Table 3. Phenotypic correlations between symptoms and functional impairment

	PHQ9	GAD7	WSAS
PHQ9	1		
GAD7	0.69 * (0.69 - 0.70)	1	
WSAS	0.63 * (0.62 - 0.64)	0.49 * (0.48 - 0.51)	1

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment. Values in brackets are 95% confidence intervals. * Significance at $p < 0.05$ (all < 0.001). Phenotypic correlation between the 5 and 4-item WSAS = 0.98 but as this is restricted to participants with data on both (i.e. 5-item completers) this is correlating the same participants' 4-item and 5-item scores which are highly similar and increase in parallel. Similarly, complete case 4-item WSAS correlations with depression and anxiety are 0.62 and 0.49 respectively.

Supplementary Table 4. Bivariate-GREML Genetic correlation (r_g) estimates between symptoms and functional impairment

	PHQ9	GAD7	WSAS
PHQ9	1		
GAD7	0.87 * (0.77 - 0.97)	1	
WSAS	0.84 * (0.71 - 0.98)	0.77 * (0.59 - 0.96)	1

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment. Values in brackets are 95% confidence intervals calculated as r_g +/- (1.96*SE). * Significance at $p < 0.05$ (all < 0.001).

Differences from $r_g = 1$ were assessed using the reml-bivar-lrt-rg flag in GCTA to perform a likelihood ratio test and generate a *p*-value. This was also used to generate a *p*-value for the default test of difference from $r_g = 0$, which is not provided by default.

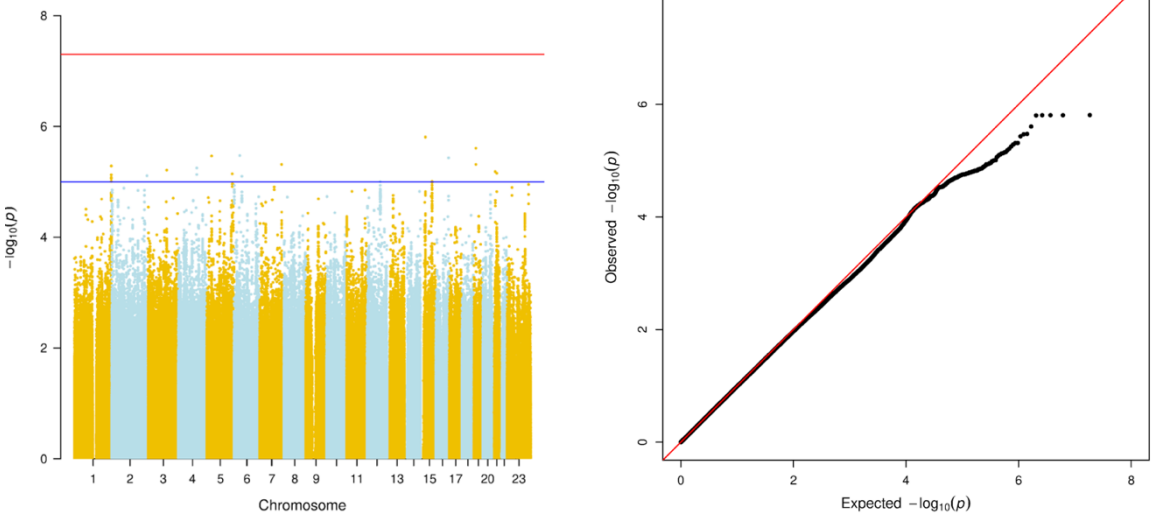
Supplementary Table 5. LDSC estimated genetic correlations (r_g) with existing phenotypes

Phenotype	Existing Phenotype	r_g	95% CI	*z*	*p*
PHQ9	Depression	0.364	0.147 - 0.582	3.279	0.001*
	Anxiety	0.165	-0.046 - 0.377	1.533	0.125
	Schizophrenia	-0.042	-0.190 - 0.105	-0.560	0.575
	Years of education	-0.612	-0.789 - -0.435	-6.779	$1x10^{-11}$*
	Self-rated health	-0.819	-1.074 - -0.563	-6.285	$3x10^{-10}$*
GAD7	Depression	0.387	0.136 - 0.637	3.028	0.002*
	Anxiety	0.202	-0.041 - 0.445	1.632	0.103
	Schizophrenia	0.140	-0.050 - 0.329	1.445	0.148
	Years of education	-0.603	-0.829 - -0.377	-5.225	$2x10^{-7}$*
	Self-rated health	-0.740	-1.048 - -0.432	-4.706	$3x10^{-6}$*
WSAS	Depression	0.419	0.202 - 0.637	3.778	0.0002*
	Anxiety	0.122	-0.075 - 0.319	1.214	0.225
	Schizophrenia	0.040	-0.097 - 0.178	0.573	0.566
	Years of education	-0.338	-0.460 - -0.215	-5.423	$6x10^{-8}$*
	Self-rated health	-0.605	-0.823 - -0.387	-5.437	$5x10^{-8}$*

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = functional impairment. 95% CI = 95% confidence interval. * Significance at $p < 0.05$

Supplementary Information 2. Complete case WSAS (WSAS-cc; N = 15,065) and 4-item WSAS score excluding the work item (WSAS-4, N = 17,080), including LDSC heritability and genetic correlation estimates

WSAS-cc



WSAS-4

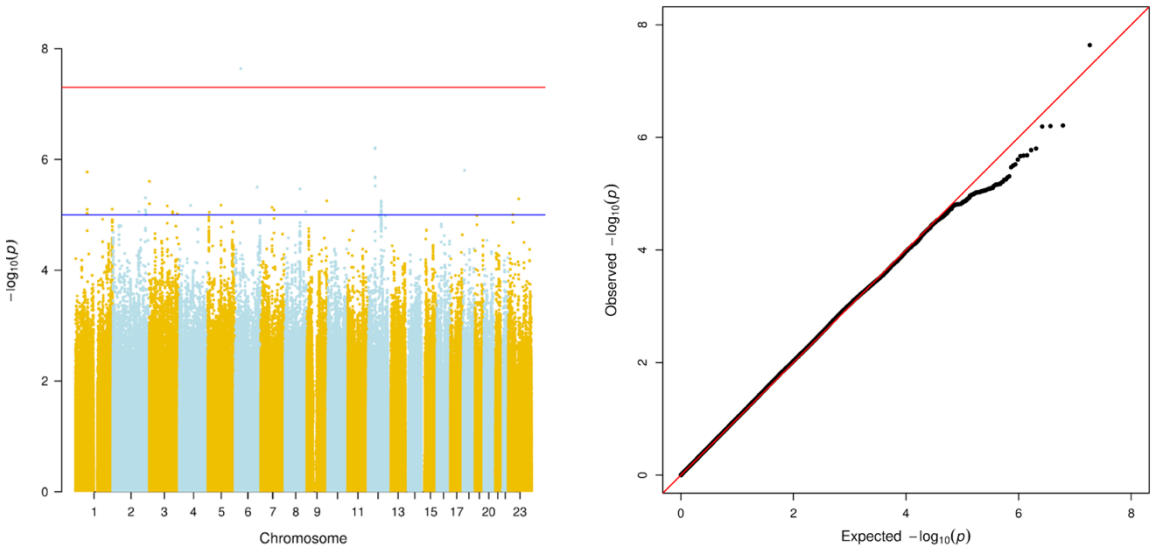


Figure *I*. QQ and Manhattan plots from GWAS of complete case WSAS (WSAS-cc) and 4-item WSAS (WSAS-4)

The red line on the QQ plot presents the null hypothesis of no significantly associated variants. A tail of points above this line in the top right of the plot would represent deviations from the null indicating true genetic signal. The red line on the Manhattan plot indicates the genome-wide significance threshold, corrected for multiple testing ($p < 5x10^{-8}$); points above this line represent variants significantly associated with the phenotype. The blue line indicates the threshold for suggestive significance ($p < 1x10^{-5}$).

Table *I*. GCTA-GREML SNP-based heritability estimates

Trait	h^2	h^2 SE	95% CI	*p*	n *
PHQ9	0.192	0.035	0.123 - 0.260	$6x10^{-9}$	16709
GAD7	0.173	0.035	0.104 - 0.241	$2x10^{-7}$	16725
WSAS-cc	0.133	0.038	0.059 - 0.208	$1x10^{-4}$	14744
WSAS-4	0.111	0.033	0.045 - 0.176	$2x10^{-4}$	16710

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS-cc = complete case 5-item functional impairment score, WSAS-4 = complete case 4-item functional impairment score. h^2 = heritability, SE = standard error, 95% CI = 95% confidence interval.

Table *II*. LDSC SNP-based heritability estimates

LDSC uses summary statistics and an external reference panel for LD scores making it less powerful than GCTA-GREML (van Rheenen et al., 2019; Yang et al., 2017).

Trait	h^2	h^2SE	95% CI	*p**	*z*	λ	Mean χ^2	Intercept	Intercept SE
PHQ9	0.104	0.029	0.048 - 0.160	$3x10^{-4}$	3.660	1.047	1.046	1.010	0.007
GAD7	0.082	0.030	0.024 - 0.141	0.006	2.774	1.028	1.034	1.007	0.007
WSAS	0.117	0.031	0.056 - 0.177	$2x10^{-4}$	3.761	1.032	1.033	0.993	0.007
WSAS-cc	0.125	0.032	0.062 - 0.188	$1x10^{-4}$	3.873	1.005	1.001	0.963	0.007
WSAS-4	0.111	0.030	0.053 - 0.170	$2x10^{-4}$	3.735	1.034	1.034	0.995	0.007

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = imputed 5-item functional impairment score used in main analysis, WSAS-cc = complete case 5-item functional impairment score, WSAS-4 = complete case 4-item functional impairment score. h^2 = heritability, SE = standard error, 95% CI = 95% confidence interval. *Calculated in R as pchisq((h^2/se)^2,1,F).

Table *III*. Bivariate-GREML genetic correlation (r_g) estimates

	PHQ9	GAD7	WSAS-cc	WSAS-4
PHQ9	1			
GAD7	0.87 (0.77 - 0.97)	1		
WSAS-cc	0.80 (0.65 - 0.94)	0.70 (0.50 - 0.89)	1	
WSAS-4	0.89 (0.75 - 1.02)	0.78 (0.60 - 0.97)	0.99 (0.98 - 1.00)	1

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS-cc = complete case 5-item functional impairment score, WSAS-4 = complete case 4-item functional impairment score. Values in brackets are 95% confidence intervals calculated as r_g +/- (1.96*SE).

Table *IV*. LDSC genetic correlation (r_g) estimates

	r_g	95% CI	*p*	*z*
PHQ9 - GAD7	0.85	0.63 - 1.07	$5x10^{-14}$	7.52
PHQ9 - WSAS	0.82	0.61 - 1.03	$1x10^{-14}$	7.70
GAD7 - WSAS	0.77	0.46 - 1.08	$1x10^{-6}$	4.82
PHQ9 - WSAS-cc	0.85	0.63 - 1.08	$6x10^{-14}$	7.51
GAD7 - WSAS-cc	0.79	0.45 - 1.13	$5x10^{-6}$	4.58
PHQ9 - WSAS-4	0.83	0.61 - 1.04	$5x10^{-14}$	7.54
GAD7 - WSAS-4	0.72	0.38 - 1.05	$3x10^{-5}$	4.19
WSAS-cc - WSAS-4	1.02	0.97 - 1.06	0	44.91

PHQ9 = depression symptoms, GAD7 = anxiety symptoms, WSAS = imputed 5-item functional impairment score used in main analysis, WSAS-cc = complete case 5-item score, WSAS-4 = complete case 4-item WSAS score. 95% CI = 95% confidence interval calculated as r_g +/- (1.96*SE).

LDSC heritability estimates indicated some lack of power for genetic correlations; z-scores < 4 are often too noisy (Bulik-Sullivan, Finucane, et al., 2015). LDSC jack-knife analyses revealed that the genetic correlation between the PHQ9 and imputed WSAS was not significantly different from 1 (p = 0.15), and nor was the correlation between the GAD7 and imputed WSAS (p = 0.18). This difference from analyses using GCTA is likely due to larger standard errors from LDSC estimates with summary statistics.

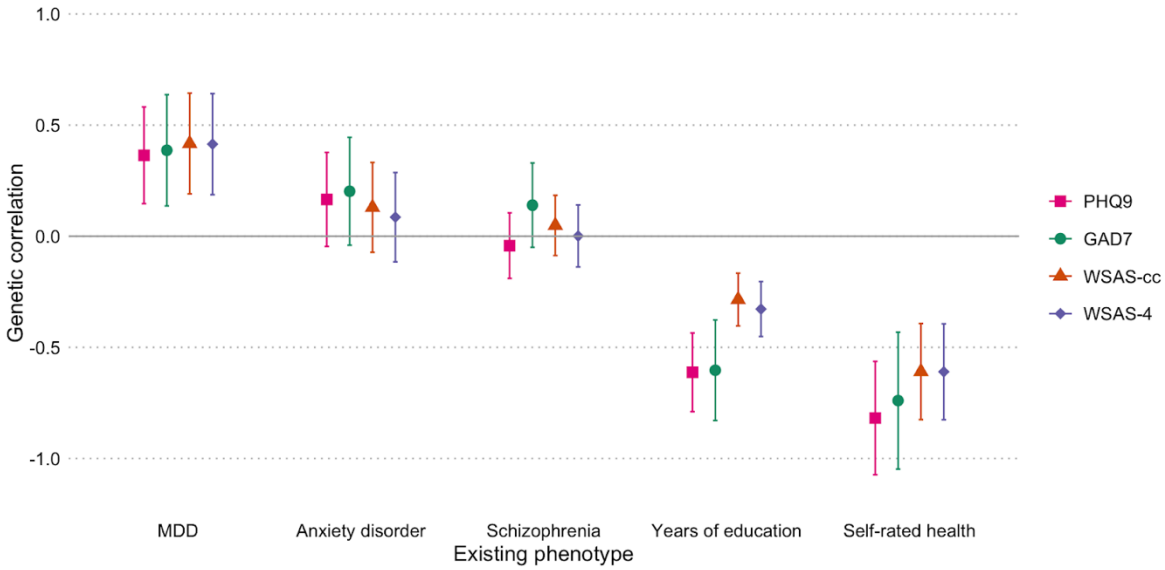


Figure *II*. Genetic correlations (LDSC SNP-based) between the traits analysed in the present study and five existing phenotypes

PHQ9 = depression symptom score, GAD7 = anxiety symptom score, WSAS-cc = complete case 5-item functional impairment score, WSAS-4 = complete case 4-item functional impairment score. Anxiety (Purves et al., 2020), depression (Wray et al., 2018), without 23andme or UKBiobank), schizophrenia (Pardiñas et al., 2018), self-rated health (Harris et al., 2017), years of education (J. J. Lee et al., 2018). See Supplementary Table 4 for further details of these phenotypes. Error bars represent 95% confidence intervals. * Significance at $p < 0.01$.

References

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Harris, S. E., Hagenaars, S. P., Davies, G., David Hill, W., Liewald, D. C. M., Ritchie, S. J., … Deary, I. J. (2017). Molecular genetic contributions to self-rated health. *International Journal of Epidemiology*, *46*(3), 994–1009. https://doi.org/10.1093/ije/dyw219

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., … Walters, J. T. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. https://doi.org/10.1038/s41588-018-0059-2

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A Major Role for Common Genetic Variation in Anxiety Disorders. *Molecular Psychiatry*, *25*(12), 3292–3303. https://doi.org/10.1038/s41380-019-0559-1

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews. Genetics*. https://doi.org/10.1038/s41576-019-0137-z

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*(9), 1304–1310. https://doi.org/10.1038/ng.3941

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., … Neale, B. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*(2), 272–279. https://doi.org/10.1093/bioinformatics/btw613

Appendix C - Supplementary materials for Chapter 4

Supplementary materials for "Trajectories of depression and anxiety symptom severity during psychological therapy for common mental health problems"

Grant Numbers and Disclosures
MS was funded by an NIHR Maudsley Biomedical Research Centre studentship. JEJB was funded on a fellowship from the Wellcome Trust (201292/Z/16/Z). CR was supported by a grant from Fondation Peters to TCE and GB. AJP was supported by an ESRC studentship. GB has sat on advisory boards for Otsuka and Compass Pathways. The authors declare no conflict of interest.

Acknowledgements
This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, and the Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or King's College London. The authors acknowledge the valuable work of the NIHR Maudsley Biomedical Research Centre Clinical Record Interactive Search team, especially Matthew Broadbent, who performed the data extraction.

Author contribution
All authors were involved in conceptualising the study. MS performed the data pre-processing and analysis with technical input from EC, JEJB, KAG, and RS. Clinical expertise was provided by JEJB, CRH, KAR, RS and JW. MS drafted the manuscript and all other authors provided critical feedback. All authors approved the final version prior to submission. TCE and GB supervised this work.

Ethical statement
We assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The use of anonymised data in this study was approved by the South Central - Oxford C Research Ethics Committee (15/SC/0586) and CRIS Oversight Committee (16010).

Data availability
This study used anonymised patient data from the South London and Maudsley NHS Foundation Trust that is not publicly available. The data can be accessed within a secure

firewall via the Clinical Record Interactive Search (CRIS) tool by persons with the appropriate permissions. For more information please contact: cris.administrator@slam.nhs.uk.

1. Supplementary Information 1. Data pre-processing and additional exclusion criteria
2. Supplementary Table 1. Number of time points (sessions) for patients in the analysis
3. Supplementary Table 2. Descriptives of time intervals between time points (sessions) for patients in the analysis
4. Supplementary Information 2: Growth mixture model method
5. Supplementary Figure 1 (a - c). Latent growth curves of depression and anxiety symptoms by the four treatment sites, and latent class growth analysis of depression symptoms for each site
6. Supplementary Figure 2. Histograms of observed depression and anxiety symptoms per time point (session)
7. Supplementary Table 3. Proportion of patients in treatment at each time point (session) with observed depression and anxiety symptoms
8. Supplementary Table 4. Comparison of baseline characteristics of patients from the original data extraction who were excluded from analysis to patients included in the analysis
9. Supplementary Information 3. Latent growth curves of depression and anxiety symptoms
10. Supplementary Information 4. Latent class growth analysis of depression and anxiety symptoms
11. Supplementary Information 5. Selection of depression symptoms growth mixture model
12. Supplementary Information 6. Descriptives of four-class growth mixture model of depression symptoms
13. Supplementary Information 7. Selection of anxiety symptoms growth mixture model
14. Supplementary Information 8. Descriptives of four-class growth mixture model of anxiety symptoms
15. Supplementary Table 5. Multinomial regression output for conditional four-class growth mixture model of depression symptoms (PHQ9)
16. Supplementary Table 6. Multinomial regression output for conditional four-class growth mixture model of anxiety symptoms (GAD7)

Supplementary Information 1. Data pre-processing and additional exclusion criteria

We initially extracted treatment records for 115,304 patients covering 587,120 sessions. We performed extensive data pre-processing, as is often required for electronic health records, and imposed additional exclusion criteria, under guidance from experienced IAPT clinicians. A core component was classification of free-text responses. For example, the intervention variable exceeded 100 distinct values, which required mapping to categories of high-intensity, low-intensity, and non-intervention (e.g., triage). We were then able to restrict analysis to individuals who had received high-intensity therapy, which was required due to the difference in the average number of sessions and level of structure between intensities. We imposed additional exclusion criteria to patients and sessions to obtain valid outcome data suitable for growth mixture modelling. We did not limit analyses to individuals who scored above clinical thresholds on the PHQ9 or GAD7 at baseline, as this would not accurately reflect the IAPT population. The resulting sample for analysis consisted of 16,258 patients covering 110,773 sessions. The largest exclusion was patients who did not receive at least two sessions of high-intensity therapy (N = 94,979). We ideally would have limited our original extraction to such individuals but were unable to due to the lack of a reliable variable indicating treatment intensity. Furthermore, we could not perform this exclusion earlier in the pre-processing as valid treatment sessions had to first be identified to count the number of each intensity. Several variables of interest were not included in analyses due to high, potentially non-random missingness e.g., at baseline the social phobia, agoraphobia and specific phobia items were each 49% missing. Data pre-processing and descriptives were performed in R version 3.6.3.

Supplementary Table 1. Number of time points (sessions) for patients in the analysis (N = 16,258)

Total sessions received	Frequency	Proportion (%)	Cumulative Proportion (%)
3	1885	11.59	11.59
4	1804	11.10	22.69
5	2153	13.24	35.93
6	2706	16.64	52.58
7	2046	12.58	65.16
8	1080	6.64	71.80
9	884	5.44	77.24
10	717	4.41	81.65
11	2983	18.35	100.00

Note: To identify a sample who received high-intensity therapy, patients were included in the analysis if they had attended at least three sessions, with at least two being high-intensity treatment (and permitting one low-intensity and/or one assessment and triage session). The **mean number of sessions** was 7.3 (SD = 3.8) and ranged 3-55. The number of sessions was then limited to maximum 11, including the baseline assessment, to be reasonably representative of the number received in the sample (within 1 SD of mean**; < 15% received more than 11)**, and to provide sufficient complete data for analysis using maximum likelihood estimation for missing data (covariance coverage above Mplus default of 0.10). **Following this, the mean was 6.8 sessions** (SD = 2.9).

Supplementary Table 2. Descriptives of time intervals (days) between time points (sessions) for patients in the analysis (N = 16,258)

	Time point (Session)										
Dataset	0	1	2	3	4	5	6	7	8	9	10
Unfiltered	0.00 (0.00)	31.21 (50.82)	19.47 (33.80)	13.28 (17.03)	12.72 (15.39)	12.93 (17.90)	13.13 (15.86)	13.44 (15.27)	13.33 (14.07)	13.89 (16.54)	13.87 (17.24)
Filtered	0.00 (0.00)	12.53 (7.17)	10.41 (5.68)	10.27 (5.51)	10.16 (5.40)	10.13 (5.48)	10.23 (5.46)	10.59 (5.78)	10.46 (5.63)	10.46 (5.72)	10.46 (5.64)

Note: Mean (standard deviation) days between column-specified session and previous session. Descriptives are provided for patients *prior* to filtering out sessions occurring after intervals exceeding 30 days (unfiltered) and *following* this (filtered). The filtered dataset was used for analysis. In both datasets we had previously removed the baseline session if there was an interval exceeding 30 days to session 1 as this had the separate purpose of identifying a baseline session (the unfiltered mean interval here exceeds 30 as the baseline to session 1 filter was performed only once therefore the 'new' session 1 could have occurred after a long interval from the previous session 1).

Supplementary Information 2: Growth mixture model method

We estimated growth mixture models (GMMs) for depression symptoms and for anxiety symptoms.

Step 1: Latent growth curve analysis. Used to identify the best-fitting single, average, latent growth curve (trajectory). The latent growth curve consists of an estimated mean intercept and estimated mean slope. This trajectory describes the pattern of symptom change observed across *all* patients as though they are a homogeneous group. Linear, quadratic, and negative log-linear (base 10) latent growth curves were compared to determine which form most closely represented the observed data. Each was run with and without correlations between the residuals of adjacent timepoints. Model fit was assessed using Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Standardized Root Mean Square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA). Lower values indicate superior fit, with recommendations of SRMR ≤ 0.08 and RMSEA ≤ 0.06 (Hu & Bentler, 1999; Nylund, Asparouhov & Muthén, 2007; Schwarz, 1978). We also used the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), where values closer to 1 indicate better fit, ideally ≥ 0.95 (Hu & Bentler, 1999).

Step 2: Multiple classes of trajectories. Used to determine whether patients' observed data is better explained by multiple average latent growth curves than a single average one. This is done by introducing a latent factor of 'class' into the model. Each class of patients had a trajectory of the best-fitting form identified in Step 1, but with its own specific mean intercept and slope. In GMM, the variance within the intercept and/or slope of each trajectory class is free, allowing varying expressions of estimated individual trajectories around the mean. A trajectory class therefore represents multiple similar trajectories. In a restricted version of GMM called latent class growth analysis (LCGA), the within-class variance is fixed to zero meaning individuals in a class follow the same trajectory. LCGA is less computationally demanding and can provide a useful representation of the true trajectories. However, it is unlikely that individuals in a class follow one same trajectory

(Muthén, 2002; Nagin & Odgers, 2010) and LCGA can result in spurious classes, including classes that differ solely by intercept, as only classe can explain variance (Muthén, 2002; Bauer & Curran, 2004). We followed recommendations to first perform LCGA and inspect results to determine whether to run GMM.

Step 3: Run GMM with the variance freed in the intercepts. We modelled up to six trajectory classes for LCGA and GMM. Each model was estimated using a series of steps to aid global as opposed to local solutions (see below) and compared to the model with one fewer class. The model fit criteria were AIC, BIC and Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT; Mplus 'TECH11'). A significant *p*-value (< 0.05) for the VLMR-LRT indicates the current model is a significantly better fit of the observed data than a model with one fewer class. We prioritised BIC, which is preferred in this field (van de Schoot et al., 2017). However, it is possible with large sample sizes that as the number of classes increases, information criteria values fail to reach a minimum, in which case it can be useful to identify the point of diminishing gains (decreases in BIC) in an elbow plot (Meyer & Morin, 2016; Petras & Masyn, 2010).

Each individual has a likelihood, 'posterior probability', of belonging to each class i.e., following an estimated trajectory from that class. A useful model has good distinction; each patient has a high likelihood of membership to one class and low probabilities for the others. This is captured by entropy, which ranges between 0 (equal probability of belonging to each class) and 1 (distinct classifications). A common rule of thumb is that > 0.8 indicates high class separation, 0.6 medium, and 0.4 low (Clark & Muthén, 2009). Entropy was only consulted for model selection when other indices were similar between models (Petras & Masyn, 2010; van de Schoot et al., 2017). We favoured models that were clinically interpretable and reasonable in terms of theory and previous literature. More parsimonious models were preferred, especially if a model had an additional class with a similar slope to one in the previous model and only the baseline score (intercept) differed. To aid this process, we plotted the mean estimated trajectory of each class per model. Finally, we required that models converged and that estimated values were in the range of the outcome measure.

Procedure for conducting growth mixture modelling in Mplus

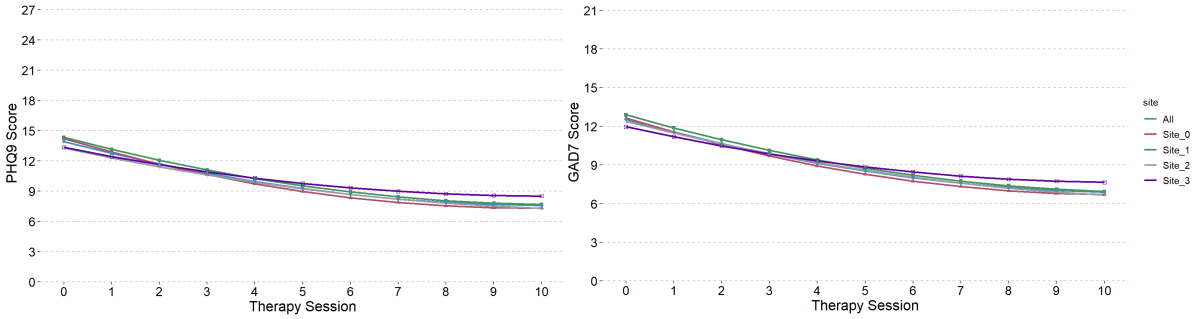
To help ensure global as opposed to local solutions, the following procedure was used in Mplus (Version 8.3) for one- to six-class GMMs (based on Asparouhov & Muthén, 2012; Jung & Wickrama, 2008; Wickrama et al., 2016):

1. Run model with 400 initial stage random starting values and 100 top log-likelihood values brought to the final stage, as recommended, with maximum 10 iterations.
2. Check whether the best log-likelihood value replicates. Check for normal termination of model estimation, negative residual variances and other warnings or errors in the output. If the log-likelihood does not replicate, increase the number of random starts in both stages until it does.

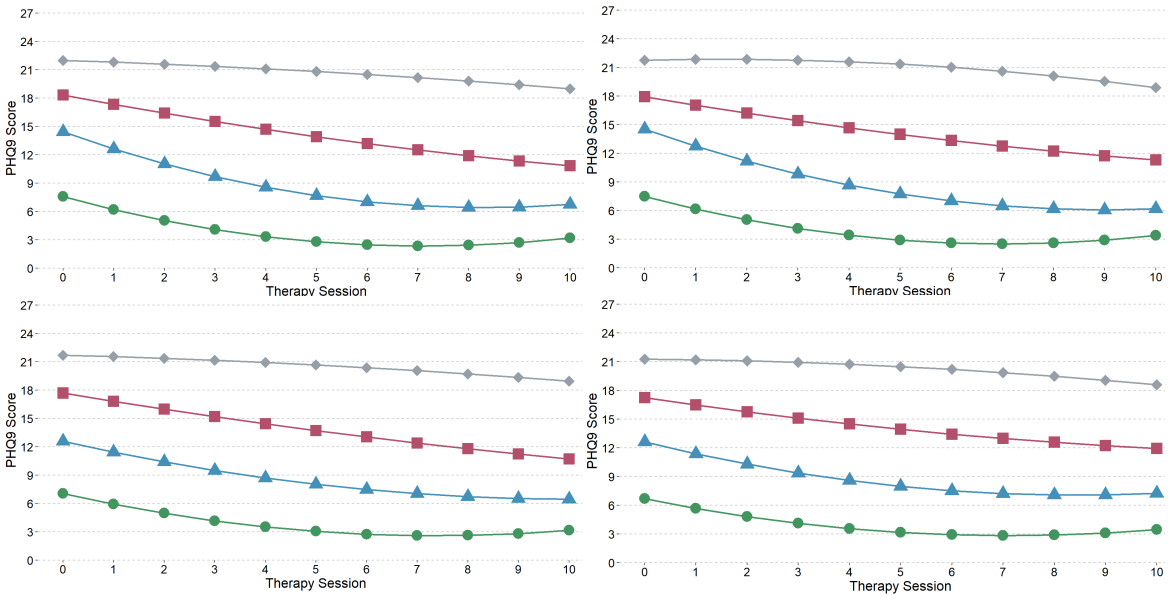
3. Once the log-likelihood replicates, check that it is not a local solution by taking the seed values from the top two log-likelihood values and re-running the model using the Mplus optseed argument and setting the starts argument to zero. At this stage, include TECH11 (the VLMR test).
4. Ensure that the log-likelihood value for the K-1 class in the TECH11 output matches that of the actual K-1 class model. If not, increase the starting values in the K-1 starts argument of the input.

Supplementary Figure 1 (a - b). Latent growth curves of depression (PHQ9) and anxiety (GAD7) symptoms per treatment site, and latent class growth analysis of depression symptoms for each site

Latent growth curves for symptoms of patients from each of the four IAPT treatment sites were highly similar (Figure a). Latent-class growth analysis was run for one outcome (PHQ9) and showed similar trajectories (see four-class models in Figure b).

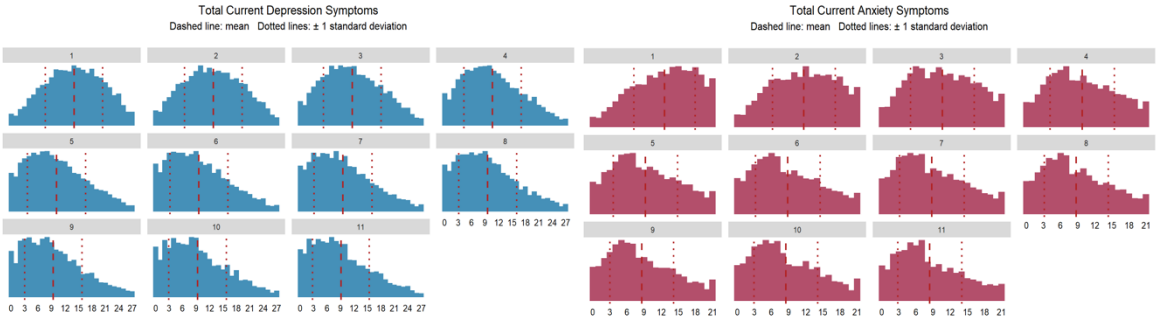


a. Latent growth curves of depression symptoms (PHQ9, left) and anxiety symptoms (GAD7; right), per treatment site and across all sites ('All')



b. Four-class latent class growth analysis of depression symptoms (PHQ9), for each IAPT treatment site

Supplementary Figure 2. Histograms of observed depression (PHQ9; left) and anxiety symptoms (GAD7; right) per time point (session)



Note: Histograms for total symptom scores measured across 11 time points (baseline and 10 treatment sessions). Observed scores for patients in treatment at each time point are shown; see Supplementary Table 3 for Ns

Supplementary Table 3. Proportion of patients in treatment at each time point (session) with observed depression (PHQ9) and anxiety (GAD7) symptoms

Time point (session)	Number of patients in treatment	Proportion of patients with PHQ9 score	Proportion of patients with GAD7 score
0	16258	99.06%	99.05%
1	16258	99.48%	99.48%
2	16258	99.38%	99.37%
3	14373	99.52%	99.49%
4	12569	99.53%	99.51%
5	10416	99.65%	99.64%
6	7710	99.52%	99.49%
7	5664	99.35%	99.35%
8	4584	99.50%	99.48%
9	3700	99.70%	99.70%
10	2983	99.63%	99.63%

Supplementary Table 4. Comparison of baseline characteristics of patients from the original data extraction who were excluded from analysis to patients included in the analysis

Variable	Excluded from analysis (n = 94,979)	Included in analysis (n = 16,258)	*p*-value	Effect size
Age (years)			< 0.001	0.12
Mean (SD; range)	35.96 (13.47; 16.00-100.00)	37.55 (13.36; 16.00-94.00)		
Gender			< 0.001	0.02
Female	61801 (65.18%)	10979 (67.60%)		
Male	33019 (34.82%)	5262 (32.40%)		
Missing	159	17		
Depression symptoms (PHQ9)			0.239	0.01
Mean (SD; range)	13.91 (6.48; 0.00-27.00)	13.98 (6.39; 0.00-27.00)		
Missing	2829	153		
Anxiety symptoms (GAD7)			0.771	< 0.01
Mean (SD)	12.52 (5.46; 0.00 - 21.00)	12.53 (5.39; 0.00 - 21.00)		
Missing	2885	154		
Case on PHQ9 and/or GAD7 [1]			0.679	< 0.01
Yes	78434 (85.17%)	13735 (85.29%)		
No	13658 (14.83%)	2368 (14.71%)		
Missing	2887	155		
Functional impairment score (WSAS)			< 0.001	0.06
Mean (SD; range)	18.15 (9.56; 0.00 - 40.00)	17.58 (9.31; 0.00 - 40.00)		
Missing	25282	5447		
Problem descriptor [2]			< 0.001	0.14
Depression	35509 (43.62%)	6703 (45.74%)		
GAD	17181 (21.11%)	1393 (9.51%)		
Other	7240 (8.89%)	1423 (9.71%)		
MADD	6789 (8.34%)	1129 (7.70%)		
Panic/phobia	6329 (7.77%)	1003 (6.85%)		
Adjustment disorder	4429 (5.44%)	1320 (9.01%)		
PTSD	2761 (3.39%)	1132 (7.73%)		
OCD	1169 (1.44%)	550 (3.75%)		
Missing	13572	1605		
Psychotropic medication			0.466	< 0.01
Prescribed	31024 (35.42%)	5545 (35.72%)		
Not prescribed	56565 (64.58%)	9977 (64.28%)		
Missing	7390	736		
Ethnicity			0.268	< 0.01
White	56994 (62.92%)	9789 (63.64%)		
Black	17654 (19.49%)	2964 (19.27%)		
Mixed	6565 (7.25%)	1111 (7.22%)		

Asian	6050 (6.68%)	961 (6.25%)		
Other	3320 (3.67%)	557 (3.62%)		
Missing	4396	876		
Employment status			0.002	0.01
Employed	57249 (64.21%)	10033 (63.39%)		
Unemployed	19009 (21.32%)	3572 (22.57%)		
Non-worker [3]	12897 (14.47%)	2222 (14.04%)		
Missing	5824	431		
Disability [4]			0.004	< 0.01
Yes	8527 (8.98%)	1575 (9.69%)		
No	86452 (91.02%)	14683 (90.31%)		
Number of sessions (including baseline assessment)			< 0.001	0.96
Mean (SD)	4.01 (4.53; 1.00 - 73.00)	8.35 (4.53; 3.00 - 75.00)		
Recovered [5]			< 0.001	0.20
Yes	20424 (26.72%)	7028 (51.98%)		
No	56008 (73.28%)	6493 (48.02%)		
Missing	18547	2737		
Reason for end of treatment			< 0.001	0.30
Discharged	36322 (38.24%)	13138 (80.81%)		
Dropout	36821 (38.77%)	2328 (14.32%)		
Referred	21836 (22.99%)	792 (4.87%)		
Service			< 0.001	0.13
0	26143 (27.53%)	7027 (43.22%)		
1	21573 (22.71%)	3402 (20.93%)		
2	21579 (22.72%)	3244 (19.95%)		
3	25684 (27.04%)	2585 (15.90%)		

Note: Percentages were calculated using the available sample for each variable, after excluding missing values. The "Missing" row represents the number of missing values and is omitted if there was no missing data. Effect sizes are Cohen's *d* for continuous variables and Cramer's *V* for categorical. *p*-values are from Chi-Square tests for categorical variables and ANOVAs/t-tests for continuous variables. **The descriptives of the analytical sample here were created prior to any filtering of sessions, hence differences in e.g., number of sessions and recovery rates to those reported in Table 1 for the cleaned analytical sample**. [1] Case thresholds were PHQ9 ≥10, GAD7 ≥8. [2] GAD = generalised anxiety disorder; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; MADD = mixed anxiety and depressive disorder; Panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; 'Other' included somatoform disorder, severe mental illness. Differences in problem descriptor frequencies partly reflect inclusion criteria, as some disorder-specific treatments are more likely to be high-intensity. [3] 'Non-worker' included homemaker, carer, retired, student. [4] No negative responses were recorded, therefore the absence of any value was taken as a negative response rather than missing. [5] Only calculated for patients who scored above case thresholds on either the PHQ9 or GAD7 at the start of treatment and had an observed score for their final session, otherwise were coded as missing. Represents whether the patient reached recovery within the 10 treatment sessions modelled; if they received more sessions and then recovered, they would appear unrecovered here.

Supplementary Information 3. Latent growth curves of depression and anxiety symptoms

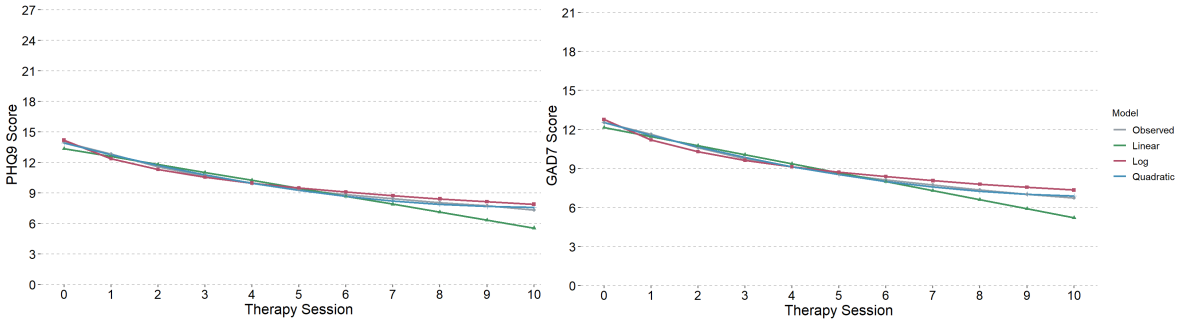
Fit indices and plots of single latent growth curves of different form.

Fit indices for latent growth curves of depression and anxiety symptoms during psychological therapy

Depression Symptoms (PHQ9) Latent Growth Curve Form	**Parameters**	**AIC**	**BIC**	**CFI**	**TLI**	**SRMR**	**RMSEA estimate**
Quadratic with pairwise residual correlations	30	609627	609858	0.994	0.993	0.019	0.024
Logarithmic with pairwise residual correlations	26	611250	611450	0.975	0.973	0.041	0.047
Quadratic	20	611570	611724	0.973	0.974	0.030	0.046
Linear with pairwise residual correlations	26	611771	611971	0.969	0.967	0.109	0.052
Logarithmic	16	615385	615508	0.933	0.939	0.057	0.070
Linear	16	615502	615625	0.930	0.937	0.149	0.071
Anxiety Symptoms (GAD7) Latent Growth Curve Form	**Parameters**	**AIC**	**BIC**	**CFI**	**TLI**	**SRMR**	**RMSEA estimate**
Quadratic with pairwise residual correlations	30	591154	591385	0.994	0.992	0.021	0.024
Logarithmic with pairwise residual correlations	26	592934	593134	0.971	0.969	0.046	0.049
Quadratic	20	593028	593181	0.971	0.972	0.034	0.046
Linear with pairwise residual correlations	26	593040	593240	0.969	0.967	0.115	0.050
Linear	16	596403	596526	0.930	0.937	0.154	0.069
Logarithmic	16	596974	597097	0.924	0.932	0.064	0.072

Note: Rows are ordered by BIC, with the optimal model first. AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, RMSEA = Root Mean Square Error of Approximation. Lower values indicate superior fit for AIC, BIC, SRMR, RMSEA. Higher values closer to 1 indicate better fit for the CFI and TLI.

Latent growth curves of depression (PHQ9; left) and anxiety (GAD7; right) symptoms during psychological therapy



Note: The intercept represents the estimated mean outcome score across the whole sample at baseline (session 0). Only trajectories with the residuals correlated are shown as were a better fit than forms with uncorrelated residuals.

Supplementary Information 4. Latent class growth analysis of depression and anxiety symptoms

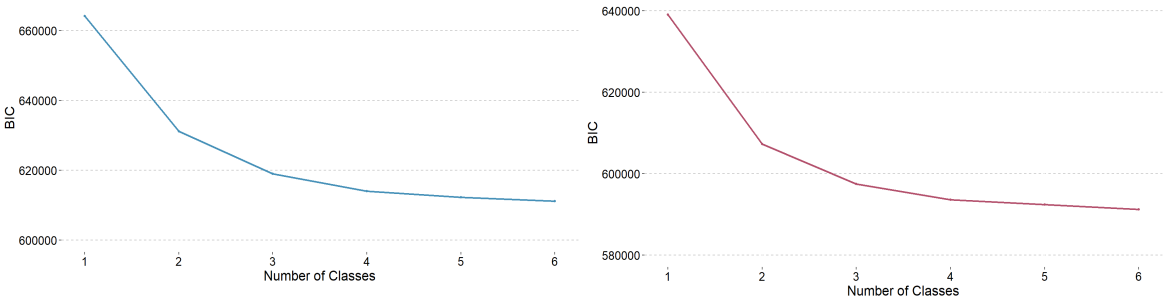
This section describes the model selection for latent class growth analysis (LCGA), where the variance within classes is restricted to zero. It includes fit indices for the estimated models, and plots of the selected models. The LCGA of depression symptoms and LCGA of anxiety symptoms suggested a four-class model. All models had class trajectories that primarily differed in baseline severity and classes with more than 1% of the sample in each class. For the depression models, entropy values were good for the two- to four-class models (and acceptable for the others). The anxiety models showed good entropy for the two- to three-class models (and acceptable for four- to six-class). The VLMR-LRT *p*-values were significant for all models besides the six-class depression model, indicating that it was a poorer fit of the data than a five class-model. The information criteria continued to decrease up to the six-class model for both depression and anxiety symptoms, however, each of the elbow plots of BIC values showed a plateau around four classes followed by a negligible decrease for the five-class model. Therefore, the four class models were selected. For each of the four class models, there was some indication of a moderate-severe plateau class (grey diamonds), moderate-severe with gradual improvement (pink squares), slightly faster improvement to plateau (blue triangles) and a class with mild symptoms that showed small improvement (green circles).

Fit indices for latent class growth analysis of depression and anxiety symptoms

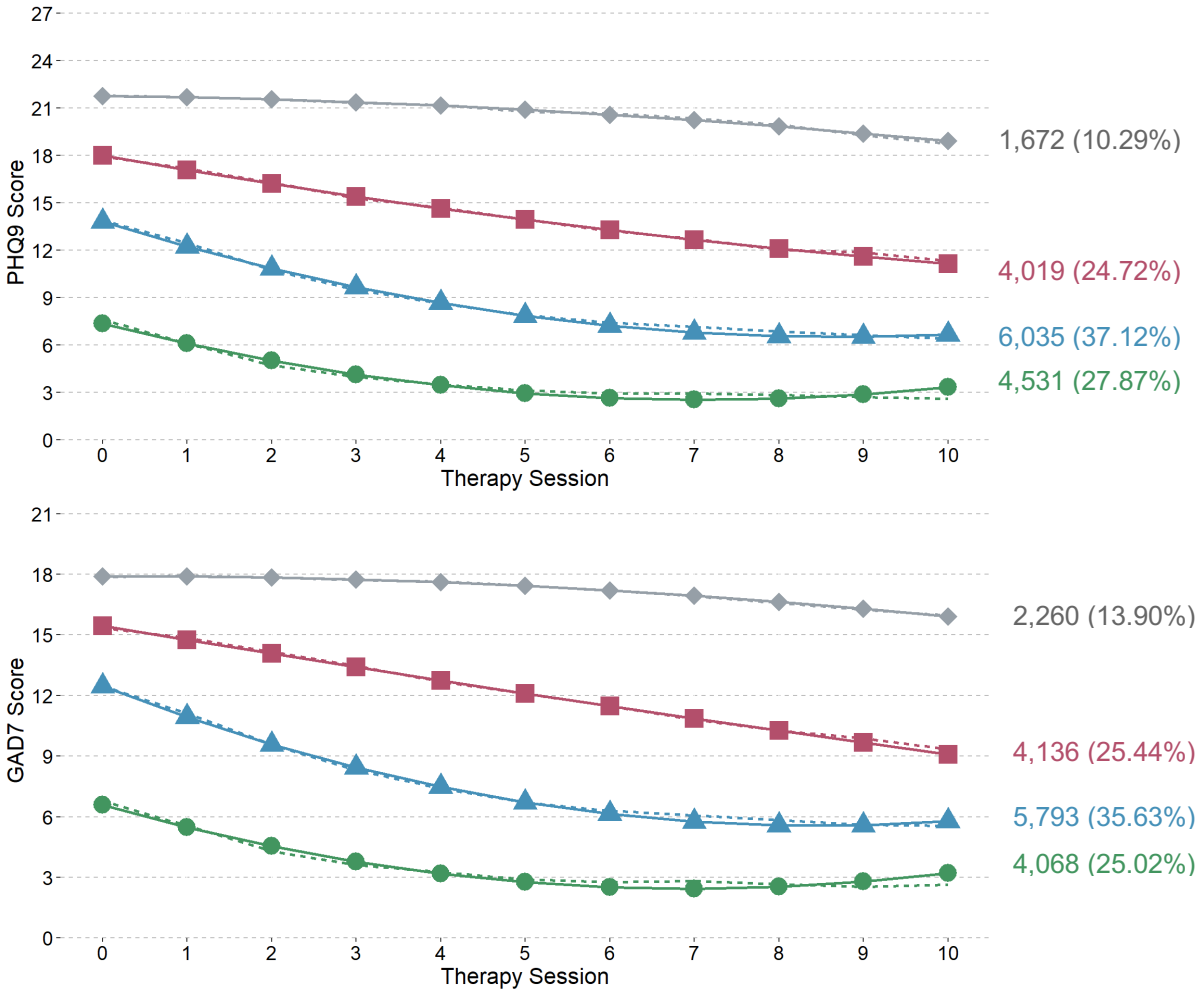
Depression symptoms (PHQ9) LCGA	**Params**	**AIC**	**BIC**	**Entropy**	**VLMR LRT *p*-value**	**Individuals per class (%)**
Growth Curve	24	664047	664232	NA	NA	100
Two Class	28	630926	631142	0.847	< 0.001	62.1, 37.9
Three Class	32	618738	618984	0.826	< 0.001	17.8, 41.8, 40.4
Four Class	36	613685	613962	0.799	< 0.001	27.9, 10.3, 37.1, 24.7
Five Class	40	611898	612206	0.763	0.004	32.9, 6.5, 24.2, 21.4, 15.0
Six Class	44	610804	611142	0.740	0.085	18.3, 30.8, 8.2, 17.6, 17.8, 7.2
Anxiety symptoms (GAD7) LCGA	**Params**	**AIC**	**BIC**	**Entropy**	**VLMR LRT *p*-value**	**Individuals per class (%)**
Growth Curve	24	638901	639086	NA	NA	100
Two Class	28	607070	607285	0.839	< 0.001	58.7, 41.3
Three Class	32	597165	597411	0.801	< 0.001	23.3, 40.5, 36.2
Four Class	36	593260	593537	0.766	< 0.001	13.9, 25.4, 35.6, 25.0
Five Class	40	592043	592351	0.720	< 0.001	15.7, 10.8, 19.1, 24.7, 29.7
Six Class	44	590846	591185	0.707	0.0002	10.7, 29.1, 18.9, 15.0, 12.0, 14.3

Note: A quadratic form with correlations between the residuals of adjacent time points was specified for all classes. Params = parameters, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, VLMR LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. Number of parameters in the growth curve (single class) differs from the latent growth curve (24 compared with 30) as the slope and intercept variance and covariance are fixed here. Individuals per class is based on a patient's highest posterior probability of belonging to a class.

Elbow plots of Bayesian Information Criterion values for latent class growth analysis of depression (PHQ9; left) and anxiety (GAD7; right) symptoms



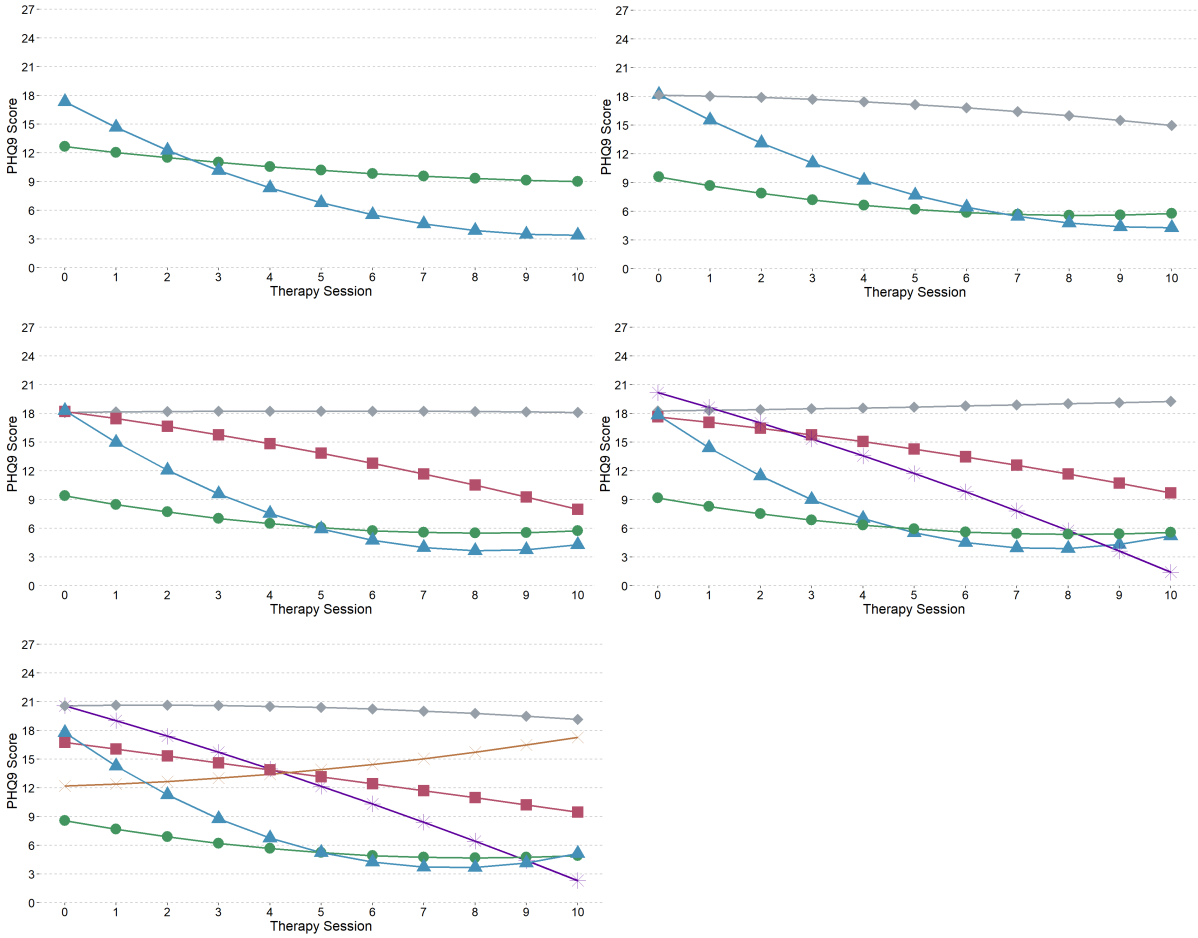
Four-class latent class growth models of depression (PHQ9; top) and anxiety (GAD7; bottom) symptoms during psychological therapy (N = 16,258)



Supplementary Information 5. Selection of depression symptoms growth mixture model

This section describes the model selection for the growth mixture model of depression symptoms (PHQ9), including fit indices (provided in the main text) and trajectory plots. Information criteria continued to decrease up to a six-class model and the VLMR LRT *p*-value did not become non-significant, however, six classes were unrealistically high compared with existing studies (hence chosen as the upper number of classes to test). The BIC values elbow-plot suggested a four-class model, consistent with previous literature, and this was selected as the optimal model. Entropy of the four-class model was medium (0.60).

Two- to six-class growth mixture models of depression symptoms during psychological therapy



Supplementary Information 6. Descriptives of the selected four-class growth mixture model of depression symptoms (PHQ9)

Descriptives of the growth factors for the four-class growth mixture model of depression symptoms

Class	Parameter	Factor	Estimate	SE	Est SE	*p*-value
Moderate-severe plateau	Means	Intercept	18.11	0.18	102.677	0
		Linear	0.06	0.06	1.047	0.295
		Quadratic	-0.01	0.01	-0.851	0.395
	Variances	Intercept	11.6	0.24	48.404	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Moderate-severe, gradual improvement	Means	Intercept	18.22	0.18	103.533	0
		Linear	-0.72	0.10	-7.17	0
		Quadratic	-0.03	0.01	-3.854	0
	Variances	Intercept	11.6	0.24	48.404	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Moderate-severe, fast improvement	Means	Intercept	18.31	0.14	127	0
		Linear	-3.54	0.10	-34.759	0
		Quadratic	0.21	0.01	16.972	0
	Variances	Intercept	11.6	0.24	48.404	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Mild, small improvement	Means	Intercept	9.41	0.11	85.924	0
		Linear	-0.97	0.03	-31.445	0
		Quadratic	0.06	0.00	20.742	0
	Variances	Intercept	11.6	0.24	48.404	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999

Model estimated depression symptom scores per therapy session and trajectory class

	Session										
Class	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Moderate-severe plateau	18.11	18.16	18.20	18.23	18.25	18.26	18.25	18.24	18.21	18.17	18.13
Moderate-severe, gradual improvement	18.22	17.47	16.65	15.78	14.85	13.86	12.80	11.69	10.52	9.29	7.99
Moderate-severe, fast improvement	18.31	14.99	12.09	9.62	7.57	5.96	4.77	4.01	3.68	3.77	4.29
Mild, small improvement	9.41	8.50	7.71	7.04	6.50	6.07	5.76	5.58	5.51	5.57	5.74

Descriptives of patients in each class of the depression symptoms model (assigned to their most likely trajectory class)

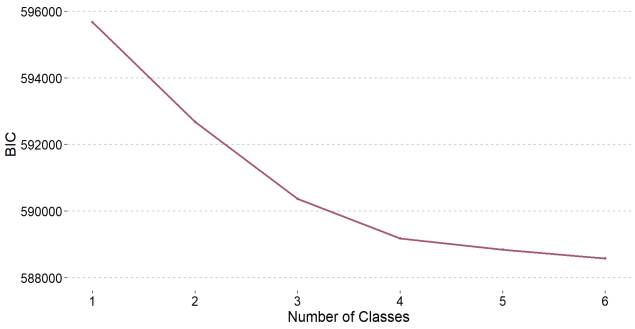
	Moderate-severe plateau (N=2200)	Moderate-severe, gradual improvement (N=2857)	Moderate-severe, fast improvement (N=2670)	Mild, small improvement (N=8530)
Depression symptoms (PHQ9)				
Mean (SD)	18.55 (5.04)	19.13 (3.80)	19.24 (3.58)	9.43 (4.34)
Range	0.00 - 27.00	0.00 - 27.00	6.00 - 27.00	0.00 - 25.00
Missing	26	32	14	80
Anxiety symptoms (GAD7)				
Mean (SD)	15.75 (4.30)	15.81 (3.86)	15.44 (4.14)	9.69 (4.82)
Range	0.00 - 21.00	0.00 - 21.00	0.00 - 21.00	0.00 - 21.00
Missing	26	32	15	80
Case on PHQ9 and/or GAD7 [1]				
Yes	2128 (97.88%)	2816 (99.68%)	2653 (99.92%)	6138 (72.65%)
No	46 (2.12%)	9 (0.32%)	2 (0.08%)	2311 (27.35%)
Missing	26	32	15	81
Functional impairment score (WSAS)				
Mean (SD)	23.11 (9.42)	22.98 (8.46)	20.80 (8.70)	13.74 (7.72)
Range	0.00 - 40.00	0.00 - 40.00	0.00 - 40.00	0.00 - 40.00
Missing	740	1032	1107	2567
Psychotropic medication				
Prescribed	1095 (52.54%)	1301 (48.06%)	1015 (39.26%)	2134 (26.20%)
Not prescribed	989 (47.46%)	1406 (51.94%)	1570 (60.74%)	6012 (73.80%)
Missing	116	150	85	384
Employment status				
Employed	956 (44.61%)	1531 (55.01%)	1658 (63.35%)	5888 (71.08%)
Unemployed	884 (41.25%)	875 (31.44%)	612 (23.39%)	1201 (14.50%)
Non-worker [2]	303 (14.14%)	377 (13.55%)	347 (13.26%)	1195 (14.43%)
Missing	57	74	53	246
Disability [3]				
Yes	367 (16.68%)	344 (12.04%)	254 (9.51%)	610 (7.15%)
No	1833 (83.32%)	2513 (87.96%)	2416 (90.49%)	7920 (92.85%)
Ethnicity				
White	1156 (56.53%)	1566 (58.43%)	1464 (57.71%)	5603 (69.01%)
Black	443 (21.66%)	555 (20.71%)	626 (24.67%)	1339 (16.49%)
Mixed	157 (7.68%)	205 (7.65%)	209 (8.24%)	540 (6.65%)
Asian	189 (9.24%)	231 (8.62%)	152 (5.99%)	389 (4.79%)
Other	100 (4.89%)	123 (4.59%)	86 (3.39%)	248 (3.05%)
Missing	155	177	133	411
Problem descriptor [4]				
Depression	968 (49.54%)	1396 (54.42%)	1362 (55.55%)	2976 (38.74%)
GAD	133 (6.81%)	170 (6.63%)	171 (6.97%)	919 (11.96%)
Other	128 (6.55%)	169 (6.59%)	198 (8.08%)	928 (12.08%)
MADD	167 (8.55%)	201 (7.84%)	202 (8.24%)	559 (7.28%)
Panic/phobia	103 (5.27%)	122 (4.76%)	105 (4.28%)	673 (8.76%)

Adjustment disorder	134 (6.86%)	146 (5.69%)	188 (7.67%)	852 (11.09%)
PTSD	255 (13.05%)	277 (10.80%)	175 (7.14%)	425 (5.53%)
OCD	66 (3.38%)	84 (3.27%)	51 (2.08%)	349 (4.54%)
Missing	246	292	218	849
Age (years)				
Mean (SD)	39.13 (13.36)	37.47 (13.44)	37.58 (13.31)	37.16 (13.31)
Range	17.00 - 90.00	17.00 - 91.00	17.00 - 89.00	16.00 - 94.00
Gender				
Female	1453 (66.11%)	1923 (67.43%)	1808 (67.74%)	5794 (68.00%)
Male	745 (33.89%)	929 (32.57%)	861 (32.26%)	2727 (32.00%)
Missing	2	5	1	9
Number of sessions (including baseline assessment)				
Mean (SD)	7.48 (2.79)	7.13 (2.86)	6.35 (2.38)	6.68 (2.62)
Range	3.00 - 11.00	3.00 - 11.00	3.00 - 11.00	3.00 - 11.00
Recovered [5]				
Yes	12 (0.57%)	522 (18.70%)	1790 (67.83%)	3579 (58.78%)
No	2101 (99.43%)	2270 (81.30%)	849 (32.17%)	2510 (41.22%)
Missing	87	65	31	2441
Reason for end of treatment				
Discharged	1532 (69.64%)	2115 (74.03%)	2187 (81.91%)	7303 (85.62%)
Dropout	416 (18.91%)	581 (20.34%)	393 (14.72%)	938 (11.00%)
Referred	252 (11.45%)	161 (5.64%)	90 (3.37%)	289 (3.39%)
Service				
0	880 (40.00%)	1210 (42.35%)	1377 (51.57%)	3560 (41.74%)
1	532 (24.18%)	633 (22.16%)	567 (21.24%)	1670 (19.58%)
2	371 (16.86%)	563 (19.71%)	424 (15.88%)	1886 (22.11%)
3	417 (18.95%)	451 (15.79%)	302 (11.31%)	1414 (16.58%)

Note: Percentages were calculated using the available sample for each variable, after excluding missing values. The "Missing" row represents the number of missing values and was omitted if there was no missing data. [1] Case thresholds: PHQ9 ≥10, GAD7 ≥8. [2] 'Non-worker' included homemaker, carer, retired, student. [3] No negative responses were recorded, therefore the absence of any value was taken as a negative response rather than missing. [4] GAD = generalised anxiety disorder; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; MADD = mixed anxiety and depressive disorder; Panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; 'Other' included somatoform disorder, severe mental illness. [5] Calculated for patients who scored above the case threshold on either/both the PHQ9 or GAD7 at the start of treatment and had an observed score for their final session, otherwise coded as missing. Represents whether the patient reached recovery within the 10 treatment sessions modelled; if a patient received more sessions and then recovered, they would appear unrecovered here.

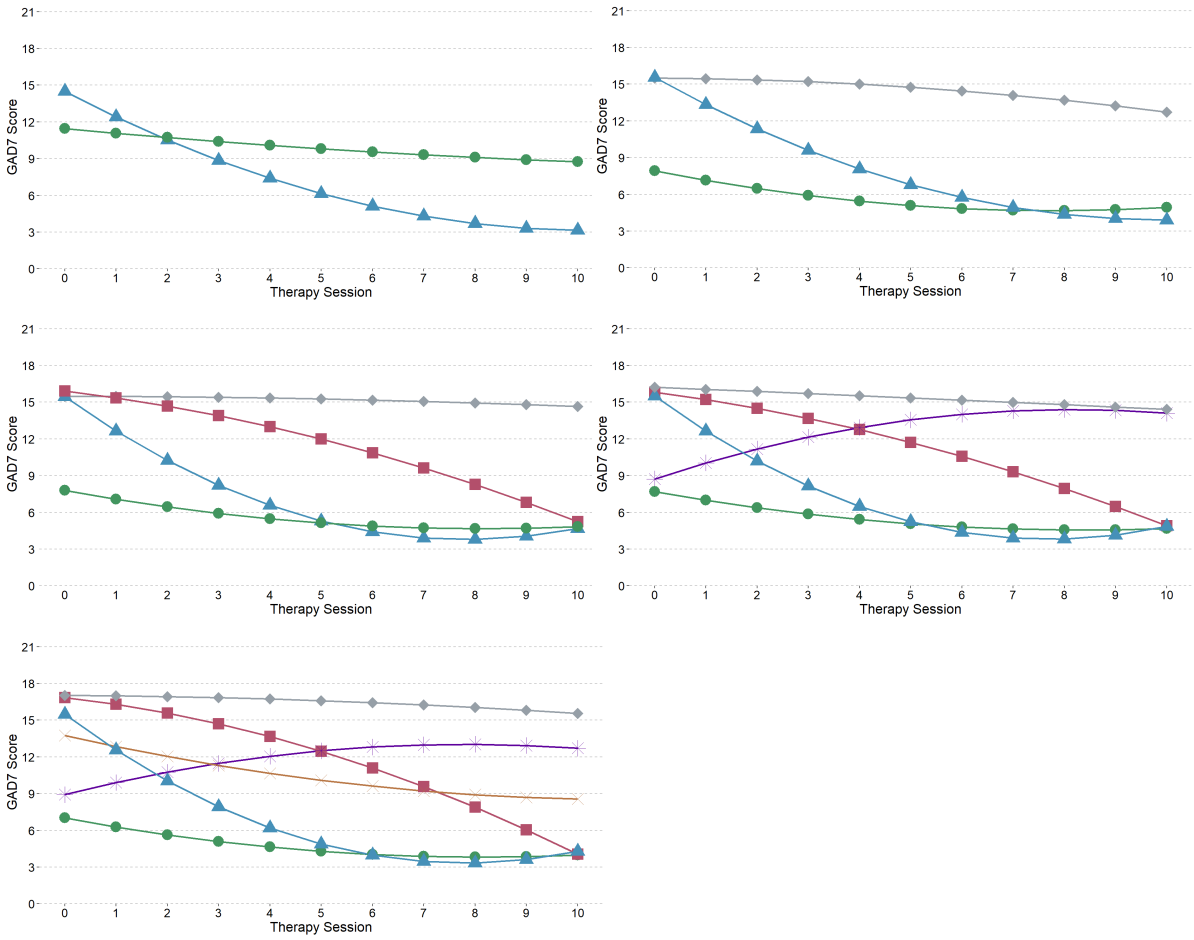
Supplementary Information 7. Selection of anxiety symptoms growth mixture model

This section describes the model selection for the growth mixture model of anxiety symptoms, including fit indices (table in the main text) and trajectory plots. As in the depression model, the VLMR LRT failed to reach non-significance and the information criteria decreased up to six classes. A four-class model was supported by the BIC elbow-plot and was chosen as the best-fitting model. Entropy in the four-class model was low (0.59).



Elbow plot of Bayesian Information Criterion values for growth mixture models of anxiety symptoms

Two- to six-class growth mixture models of anxiety symptoms during psychological therapy



Supplementary Information 8. Descriptives of the selected four-class growth mixture model of anxiety symptoms (GAD7)

Descriptives of the growth factors for the four-class growth mixture model of anxiety symptoms

Class	Parameter	Factor	Estimate	SE	Est SE	*p*-value
Moderate-severe plateau	Means	Intercept	15.475	0.094	164.421	0
		Linear	-0.003	0.033	-0.08	0.936
		Quadratic	-0.008	0.004	-2.263	0.024
	Variances	Intercept	6.811	0.125	54.584	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Moderate-severe, gradual improvement	Means	Intercept	15.908	0.12	132.867	0
		Linear	-0.504	0.078	-6.436	0
		Quadratic	-0.056	0.008	-7.128	0
	Variances	Intercept	6.811	0.125	54.584	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Moderate-severe, fast improvement	Means	Intercept	15.436	0.114	134.932	0
		Linear	-2.977	0.06	-49.832	0
		Quadratic	0.19	0.007	28.151	0
	Variances	Intercept	6.811	0.125	54.584	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999
Mild, small improvement	Means	Intercept	7.797	0.095	82.011	0
		Linear	-0.765	0.031	-24.642	0
		Quadratic	0.047	0.003	15.784	0
	Variances	Intercept	6.811	0.125	54.584	0
		Linear	0	0	999	999
		Quadratic	0	0	999	999

Model estimated anxiety symptom scores (GAD7) per therapy session and class

	Session										
Class	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Moderate-severe plateau	15.47	15.46	15.44	15.39	15.34	15.26	15.17	15.06	14.94	14.80	14.65
Moderate-severe, gradual improvement	15.91	15.35	14.68	13.89	12.99	11.98	10.86	9.63	8.28	6.82	5.25
Moderate-severe, fast improvement	15.44	12.65	10.24	8.22	6.57	5.30	4.42	3.91	3.79	4.05	4.68
Mild, small improvement	7.80	7.08	6.45	5.92	5.49	5.14	4.89	4.73	4.67	4.70	4.82

Descriptives of patients in each class of the anxiety symptoms model (assigned to their most likely trajectory class)

	Moderate-severe plateau (N=4035)	Moderate-severe, gradual improvement (N=1931)	Moderate-severe, fast improvement (N=3537)	Mild, small improvement (N=6754)
Anxiety symptoms (GAD7)				
Mean (SD)	15.74 (3.87)	16.31 (3.27)	16.07 (2.91)	7.67 (3.61)
Range	0.00 - 21.00	0.00 - 21.00	7.00 - 21.00	0.00 - 20.00
Missing	47	21	18	67
Depression symptoms (PHQ9)				
Mean (SD)	17.92 (5.33)	17.37 (5.21)	16.02 (5.26)	9.59 (5.02)
Range	0.00 - 27.00	0.00 - 27.00	0.00 - 27.00	0.00 - 27.00
Missing	47	21	18	66
Case on PHQ9 and/or GAD7 [1]				
Yes	3934 (98.65%)	1895 (99.21%)	3515 (99.89%)	4391 (65.67%)
No	54 (1.35%)	15 (0.79%)	4 (0.11%)	2295 (34.33%)
Missing	47	21	18	68
Functional impairment score (WSAS)				
Mean (SD)	22.21 (9.39)	21.06 (8.60)	18.84 (8.53)	13.44 (7.91)
Range	0.00 - 40.00	0.00 - 40.00	0.00 - 40.00	0.00 - 40.00
Missing	1405	653	1311	2077
Psychotropic medication				
Prescribed	1831 (47.79%)	766 (41.38%)	1186 (34.96%)	1762 (27.33%)
Not prescribed	2000 (52.21%)	1085 (58.62%)	2206 (65.04%)	4686 (72.67%)
Missing	204	80	145	306
Employment status				
Employed	1958 (49.90%)	1162 (61.55%)	2309 (66.75%)	4604 (70.23%)
Unemployed	1413 (36.01%)	468 (24.79%)	678 (19.60%)	1013 (15.45%)
Non-worker [2]	553 (14.09%)	258 (13.67%)	472 (13.65%)	939 (14.32%)
Missing	111	43	78	198
Disability [3]				
Yes	570 (14.13%)	190 (9.84%)	289 (8.17%)	526 (7.79%)
No	3465 (85.87%)	1741 (90.16%)	3248 (91.83%)	6228 (92.21%)
Ethnicity				
White	2225 (58.55%)	1149 (62.75%)	2066 (61.40%)	4349 (68.11%)
Black	806 (21.21%)	325 (17.75%)	720 (21.40%)	1112 (17.42%)
Mixed	275 (7.24%)	138 (7.54%)	271 (8.05%)	427 (6.69%)
Asian	320 (8.42%)	145 (7.92%)	198 (5.88%)	298 (4.67%)
Other	174 (4.58%)	74 (4.04%)	110 (3.27%)	199 (3.12%)
Missing	235	100	172	369
Problem descriptor [4]				
Depression	1705 (47.00%)	804 (46.02%)	1481 (45.75%)	2712 (44.90%)

GAD	323 (8.90%)	172 (9.85%)	378 (11.68%)	520 (8.61%)
Other	262 (7.22%)	106 (6.07%)	296 (9.14%)	759 (12.57%)
MADD	316 (8.71%)	145 (8.30%)	269 (8.31%)	399 (6.61%)
Panic/phobia	206 (5.68%)	117 (6.70%)	202 (6.24%)	478 (7.91%)
Adjustment disorder	265 (7.30%)	112 (6.41%)	278 (8.59%)	665 (11.01%)
PTSD	398 (10.97%)	195 (11.16%)	227 (7.01%)	312 (5.17%)
OCD	153 (4.22%)	96 (5.50%)	106 (3.27%)	195 (3.23%)
Missing	407	184	300	714
Age (years)				
Mean (SD)	38.03 (13.14)	36.87 (12.95)	36.61 (13.07)	37.96 (13.71)
Range	16.00 - 90.00	17.00 - 91.00	16.00 - 89.00	16.00 - 94.00
Gender				
Female	2745 (68.08%)	1337 (69.35%)	2440 (69.04%)	4456 (66.05%)
Male	1287 (31.92%)	591 (30.65%)	1094 (30.96%)	2290 (33.95%)
Missing	3	3	3	8
Number of sessions (including baseline assessment)				
Mean (SD)	6.86 (2.93)	8.15 (2.34)	6.36 (2.43)	6.64 (2.60)
Range	3.00 - 11.00	3.00 - 11.00	3.00 - 11.00	3.00 - 11.00
Recovered [5]				
Yes	40 (1.03%)	659 (34.92%)	2414 (69.11%)	2790 (64.01%)
No	3854 (98.97%)	1228 (65.08%)	1079 (30.89%)	1569 (35.99%)
Missing	141	44	44	2395
Reason for end of treatment				
Discharged	2745 (68.03%)	1623 (84.05%)	2940 (83.12%)	5829 (86.30%)
Dropout	917 (22.73%)	222 (11.50%)	480 (13.57%)	709 (10.50%)
Referred	373 (9.24%)	86 (4.45%)	117 (3.31%)	216 (3.20%)
Service				
0	1645 (40.77%)	829 (42.93%)	1683 (47.58%)	2870 (42.49%)
1	891 (22.08%)	443 (22.94%)	759 (21.46%)	1309 (19.38%)
2	780 (19.33%)	405 (20.97%)	654 (18.49%)	1405 (20.80%)
3	719 (17.82%)	254 (13.15%)	441 (12.47%)	1170 (17.32%)

Note: Percentages were calculated using the available sample for each variable, after excluding missing values. The "Missing" row represents the number of missing values and was omitted if there was no missing data. [1] Case thresholds: PHQ9 ≥10, GAD7 ≥8. [2] 'Non-worker' included homemaker, carer, retired, student. [3] No negative responses were recorded, therefore the absence of any value was taken as a negative response rather than missing. [4] GAD = generalised anxiety disorder; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; MADD = mixed anxiety and depressive disorder; Panic/phobia = panic disorder, agoraphobia, social phobia, specific phobia; 'Other' included somatoform disorder, severe mental illness. [5] Calculated for patients who scored above the case threshold on either/both the PHQ9 or GAD7 at the start of treatment and had an observed score for their final session, otherwise coded as missing. Represents whether the patient reached recovery within the 10 treatment sessions modelled; if a patient received more sessions and then recovered, they would appear unrecovered here.

Supplementary Table 5. Multinomial regression output for conditional four-class growth mixture model of depression symptoms (PHQ9)

Reference class: Moderate-severe plateau. Covariate: Service

Class	Baseline Variable	OR	Lower CI	Upper CI	*p*-value	Statistic	df
Moderate-severe, gradual improvement	(Intercept)	2.04	1.49	2.79	0.00	4.46	11799
Moderate-severe, gradual improvement	Anxiety symptoms (GAD7)	1.01	1.00	1.03	0.14	1.47	6716
Moderate-severe, gradual improvement	Functional impairment score (WSAS)	1.00	0.99	1.01	0.79	-0.27	335
Moderate-severe, gradual improvement	Psychotropic medication (Prescribed)	0.90	0.80	1.02	0.10	-1.66	3287
Moderate-severe, gradual improvement	Employment status (Unemployed)	0.68	0.60	0.78	0.00	-5.64	11374
Moderate-severe, gradual improvement	Employment status (Non-worker)	0.84	0.70	1.00	0.05	-1.95	6152
Moderate-severe, gradual improvement	Disability (Yes)	0.79	0.67	0.94	0.01	-2.73	16142
Moderate-severe, gradual improvement	Problem descriptor (GAD)	0.82	0.65	1.05	0.11	-1.59	1353
Moderate-severe, gradual improvement	Problem descriptor (Other)	0.89	0.70	1.14	0.37	-0.89	2924
Moderate-severe, gradual improvement	Problem descriptor (MADD)	0.82	0.65	1.03	0.08	-1.75	1774
Moderate-severe, gradual improvement	Problem descriptor (Panic/phobia)	0.79	0.60	1.04	0.10	-1.67	2011
Moderate-severe, gradual improvement	Problem descriptor (Adjustment)	0.77	0.60	0.98	0.04	-2.09	1441
Moderate-severe, gradual improvement	Problem descriptor (PTSD)	0.80	0.66	0.97	0.02	-2.28	3381
Moderate-severe, gradual improvement	Problem descriptor (OCD)	0.79	0.56	1.10	0.17	-1.39	2861
Moderate-severe, gradual improvement	Ethnicity (Black)	0.97	0.83	1.12	0.64	-0.47	5708
Moderate-severe, gradual improvement	Ethnicity (Mixed)	0.93	0.74	1.16	0.50	-0.67	4232
Moderate-severe, gradual improvement	Ethnicity (Asian)	0.95	0.77	1.17	0.63	-0.49	5916
Moderate-severe, gradual improvement	Ethnicity (Other)	0.95	0.72	1.25	0.71	-0.37	7440
Moderate-severe, gradual improvement	Age (10 Years)	1.00	0.99	1.00	0.03	-2.23	15899
Moderate-severe, gradual improvement	Gender (Male)	1.00	0.88	1.13	0.95	-0.06	16031
Moderate-severe, fast improvement	(Intercept)	3.65	2.64	5.04	0.00	7.87	5720
Moderate-severe, fast improvement	Anxiety symptoms (GAD7)	1.01	0.99	1.02	0.33	0.97	6726

Moderate-severe, fast improvement	Functional impairment score (WSAS)	0.98	0.97	0.99	0.00	-3.85	141
Moderate-severe, fast improvement	Psychotropic medication (Prescribed)	0.73	0.65	0.83	0.00	-5.03	7486
Moderate-severe, fast improvement	Employment status (Unemployed)	0.48	0.42	0.55	0.00	-10.34	11210
Moderate-severe, fast improvement	Employment status (Non-worker)	0.76	0.64	0.92	0.00	-2.93	8039
Moderate-severe, fast improvement	Disability (Yes)	0.68	0.56	0.81	0.00	-4.24	15949
Moderate-severe, fast improvement	Problem descriptor (GAD)	0.85	0.66	1.09	0.20	-1.30	936
Moderate-severe, fast improvement	Problem descriptor (Other)	0.99	0.77	1.26	0.90	-0.12	2352
Moderate-severe, fast improvement	Problem descriptor (MADD)	0.97	0.76	1.23	0.77	-0.29	677
Moderate-severe, fast improvement	Problem descriptor (Panic/phobia)	0.78	0.58	1.06	0.11	-1.60	614
Moderate-severe, fast improvement	Problem descriptor (Adjustment)	0.99	0.78	1.26	0.94	-0.08	1538
Moderate-severe, fast improvement	Problem descriptor (PTSD)	0.56	0.45	0.70	0.00	-5.16	1020
Moderate-severe, fast improvement	Problem descriptor (OCD)	0.52	0.35	0.75	0.00	-3.42	2213
Moderate-severe, fast improvement	Ethnicity (Black)	1.18	1.02	1.37	0.03	2.19	11449
Moderate-severe, fast improvement	Ethnicity (Mixed)	1.01	0.80	1.26	0.94	0.07	5018
Moderate-severe, fast improvement	Ethnicity (Asian)	0.73	0.57	0.92	0.01	-2.68	3641
Moderate-severe, fast improvement	Ethnicity (Other)	0.73	0.53	0.99	0.04	-2.04	2759
Moderate-severe, fast improvement	Age (10 Years)	1.00	0.99	1.00	0.43	-0.79	14694
Moderate-severe, fast improvement	Gender (Male)	1.05	0.92	1.19	0.49	0.70	15882
Mild, small improvement	(Intercept)	491.73	364.65	663.09	0.00	40.64	3594
Mild, small improvement	Anxiety symptoms (GAD7)	0.79	0.77	0.80	0.00	-33.74	4554
Mild, small improvement	Functional impairment score (WSAS)	0.95	0.94	0.95	0.00	-13.68	219
Mild, small improvement	Psychotropic medication (Prescribed)	0.51	0.45	0.57	0.00	-11.39	3767
Mild, small improvement	Employment status (Unemployed)	0.42	0.37	0.48	0.00	-12.44	4064
Mild, small improvement	Employment status (Non-worker)	0.70	0.59	0.83	0.00	-4.21	4931

Mild, small improvement	Disability (Yes)	0.65	0.55	0.77	0.00	-4.86	12042
Mild, small improvement	Problem descriptor (GAD)	2.48	1.99	3.10	0.00	8.05	736
Mild, small improvement	Problem descriptor (Other)	1.88	1.49	2.36	0.00	5.40	1700
Mild, small improvement	Problem descriptor (MADD)	1.23	0.98	1.56	0.08	1.75	427
Mild, small improvement	Problem descriptor (Panic/phobia)	2.17	1.69	2.78	0.00	6.06	1255
Mild, small improvement	Problem descriptor (Adjustment)	1.75	1.40	2.18	0.00	4.89	1320
Mild, small improvement	Problem descriptor (PTSD)	0.86	0.70	1.05	0.14	-1.50	1639
Mild, small improvement	Problem descriptor (OCD)	1.92	1.40	2.63	0.00	4.03	1184
Mild, small improvement	Ethnicity (Black)	0.71	0.61	0.82	0.00	-4.53	2532
Mild, small improvement	Ethnicity (Mixed)	0.77	0.62	0.96	0.02	-2.29	1620
Mild, small improvement	Ethnicity (Asian)	0.66	0.53	0.82	0.00	-3.68	2032
Mild, small improvement	Ethnicity (Other)	0.61	0.45	0.81	0.00	-3.39	1494
Mild, small improvement	Age (10 Years)	0.99	0.99	1.00	0.00	-2.92	11068
Mild, small improvement	Gender (Male)	0.92	0.81	1.03	0.14	-1.47	10138

Supplementary Table 6. Multinomial regression output for conditional four-class growth mixture model of anxiety symptoms (GAD7)

Reference class: Moderate-severe plateau. Covariate: Service

Class	Baseline Variable	OR	Lower CI	Upper CI	*p*-value	Statistic	df
Moderate-severe, gradual improvement	(Intercept)	0.81	0.61	1.09	0.16	-1.41	3800
Moderate-severe, gradual improvement	Depression symptoms (PHQ9)	1.00	0.99	1.01	0.78	-0.29	2605
Moderate-severe, gradual improvement	Functional impairment score (WSAS)	1.00	0.99	1.00	0.26	-1.13	157
Moderate-severe, gradual improvement	Psychotropic medication (Prescribed)	0.84	0.75	0.94	0.00	-2.93	11017
Moderate-severe, gradual improvement	Employment status (Unemployed)	0.63	0.55	0.72	0.00	-6.59	9484
Moderate-severe, gradual improvement	Employment status (Non-worker)	0.83	0.70	0.98	0.03	-2.16	4735
Moderate-severe, gradual improvement	Disability (Yes)	0.79	0.66	0.94	0.01	-2.58	16015
Moderate-severe, gradual improvement	Problem descriptor (GAD)	1.03	0.83	1.27	0.79	0.26	1614
Moderate-severe, gradual improvement	Problem descriptor (Other)	0.83	0.65	1.06	0.13	-1.50	4721
Moderate-severe, gradual improvement	Problem descriptor (MADD)	0.96	0.77	1.19	0.72	-0.36	3979
Moderate-severe, gradual improvement	Problem descriptor (Panic/phobia)	1.17	0.92	1.50	0.20	1.27	2514
Moderate-severe, gradual improvement	Problem descriptor (Adjustment)	0.90	0.70	1.15	0.38	-0.87	1288
Moderate-severe, gradual improvement	Problem descriptor (PTSD)	1.10	0.91	1.34	0.32	0.99	2268
Moderate-severe, gradual improvement	Problem descriptor (OCD)	1.17	0.88	1.56	0.27	1.10	1708
Moderate-severe, gradual improvement	Ethnicity (Black)	0.84	0.72	0.97	0.02	-2.31	2696
Moderate-severe, gradual improvement	Ethnicity (Mixed)	0.97	0.78	1.20	0.75	-0.32	7478
Moderate-severe, gradual improvement	Ethnicity (Asian)	0.94	0.75	1.16	0.54	-0.62	4219
Moderate-severe, gradual improvement	Ethnicity (Other)	0.91	0.69	1.21	0.52	-0.64	3134
Moderate-severe, gradual improvement	Age (10 Years)	1.00	1.00	1.00	0.93	-0.09	15735
Moderate-severe, gradual improvement	Gender (Male)	0.99	0.88	1.11	0.84	-0.21	16067
Moderate-severe, fast improvement	(Intercept)	4.11	3.23	5.22	0.00	11.54	8323
Moderate-severe, fast improvement	Depression symptoms (PHQ9)	0.96	0.95	0.97	0.00	-6.82	2011

Moderate-severe, fast improvement	Functional impairment score (WSAS)	0.98	0.98	0.99	0.00	-4.50	220
Moderate-severe, fast improvement	Psychotropic medication (Prescribed)	0.77	0.70	0.86	0.00	-4.95	4538
Moderate-severe, fast improvement	Employment status (Unemployed)	0.52	0.46	0.59	0.00	-10.67	7746
Moderate-severe, fast improvement	Employment status (Non-worker)	0.82	0.71	0.94	0.01	-2.79	9144
Moderate-severe, fast improvement	Disability (Yes)	0.73	0.62	0.86	0.00	-3.86	15890
Moderate-severe, fast improvement	Problem descriptor (GAD)	1.06	0.89	1.26	0.50	0.68	1946
Moderate-severe, fast improvement	Problem descriptor (Other)	1.10	0.91	1.32	0.34	0.96	5749
Moderate-severe, fast improvement	Problem descriptor (MADD)	1.01	0.84	1.21	0.94	0.08	2249
Moderate-severe, fast improvement	Problem descriptor (Panic/phobia)	1.00	0.81	1.25	0.97	0.04	1440
Moderate-severe, fast improvement	Problem descriptor (Adjustment)	1.10	0.91	1.33	0.34	0.97	1513
Moderate-severe, fast improvement	Problem descriptor (PTSD)	0.73	0.61	0.88	0.00	-3.40	3169
Moderate-severe, fast improvement	Problem descriptor (OCD)	0.61	0.47	0.80	0.00	-3.55	2714
Moderate-severe, fast improvement	Ethnicity (Black)	1.08	0.95	1.22	0.24	1.18	3553
Moderate-severe, fast improvement	Ethnicity (Mixed)	1.09	0.91	1.31	0.36	0.92	8159
Moderate-severe, fast improvement	Ethnicity (Asian)	0.80	0.66	0.97	0.02	-2.29	6895
Moderate-severe, fast improvement	Ethnicity (Other)	0.79	0.61	1.02	0.07	-1.82	2939
Moderate-severe, fast improvement	Age (10 Years)	1.00	1.00	1.00	0.37	-0.90	15427
Moderate-severe, fast improvement	Gender (Male)	1.07	0.97	1.19	0.19	1.32	15624
Mild, small improvement	(Intercept)	166.00	129.61	212.62	0.00	40.50	2622
Mild, small improvement	Depression symptoms (PHQ9)	0.76	0.75	0.77	0.00	-43.90	1287
Mild, small improvement	Functional impairment score (WSAS)	0.98	0.97	0.98	0.00	-6.53	189
Mild, small improvement	Psychotropic medication (Prescribed)	0.84	0.75	0.93	0.00	-3.34	4702
Mild, small improvement	Employment status (Unemployed)	0.61	0.54	0.69	0.00	-7.73	7037
Mild, small improvement	Employment status (Non-worker)	0.84	0.73	0.97	0.02	-2.33	10852

Mild, small improvement	Disability (Yes)	0.90	0.77	1.06	0.21	-1.25	14246
Mild, small improvement	Problem descriptor (GAD)	0.38	0.31	0.45	0.00	-10.59	1468
Mild, small improvement	Problem descriptor (Other)	0.82	0.68	0.99	0.04	-2.11	3400
Mild, small improvement	Problem descriptor (MADD)	0.55	0.45	0.66	0.00	-6.09	2099
Mild, small improvement	Problem descriptor (Panic/phobia)	0.58	0.47	0.73	0.00	-4.83	907
Mild, small improvement	Problem descriptor (Adjustment)	0.74	0.61	0.89	0.00	-3.21	2384
Mild, small improvement	Problem descriptor (PTSD)	0.50	0.41	0.61	0.00	-6.94	3582
Mild, small improvement	Problem descriptor (OCD)	0.23	0.17	0.30	0.00	-10.52	1232
Mild, small improvement	Ethnicity (Black)	0.93	0.82	1.06	0.27	-1.10	2686
Mild, small improvement	Ethnicity (Mixed)	0.99	0.82	1.20	0.91	-0.11	6911
Mild, small improvement	Ethnicity (Asian)	0.73	0.59	0.89	0.00	-3.10	5133
Mild, small improvement	Ethnicity (Other)	0.81	0.63	1.04	0.10	-1.63	7303
Mild, small improvement	Age (10 Years)	1.00	1.00	1.01	0.06	1.86	11239
Mild, small improvement	Gender (Male)	1.24	1.12	1.38	0.00	4.16	13978

References

Asparouhov, T., & Muthen, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, 14(22), 1–17.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods*, 9(1), 3–29.

Clark, S. L., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis, 1–55.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.

Jung, T., & Wickrama, K. A. S. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, 2, 302–317.

Meyer, J. P., & Morin, A. J. S. (2016). A person-centered approach to commitment research: Theory, research, and methodology: Person-Centered Commitment Research. *Journal of Organizational Behavior*, 37(4), 584–612.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1), 81–117.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, 6, 109–138.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569.

Petras, H., & Masyn, K. (2010). General Growth Mixture Analysis with Antecedents and Consequences of Change. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 69–100). Springer, New York, NY.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. Retrieved 17 March 2021 from

van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 451–467.

Wickrama, K. K., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide*. Routledge.

Appendix D - Supplementary materials for Chapter 5

Supplementary materials for "Trajectories of depression symptoms, anxiety symptoms and functional impairment during internet-enabled cognitive-behavioural therapy"

Acknowledgements and declaration of interest

This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or King's College London. Megan Skelton was supported by an NIHR Maudsley Biomedical Research Centre studentship. Christopher Rayner was supported by a grant from Fondation Peters to Thalia C Eley and Gerome Breen. Alicia J Peel was supported by an ESRC studentship. Ana Catarino and Stephanie Brown are employees of ieso, all other authors declare no conflict of interest.

CRediT authorship contributions

Megan Skelton: Conceptualisation, methodology, formal analysis, writing – original draft, writing – review & editing, visualisation. Ana Catarino and Stephanie Brown: Conceptualisation, resources (patient data), writing – review & editing. Ewan Carr, Molly R Davies, Alicia J Peel and Christopher Rayner: Conceptualisation, writing – review & editing. Gerome Breen and Thalia C Eley: Conceptualisation, writing – review & editing, supervision.

Data sharing policy

This study used anonymised patient data from ieso. Data can be accessed by researchers with legitimate academic interest and the correct permissions in agreement with ieso; please contact info@iesogroup.com for further information.

Supplementary Information 1. Inclusion criteria for treatment with ieso

ieso treats individuals who are at least 18 years of age, registered with a GP in an area where the service is commissioned and have access to a suitable device and internet connection. Individuals are excluded (and signposted to appropriate services) if they express suicidal intent, have active psychosis or mania symptoms, a severe learning disability or cognitive impairment, a severe personality disorder, or are unable to read or write English on a device due to low-level literacy, language barriers, or visuomotor disability without access to assistive technology. Therapists have access to patient's self-report data and use it to monitor progress throughout therapy. They conduct risk assessments and patients at high risk at any point are signposted to more appropriate services.

Supplementary Information 2. Exclusions applied to retain a sample of patients suitable for trajectory analysis

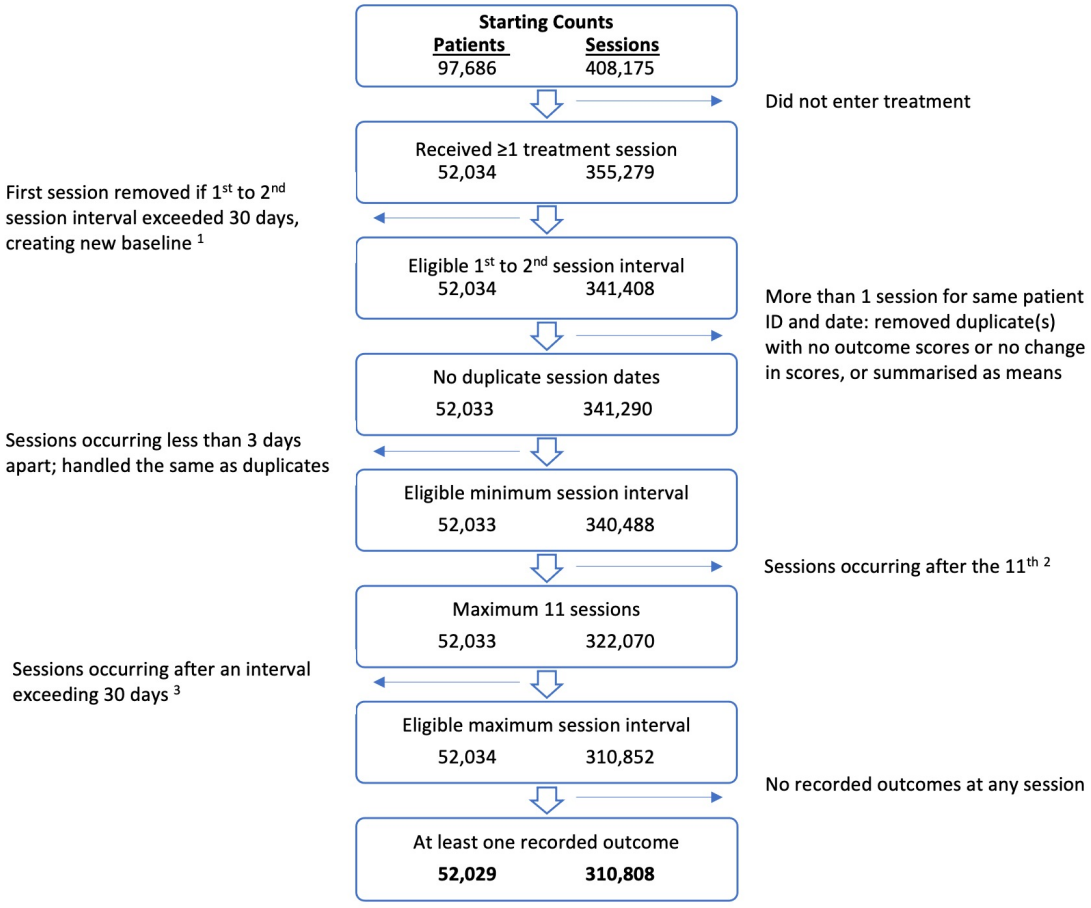


Figure *1*: Flowchart of exclusions applied to original sample to retain sample for analysis

1 The first session in IAPT is often an assessment or triage session with treatment beginning formally at the second session. If the first to second session interval exceeded 30 days, the first session was not considered to represent an accurate baseline assessment and was therefore removed.

2 A baseline assessment and ten therapy sessions. Selected for sufficient complete data across timepoints for all outcomes (using Mplus covariance coverage minimum 0.10), and as modelling a higher number than most patients received could make optimal solutions invalid and unstable [28]. Table I shows descriptives before and after this limitation.

3 The analytical method (growth mixture modelling) assumes time intervals are invariant across patients. Although therapy sessions are intended to occur on a weekly basis, the naturalistic nature of the data introduced substantial variance. In an attempt to limit this we excluded sessions occurring after a 30 day interval. Table II for session interval descriptives with and without this filtering.

Table *I*: Number of timepoints recorded for patients in the analysis sample

The mean number of timepoints before limiting the maximum was 6.5 (SD = 3.9, range = 1 - 39) and following limiting to 11 maximum it was 6.0 (SD = 3.1, range = 1 - 11).

Timepoint	Frequency	Proportion (%)	Cumulative Proportion (%)
1 (baseline)	2726	5.24	5.24
2	6348	12.20	17.44
3	5368	10.32	27.76
4	5129	9.86	37.62
5	5272	10.13	47.75
6	5048	9.70	57.45
7	5188	9.97	67.42
8	3766	7.24	74.66
9	3262	6.27	80.93
10	2826	5.43	86.36
11 (10 therapy sessions)	7096	13.64	100.00

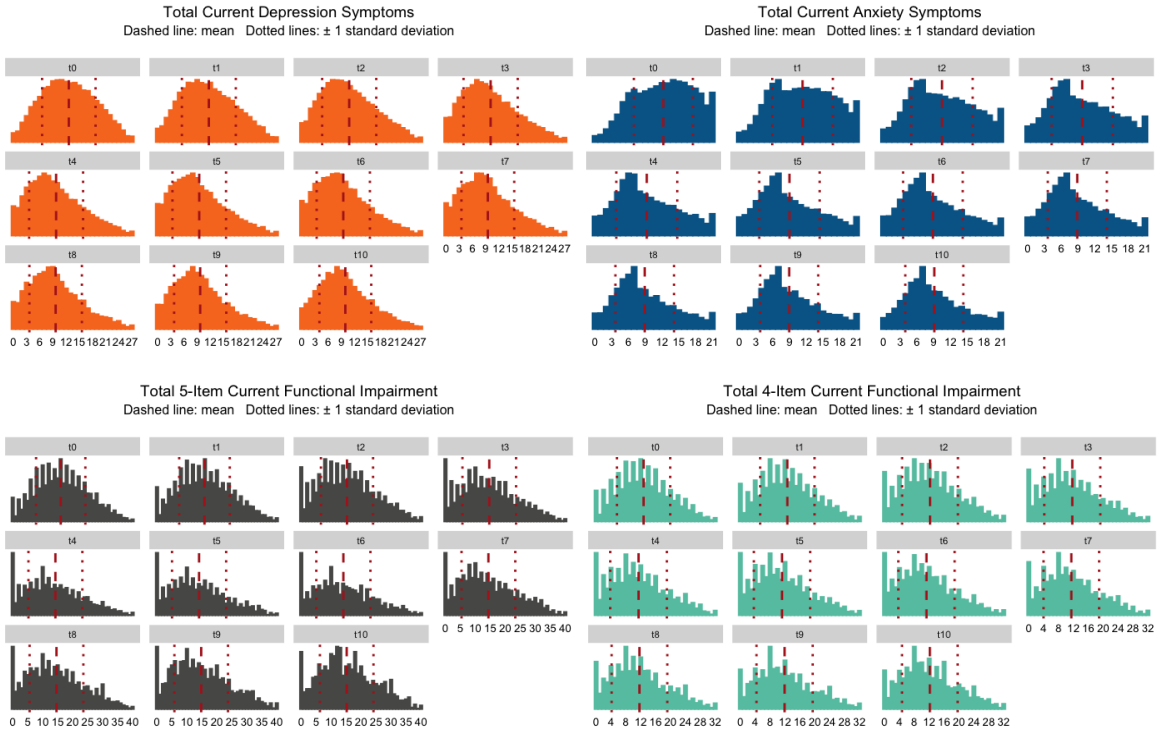
Table *II*: Descriptives of days between therapy sessions

Prior to filtering out sessions after a long (30 day) interval, the mean interval was 13 days (SD = 10, range = 0 - 527) and after filtering it was 11 days (SD = 6, range = 3 - 30).

	Therapy Session										
Dataset	0	1	2	3	4	5	6	7	8	9	10
Unfiltered	0.00 (0.00)	24.71 (14.16)	9.62 (6.35)	9.92 (6.46)	10.19 (7.06)	10.43 (7.00)	10.73 (7.32)	10.85 (7.22)	10.96 (7.18)	11.22 (7.57)	11.49 (7.87)
Filtered	0.00 (0.00)	16.41 (7.13)	9.17 (4.56)	9.43 (4.73)	9.61 (4.85)	9.72 (4.84)	9.94 (5.02)	10.06 (5.06)	10.12 (5.09)	10.36 (5.24)	10.49 (5.36)

Note: Mean (standard deviation) days between column-specified session and previous session. For patients who had entered treatment descriptives are provided *prior* to removing sessions occurring after intervals exceeding 30 days (unfiltered - upper row) and *following* this (filtered - lower row). The filtered dataset was used for analysis.

Supplementary Figure 1. Histograms of observed depression symptoms (PHQ9; top), anxiety symptoms (GAD7; middle) and functional impairment (WSAS & 4-item WSAS; bottom) at each session of psychological therapy



Note: Histograms for total symptom scores measured across the 11 timepoints (baseline and 10 therapy sessions). Only observed scores for patients who were in treatment at each timepoint were included; see Supplementary Table 1 for Ns

Supplementary Table 1. Proportion of observed scores for depression symptoms and anxiety symptoms (upper table), functional impairment and WSAS 4-item (lower table) across psychological therapy sessions, in patients in treatment at that time

Therapy session	Number of patients in treatment	Proportion (%) with PHQ9 score	Proportion (%) with GAD7 score
0 (baseline)	52029	98.97	98.93
1	49303	97.44	97.28
2	42955	96.93	96.75
3	37587	96.96	96.76
4	32458	96.97	96.75
5	27186	96.86	96.69
6	22138	96.77	96.67
7	16950	96.97	96.83
8	13184	96.96	96.81
9	9922	97.06	96.92
10	7096	97.29	97.11

Therapy session	Number of WSAS subsample patients in treatment	Proportion (%) with WSAS score	Number of WSAS 4-item subsample patients in treatment	Proportion (%) with WSAS 4-item score
0 (baseline)	32168	99.50	19293	99.63
1	30375	96.96	18606	97.61
2	26141	96.06	16633	96.61
3	22809	96.20	14640	96.37
4	19677	96.08	12679	96.26
5	16472	96.08	10630	96.13
6	13379	96.36	8688	95.90
7	10170	96.33	6722	96.30
8	7904	96.23	5242	96.36
9	5937	96.38	3958	96.46
10	4228	96.31	2852	96.74

Supplementary Table 2. Baseline descriptives of patient sample for analysis who received internet-enabled cognitive-behavioural therapy (N = 52,029) and subset with functional impairment score (N = 32,168)

Variable	Mean (SD); range or Count (Proportion %)	
	Full sample (N = 52,029)	WSAS subset (N = 32,168)
Age (years)		
Mean (SD); Range	34.3 (12.3); 18 - 94	33.9 (10.9); 18 - 92
Missing	539 (1.0%)	181 (0.6%)
Gender		
Female	38,246 (73.5%)	23,144 (71.9%)
Male	13,572 (26.1%)	8,910 (27.7%)
Missing	211 (0.4%)	114 (0.4%)
Depression Symptoms (PHQ9)		
Mean (SD); Range	12.6 (6.1); 0 - 27	12.4 (6.1); 0 - 27
Missing	537 (1.0%)	133 (0.4%)
Anxiety Symptoms (GAD7)		
Mean (SD); Range	12.2 (5.3); 0 - 21	12.0 (5.3); 0 - 21
Missing	556 (1.1%)	138 (0.4%)
Case on PHQ9 and/or GAD7 [1]		
No	8,527 (16.4%)	5,557 (17.3%)
Yes	42,984 (82.6%)	26,487 (82.3%)
Missing	518 (1.0%)	124 (0.4%)
Functional Impairment (WSAS)		
Mean (SD); Range	16.1 (8.3); 0 - 40	16.0 (8.2); 0 - 40
Missing	12,545 (24.1%)	161 (0.5%)
Functional Impairment WSAS Home Management		
Mean (SD); Range	2.9 (2.1); 0 - 8	2.9 (2.1); 0 - 8
Missing	800 (1.5%)	161 (0.5%)
Functional Impairment WSAS Social Leisure		
Mean (SD); Range	3.8 (2.3); 0 - 8	3.8 (2.3); 0 - 8
Missing	800 (1.5%)	161 (0.5%)
Functional Impairment WSAS Private Leisure		
Mean (SD); Range	2.8 (2.2); 0 - 8	2.8 (2.2); 0 - 8
Missing	800 (1.5%)	161 (0.5%)
Functional Impairment WSAS Relationships		
Mean (SD); Range	3.2 (2.2); 0 - 8	3.2 (2.2); 0 - 8
Missing	800 (1.5%)	161 (0.5%)
Functional Impairment WSAS Work		
Mean (SD); Range	3.3 (2.3); 0 - 8	3.3 (2.3); 0 - 8
Missing	12,530 (24.1%)	161 (0.5%)
Agoraphobia Item		
Mean (SD); Range	2.6 (2.5); 0 - 8	2.5 (2.5); 0 - 8
Missing	814 (1.6%)	210 (0.7%)
Social Phobia Item		

Mean (SD); Range	3.3 (2.4); 0 - 8	3.2 (2.3); 0 - 8
Missing	814 (1.6%)	210 (0.7%)
Specific Phobia Item		
Mean (SD); Range	2.5 (2.5); 0 - 8	2.4 (2.5); 0 - 8
Missing	814 (1.6%)	210 (0.7%)
Diagnosis [2]		
Depression	21,296 (40.9%)	12,962 (40.3%)
GAD	14,203 (27.3%)	9,072 (28.2%)
Other Anxiety	10,184 (19.6%)	6,303 (19.6%)
OCD	2,509 (4.8%)	1,559 (4.8%)
PTSD	2,107 (4.0%)	1,192 (3.7%)
Other	1,677 (3.2%)	1,048 (3.3%)
Missing	53 (0.1%)	32 (0.1%)
Prescribed Psychotropic Medication		
No	26,469 (50.9%)	16,742 (52.0%)
Yes	23,996 (46.1%)	14,805 (46.0%)
Missing	1,564 (3.0%)	621 (1.9%)
Ethnicity		
White	34,839 (67.0%)	21,947 (68.2%)
Minoritised Ethnic Groups	3,731 (7.2%)	1,932 (6.0%)
Missing	13,459 (25.9%)	8,289 (25.8%)
Disability Reported		
No	21,191 (40.7%)	13,699 (42.6%)
Yes	4,731 (9.1%)	2,335 (7.3%)
Missing	26,107 (50.2%)	16,134 (50.2%)
Employment Status [3]		
Employed	34,518 (66.3%)	26,537 (82.5%)
Non-worker	8,543 (16.4%)	1,994 (6.2%)
Unemployed	6,061 (11.6%)	2,345 (7.3%)
Missing	2,907 (5.6%)	1,292 (4.0%)
Number of Sessions		
Mean (SD); Range	6.0 (3.1); 1 - 11	5.9 (3.2); 1 - 11
Missing	0 (0.0%)	0 (0.0%)
Recovered (higher 'yes' if include more than the 10 treatment sessions modelled) [4]		
No	23,081 (44.4%)	14,070 (43.7%)
Yes	17,758 (34.1%)	11,060 (34.4%)
Missing (includes not a case at start of treatment)	11,190 (21.5%)	7,038 (21.9%)
Functional Impairment WSAS Model		
Four item only	19,293 (37.1%)	0 (0.0%)
All five items	32,168 (61.8%)	32,168 (100.0%)
Missing	568 (1.1%)	0 (0.0%)

Note: Only the baseline values are presented for variables that were measured at multiple timepoints. The WSAS subset was analysed for that model as they had complete scores for the measure. [1] Case thresholds were PHQ9 ≥10, GAD7 ≥8. [2] IAPT 'problem descriptor': GAD = generalised anxiety disorder; Other Anxiety = panic disorder, agoraphobia, social phobia, specific phobia; hypochondriacal, unspecified anxiety; PTSD = post-traumatic stress disorder; OCD = obsessive-compulsive disorder; 'Other' included adjustment disorders and eating disorders. [3] 'Non-worker' included homemaker, retired, student. [4] Only calculated for patients who

scored above case thresholds on either the PHQ9 or GAD7 at the start of treatment and had an observed score for their final session, otherwise coded as missing. Represents whether the patient reached recovery within the 10 treatment sessions modelled; if they received more sessions and then recovered, they would appear unrecovered here.

Supplementary Table 3. Comparison of baseline characteristics of patients who were excluded from analysis (n = 45,657) to patients included in the analysis (n = 52,029)

Variable	Excluded from analysis (N = 45,657)	Included in analysis (N = 52,029)	*p*-value	Effect size
Started Treatment	5 (0.0%)	52,029 (100.0%)	< 0.001	1.00
Age (years)			< 0.001	0.05
Mean (SD)	34.9 (13.2)	34.3 (12.3)		
Missing	480	539		
Gender			< 0.001	0.02
Female	32,618 (72.0%)	38,246 (73.8%)		
Male	12,708 (28.0%)	13,572 (26.2%)		
Missing	331	211		
Depression Symptoms (PHQ9)			< 0.001	0.29
Mean (SD)	14.4 (6.3)	12.6 (6.1)		
Missing	21,278	537		
Anxiety Symptoms (GAD7)			< 0.001	0.15
Mean (SD)	13.1 (5.4)	12.2 (5.3)		
Missing	21,453	556		
Case on PHQ9 and/or GAD7			< 0.001	0.06
No	2,996 (12.3%)	8,527 (16.6%)		
Yes	21,354 (87.7%)	42,984 (83.4%)		
Missing	21,307	518		
Functional Impairment (WSAS)			< 0.001	-0.05
Mean (SD)	15.7 (9.0)	16.1 (8.3)		
Missing	40,542	12,545		
Functional Impairment (WSAS 4-item)			< 0.001	-0.04
Mean (SD)	12.4 (7.5)	12.7 (7.0)		
Missing	38,635	800		
Functional Impairment WSAS Home			0.031	-0.01
Mean (SD)	2.8 (2.3)	2.9 (2.1)		
Missing	38,635	800		
Functional Impairment WSAS Social Leisure			< 0.001	-0.09
Mean (SD)	3.6 (2.5)	3.8 (2.3)		
Missing	38,635	800		
Functional Impairment WSAS Private Leisure			< 0.001	-0.03
Mean (SD)	2.8 (2.4)	2.8 (2.2)		
Missing	38,635	800		
Functional Impairment WSAS Relationships			0.8	0.01
Mean (SD)	3.2 (2.3)	3.2 (2.2)		
Missing	38,635	800		

Functional Impairment WSAS Work			< 0.001	-0.07
Mean (SD)	3.2 (2.4)	3.3 (2.3)		
Missing	38,475	12,530		
Agoraphobia Item			< 0.001	-0.06
Mean (SD)	2.4 (2.6)	2.6 (2.5)		
Missing	38,629	814		
Social Phobia Item			< 0.001	-0.07
Mean (SD)	3.1 (2.5)	3.3 (2.4)		
Missing	38,629	814		
Specific Phobia Item			< 0.001	-0.10
Mean (SD)	2.2 (2.5)	2.5 (2.5)		
Missing	38,629	814		
Diagnosis			< 0.001	0.06
Depression	6,063 (40.2%)	21,296 (41.0%)		
GAD	3,918 (26.0%)	14,203 (27.3%)		
Other Anxiety	2,765 (18.3%)	10,184 (19.6%)		
OCD	718 (4.8%)	2,509 (4.8%)		
PTSD	799 (5.3%)	2,107 (4.1%)		
Other	806 (5.3%)	1,677 (3.2%)		
Missing	30,588	53		
Prescribed Psychotropic Medication			0.2	0.01
No	3,549 (51.6%)	26,469 (52.5%)		
Yes	3,333 (48.4%)	23,996 (47.5%)		
Missing	38,775	1,564		
Ethnicity			< 0.001	0.03
White	16,320 (88.3%)	34,839 (90.3%)		
Minoritised Ethnic Groups	2,153 (11.7%)	3,731 (9.7%)		
Missing	27,184	13,459		
Disability Reported			< 0.001	0.07
No	8,907 (75.6%)	21,191 (81.7%)		
Yes	2,868 (24.4%)	4,731 (18.3%)		
Missing	33,882	26,107		

Note: Percentages were calculated using the available sample for each variable, after excluding missing values. *p*-values are from Pearson's Chi-Squared test for categorical variables and Wilcoxon rank sum for continuous variables. Effect sizes are Cohen's d for continuous variables and Cramer's V for categorical. [1] Case thresholds were PHQ9 ≥10, GAD7 ≥8. [2] IAPT 'problem descriptor'; GAD = generalised anxiety disorder; Other Anxiety = panic disorder, agoraphobia, social phobia, specific phobia; hypochondriacal disorder, unspecified anxiety; OCD = obsessive-compulsive disorder; PTSD = post-traumatic stress disorder; 'Other' included adjustment disorders and eating disorders. Employment and recovery are absent as were NA for excluded patients; employment data is not collected in initial assessments, and recovery requires at least two sets of scores.

Supplementary Information 3. Latent growth curve fit indices and plots

Fit indices and plots of the model-estimated and observed latent growth curves for each outcome. Fit indices as described in Chapter 4 Supplementary Information 2. Rows are ordered by BIC, with the optimal value at the top. Plot intercept represents the estimated mean score on the outcome across the whole sample at baseline (session 0). Only the trajectories with the adjacent residuals correlated ('corr resid') are plotted as all had superior fit to the forms with uncorrelated residuals.

Table *I.* Fit indices of latent growth curves

Depression symptoms (PHQ9)	**Obs**	**Params**	**AIC**	**BIC**	**CFI**	**TLI**	**SRMR**	**RMSEA Estimate**
Quadratic Growth Curve; corr resid	51683	30	1646008	1646273	0.995	0.994	0.020	0.020
Linear Growth Curve; corr resid	51683	26	1650236	1650466	0.977	0.975	0.105	0.041
Log Growth Curve; corr resid	51683	26	1651167	1651397	0.974	0.971	0.049	0.044
Quadratic Growth Curve	51683	20	1651814	1651991	0.972	0.973	0.034	0.043
Linear Growth Curve	51683	16	1661670	1661811	0.933	0.940	0.146	0.064
Log Growth Curve	51683	16	1663932	1664074	0.926	0.933	0.065	0.068
Anxiety symptoms (GAD7)	**Obs**	**Params**	**AIC**	**BIC**	**CFI**	**TLI**	**SRMR**	**RMSEA Estimate**
Quadratic Growth Curve; corr resid	51667	30	1614351	1614616	0.994	0.992	0.025	0.022
Linear Growth Curve; corr resid	51667	26	1618783	1619013	0.972	0.970	0.128	0.044
Quadratic Growth Curve	51667	20	1620077	1620254	0.968	0.969	0.042	0.045
Log Growth Curve ; corr resi	51667	26	1621324	1621554	0.960	0.957	0.056	0.052
Linear Growth Curve	51667	16	1629740	1629882	0.923	0.931	0.175	0.066
Log Growth Curve	51667	16	1634427	1634569	0.904	0.913	0.076	0.074
Functional Impairment (WSAS)	**Obs**	**Params**	**AIC**	**BIC**	**CFI**	**TLI**	**SRMR**	**RMSEA Estimate**
Quadratic Growth Curve; corr resid	32168	30	1078799	1079050	0.994	0.993	0.017	0.022
Linear Growth Curve; corr resid	32168	26	1080961	1081179	0.982	0.98	0.067	0.038
Quadratic Growth Curve	32168	20	1081861	1082029	0.978	0.979	0.031	0.04
Log Growth Curve; corr resid	32168	26	1084869	1085087	0.96	0.957	0.048	0.057
Linear Growth Curve	32168	16	1088271	1088405	0.944	0.949	0.099	0.062
Log Growth Curve	32168	16	1093112	1093246	0.919	0.927	0.049	0.074

Functional Impairment (WSAS 4-item)	Obs	Params	AIC	BIC	CFI	TLI	SRMR	RMSEA Estimate
Quadratic Growth Curve; corr resid	19293	30	662799	663035	0.994	0.993	0.017	0.023
Linear Growth Curve; corr resid	19293	26	663945	664149	0.983	0.981	0.067	0.037
Quadratic Growth Curve	19293	20	664520	664678	0.978	0.979	0.031	0.039
Log Growth Curve; corr resid	19293	26	665733	665937	0.965	0.962	0.043	0.052
Linear Growth Curve	19293	16	668093	668219	0.945	0.950	0.097	0.060
Log Growth Curve	19293	16	670060	670185	0.927	0.934	0.048	0.069

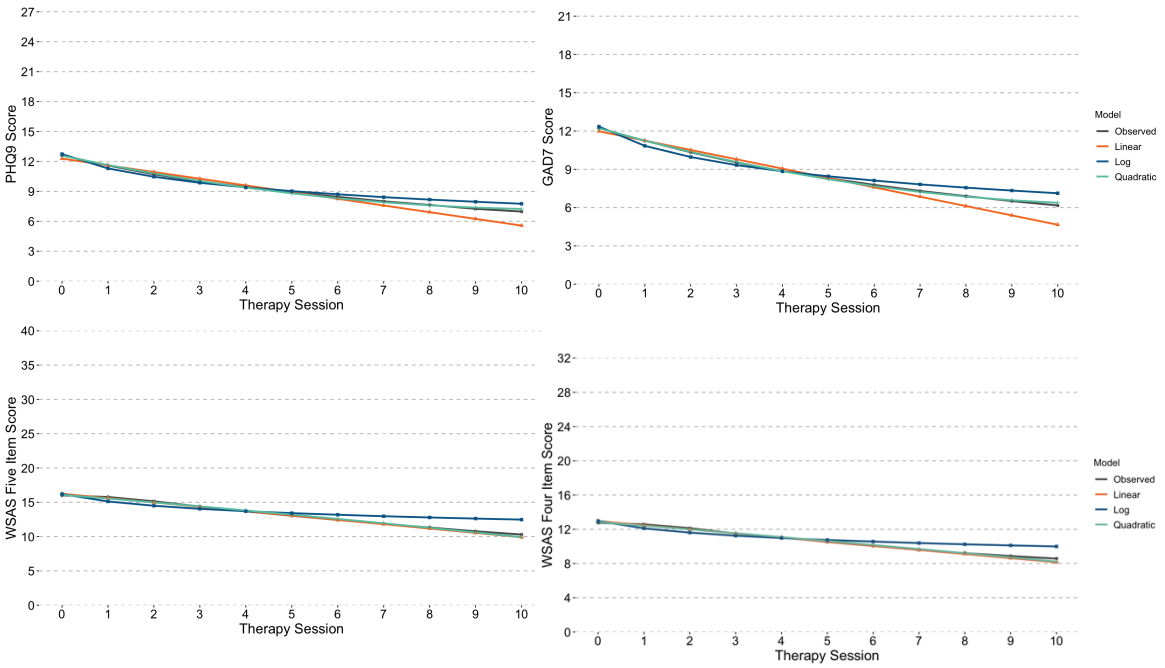


Figure *I*. Plots of estimated and observed latent growth trajectory forms

Supplementary Information 4. Latent class growth analysis

This section describes the model selection for latent class growth analysis (LCGA), a restricted version of growth mixture modelling where the variance within classes is restricted to zero.

Depression and anxiety symptoms LCGAs all had classes with over 1% of the sample and good entropy values. For depression symptoms, besides the six-class, models had trajectories that primarily differed in baseline severity. This was also observed in the two- to four-class models of anxiety symptoms. For both outcomes, the VLMR-LRT *p*-value was significant for the two-class model, despite information criteria decreasing to six-classes. Consistent with the literature, we placed emphasis on BIC which showed a plateau around three classes in the elbow plot (available upon request). In the depression model, an additional trajectory in the four-class appeared to only divide one of the three classes by intercept. For both depression and anxiety symptoms, a three-class model was selected as the optimal, and this indicated: moderate-severe late improvement (grey squares); moderate-severe with improvement (blue triangles); mild symptoms, small improvement (green circles).

The functional impairment LCGAs had good entropy and trajectory classes each had more than 1% of the sample. Trajectories only noticeably differed by intercept values. The VLMR-LRT *p*-value was not significant for any model. The information criteria decreased up to the six-class model and the BIC elbow plot (available upon request) showed a plateau around four classes. The three-class model was selected as the additional trajectory in the four-class differed only by intercept value. In patients with complete 5-item WSAS scores, the correlation between the 4-item score and the work item was r = 0.51 ($p < 0.001$). Results for the 4-item functional impairment score were very similar to the full five-item models but had substantially lower information criteria indicating better fit of the model to the data. The LCGAs had good entropy, the VLMR LRT *p*-value did not indicate an optimal model, and class proportions were all acceptable (> 1%). A three-class model was again selected based on the BIC elbow plot and the plotted estimated mean trajectories. The plots for these models are available on request.

Fit indices of latent growth trajectories

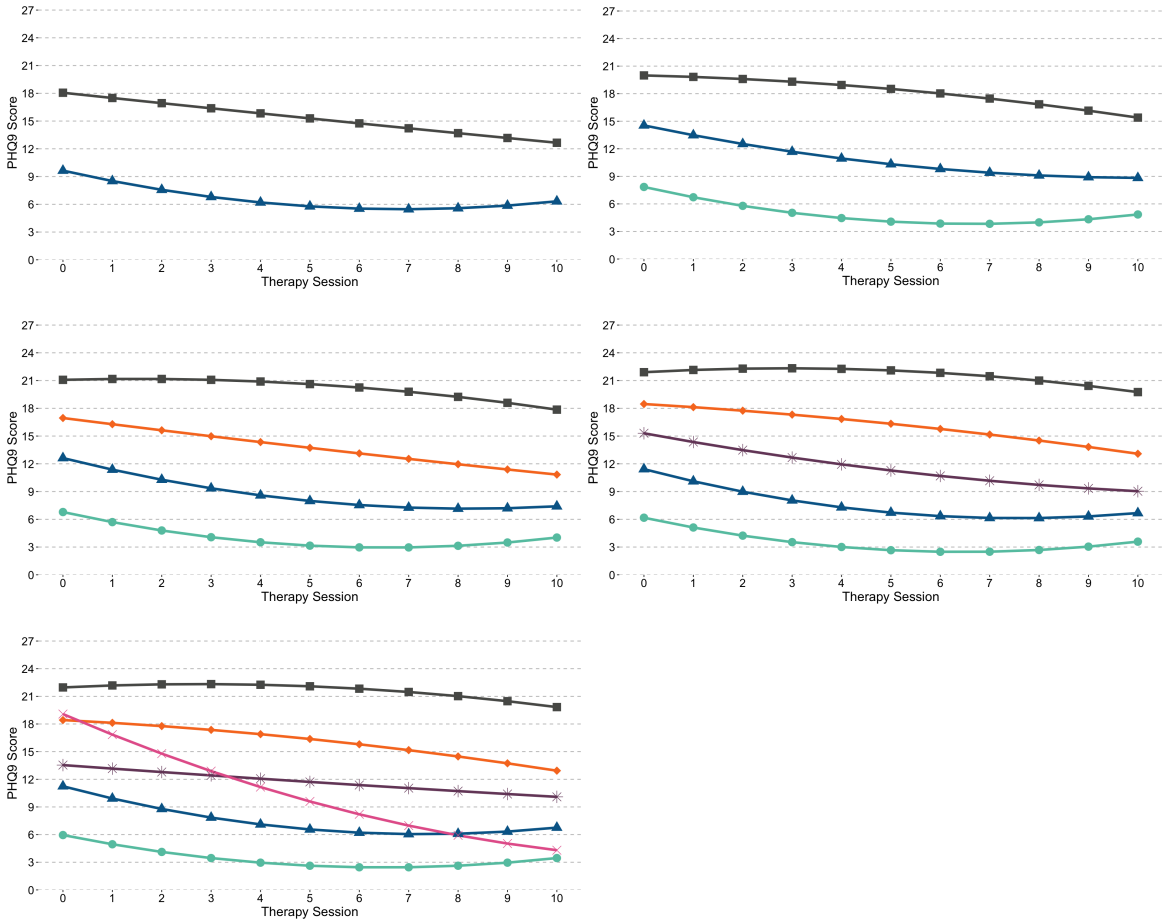
LCGA of Depression Symptoms (PHQ9)	Params	AIC	BIC	Entropy	VLMR LRT *p*-value	Class Proportions
Growth Curve	24	1786472	1786685	NA	NA	100
Two Class	28	1701981	1702229	0.814	0.333	65.3, 34.7
Three Class	32	1669846	1670129	0.782	< 0.001	43.5, 17.1, 39.5
Four Class	36	1656890	1657209	0.752	< 0.001	31.8, 36.2, 9.9, 22.2
Five Class	40	1651879	1652233	0.719	< 0.001	13.3, 32.1, 25.4, 22.9, 6.1
Six Class	44	1649035	1649424	0.708	0.001	13.0, 6.1, 17.2, 31.5, 24.4, 7.8

LCGA of Anxiety Symptoms (GAD7)	Params	AIC	BIC	Entropy	VLMR LRT p-value	Class Proportions
Growth Curve	24	1729658	1729870	NA	NA	100
Two Class	28	1654612	1654860	0.792	0.3333	37.1, 62.9
Three Class	32	1629092	1629376	0.752	< 0.001	38.6, 42.1, 19.3
Four Class	36	1620525	1620843	0.706	< 0.001	22.8, 12.5, 35.3, 29.5
Five Class	40	1617410	1617764	0.670	< 0.001	22.3, 16.2, 27.0, 12.5, 22.0
Six Class	44	1613307	1613697	0.676	< 0.001	10.0, 16.0, 24.3, 12.6, 13.3, 23.8

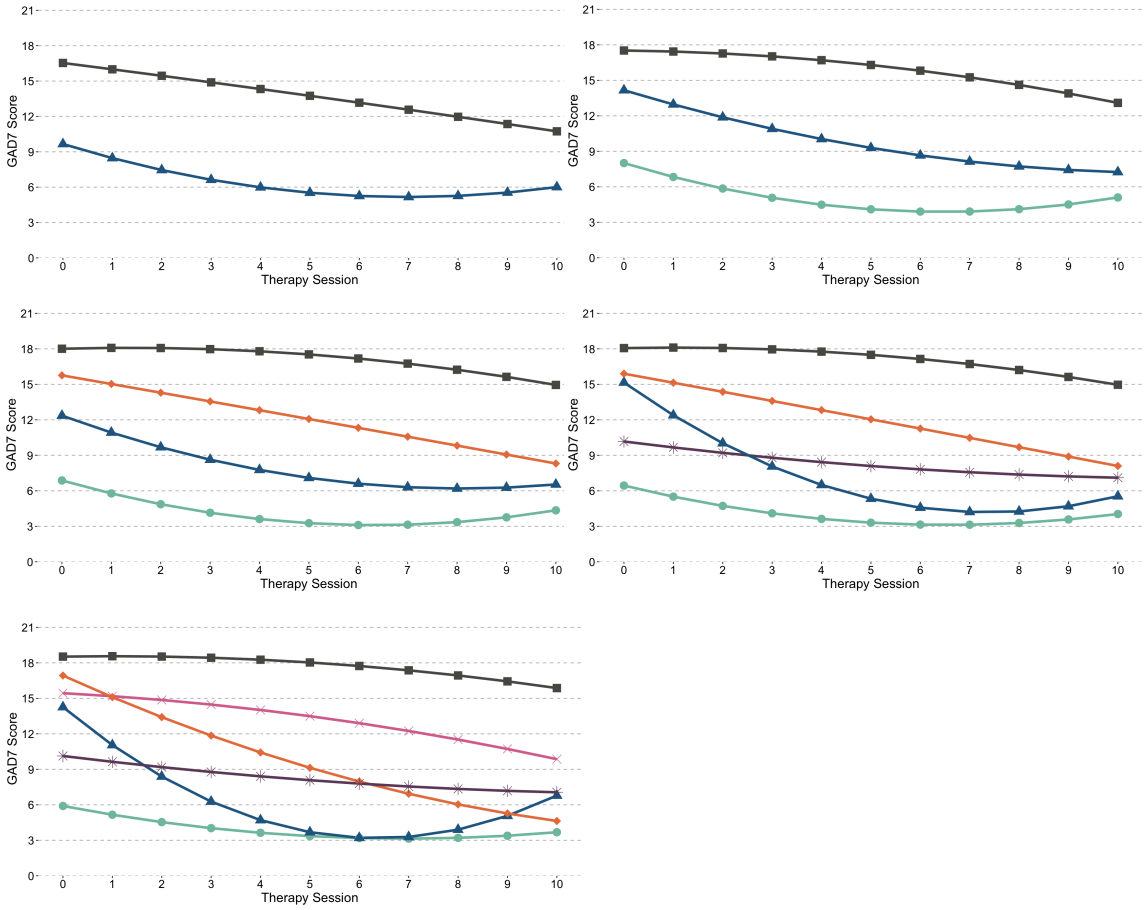
LCGA of Functional Impairment (WSAS)	Params	AIC	BIC	Entropy	VLMR LRT p-value	Class Proportions
Growth Curve	24	1188329	1188530	NA	NA	100
Two Class	28	1130897	1131131	0.817	< 0.001	34.9, 65.0
Three Class	32	1104464	1104732	0.796	< 0.001	39.8, 42.2, 18.1
Four Class	36	1092660	1092962	0.774	< 0.001	23.0, 35.9, 10.9, 30.1
Five Class	40	1086899	1087235	0.748	< 0.001	7.0, 23.9, 30.8, 14.7, 23.6
Six Class	44	1084308	1084677	0.722	0.023	23.4, 15.8, 4.8, 10.3, 19.4, 26.3

LCGA of Functional Impairment (WSAS 4-item)	Params	AIC	BIC	Entropy	VLMR LRT p-value	Class Proportions
Growth Curve	24	726558	726747	NA	NA	100
Two Class	28	690566	690786	0.829	< 0.001	34.2, 65.8
Three Class	32	675187	675439	0.803	< 0.001	43.1, 39.9, 17.0
Four Class	36	668743	669026	0.776	< 0.001	22.4, 11.3, 35.7, 30.6
Five Class	40	665919	666234	0.753	< 0.001	6.8, 12.7, 22.5, 26.4, 31.6
Six Class	44	664717	665064	0.747	0.0002	26.0, 5.2, 30.9, 6.7, 12.1, 19.1

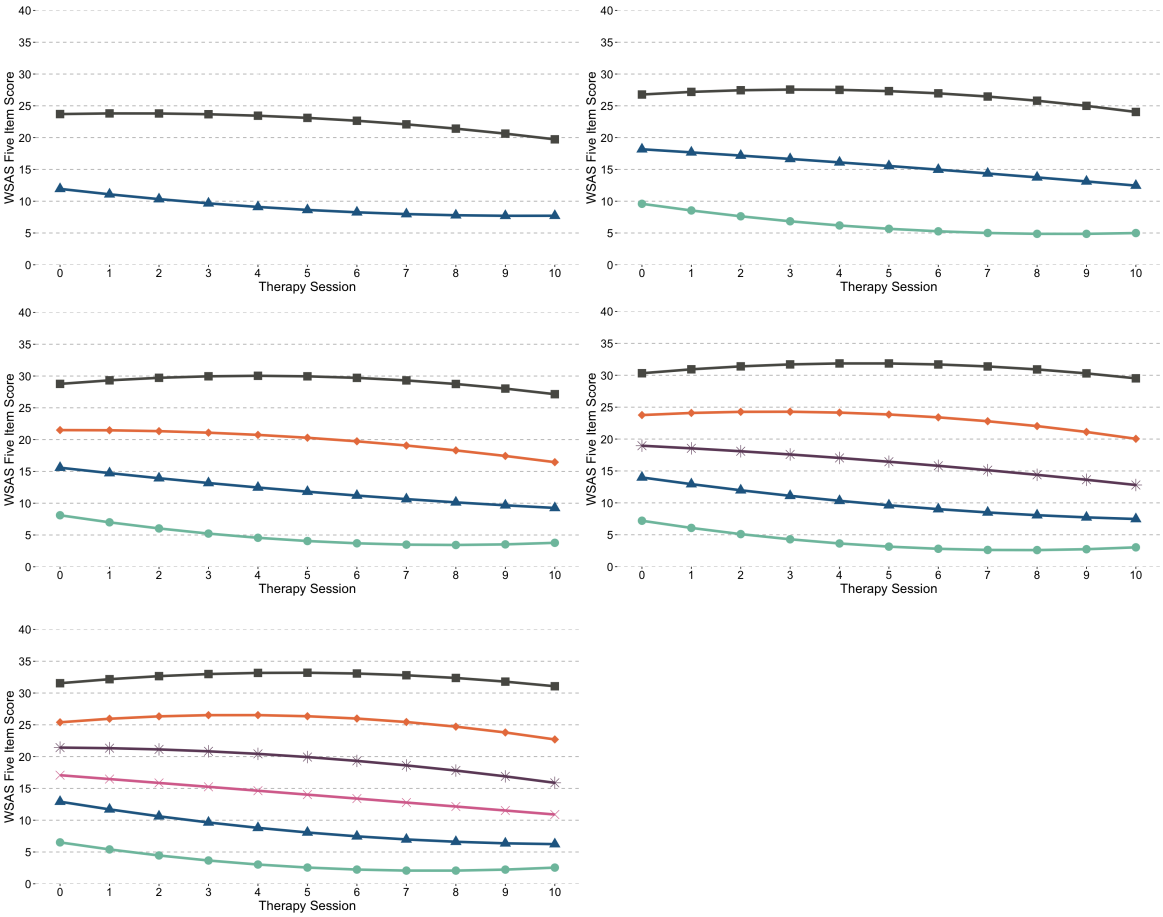
Estimated class trajectories for two- to six-class latent class growth models of depression symptoms (PHQ9) during internet-enabled cognitive-behavioural therapy (N = 51,683)



Estimated class trajectories for two- to six-class latent class growth models of anxiety symptoms (GAD7) during internet-enabled cognitive-behavioural therapy (N = 51,667)



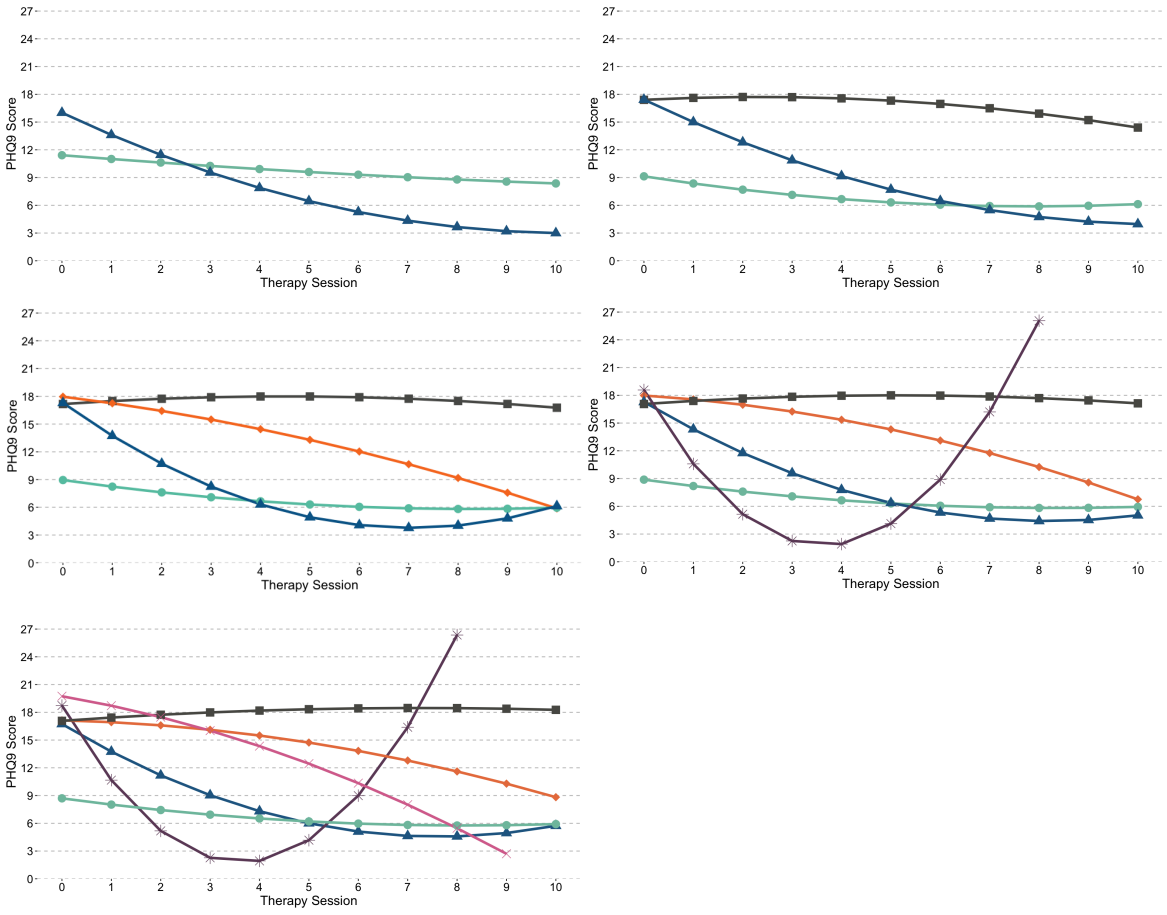
Estimated class trajectories for two- to six-class latent class analysis models of functional impairment (WSAS) during internet-enabled cognitive-behavioural therapy (N = 32,168)



Supplementary Information 5. Growth mixture model of depression symptoms (PHQ9)

This section details the model selection for the GMM of depression symptoms, including trajectory plots, and describes the selected model. The fit indices are presented in the main text. The VLMR LRT *p*-value remained significant and information criteria decreased up to six classes, however, six was unrealistically high compared with existing studies (hence it was the upper number to test). All models had classes with more than 1% of the sample. The five- and six-class models revealed trajectories outside the PHQ9 range. As per the method, an elbow plot of BIC values was used to identify a point of diminishing gains, and this suggested a four-class model with medium entropy (0.629).

Estimated class trajectories for two- to six-class growth mixture models of depression symptoms (PHQ9) during internet-enabled cognitive-behavioural therapy (N = 51,683)



Characteristics of patients within each trajectory class of a four-class depression growth mixture model (based on most likely trajectory class)

Variable	Moderate-severe plateau N = 6,892	Moderate-severe gradual improvement N = 6,829	Moderate-severe fast improvement N = 5,683	Mild, small improvement N = 32,279
Anxiety Symptoms (GAD7)				
Mean (SD)	15.6 (4.2)	16.0 (3.9)	15.1 (4.3)	10.2 (4.8)
Missing	21	16	11	168
Depression Symptoms (PHQ9)				
Mean (SD)	17.6 (4.3)	19.4 (3.6)	18.5 (3.4)	9.1 (4.1)
Missing	18	12	11	150
Functional Impairment (WSAS)				
Mean (SD)	22.6 (7.8)	22.3 (7.4)	18.1 (7.5)	13.1 (7.0)
Missing	1,741	1,604	1,375	7,484
WSAS Home Management				
Mean (SD)	4.1 (2.2)	4.1 (2.1)	3.3 (2.1)	2.2 (1.9)
Missing	41	46	36	338
WSAS Private Leisure				
Mean (SD)	4.2 (2.3)	4.2 (2.3)	3.3 (2.2)	2.2 (1.9)
Missing	41	46	36	338
WSAS Social Leisure				
Mean (SD)	5.2 (2.2)	5.1 (2.2)	4.2 (2.2)	3.2 (2.1)
Missing	41	46	36	338
WSAS Relationships				
Mean (SD)	4.3 (2.2)	4.4 (2.2)	3.5 (2.2)	2.6 (2.0)
Missing	41	46	36	338
WSAS Work				
Mean (SD)	4.6 (2.4)	4.4 (2.3)	3.7 (2.3)	2.8 (2.1)
Missing	1,741	1,602	1,374	7,475
Agoraphobia Item				
Mean (SD)	3.9 (2.6)	3.7 (2.6)	2.7 (2.5)	2.1 (2.3)
Missing	43	45	43	343
Social Phobia Item				
Mean (SD)	4.6 (2.4)	4.4 (2.4)	3.5 (2.3)	2.7 (2.1)
Missing	43	45	43	343
Specific Phobia Item				
Mean (SD)	3.4 (2.7)	3.1 (2.7)	2.3 (2.5)	2.1 (2.3)
Missing	43	45	43	343
Diagnosis				
Depression	3,408 (49.5%)	3,703 (54.3%)	3,200 (56.4%)	10,855 (33.7%)
GAD	1,282 (18.6%)	1,327 (19.5%)	1,260 (22.2%)	10,249 (31.8%)
Other Anxiety	1,147 (16.7%)	1,020 (15.0%)	737 (13.0%)	7,200 (22.3%)
OCD	360 (5.2%)	241 (3.5%)	140 (2.5%)	1,757 (5.4%)
PTSD	519 (7.5%)	374 (5.5%)	214 (3.8%)	989 (3.1%)
Other	170 (2.5%)	153 (2.2%)	124 (2.2%)	1,202 (3.7%)
Missing	6	11	8	27

Prescribed Psychotropic Medication				
No	2,589 (38.5%)	2,623 (39.4%)	2,615 (46.9%)	18,634 (59.2%)
Yes	4,132 (61.5%)	4,036 (60.6%)	2,961 (53.1%)	12,861 (40.8%)
Missing	171	170	107	784
Number of Sessions (incl baseline)				
Mean (SD)	6.9 (3.3)	6.4 (3.4)	5.7 (2.5)	5.8 (3.1)
Missing	0	0	0	0
Case on PHQ9 and/or GAD7				
No	66 (1.0%)	3 (0.0%)	2 (0.0%)	8,455 (26.3%)
Yes	6,808 (99.0%)	6,816 (100.0%)	5,671 (100.0%)	23,684 (73.7%)
Missing	18	10	10	140
Recovered				
No	6,424 (98.5%)	4,439 (68.4%)	1,619 (29.9%)	10,595 (47.3%)
Yes	95 (1.5%)	2,048 (31.6%)	3,801 (70.1%)	11,814 (52.7%)
Missing	373	342	263	9,870
Age (years)				
Mean (SD)	31.9 (11.8)	32.4 (11.8)	35.7 (12.6)	34.9 (12.4)
Missing	125	84	28	262
Gender				
Female	5,206 (76%)	5,128 (75.6%)	4,184 (73.9%)	23,496 (73.0%)
Male	1,640 (24.0%)	1,657 (24.4%)	1,478 (26.1%)	8,685 (27.0%)
Missing	46	44	21	98
Ethnicity				
White	5,041 (89.4%)	4,876 (89.0%)	3,517 (89.8%)	21,212 (91.0%)
Minoritised Ethnic Groups	600 (10.6%)	602 (11.0%)	398 (10.2%)	2,099 (9.0%)
Missing	1,251	1,351	1,768	8,968
Disability Reported				
No	2,916 (73.0%)	2,923 (76.3%)	2,067 (81.2%)	13,164 (85.5%)
Yes	1,076 (27.0%)	906 (23.7%)	479 (18.8%)	2,231 (14.5%)
Missing	2,900	3,000	3,137	16,884
Employment Status				
Employed	3,850 (58.9%)	4,126 (63.7%)	3,938 (72.6%)	22,598 (73.7%)
Non-worker	1,247 (19.1%)	1,158 (17.9%)	829 (15.3%)	5,306 (17.3%)
Unemployed	1,443 (22.1%)	1,189 (18.4%)	660 (12.2%)	2,769 (9.0%)
Missing	352	356	256	1,606

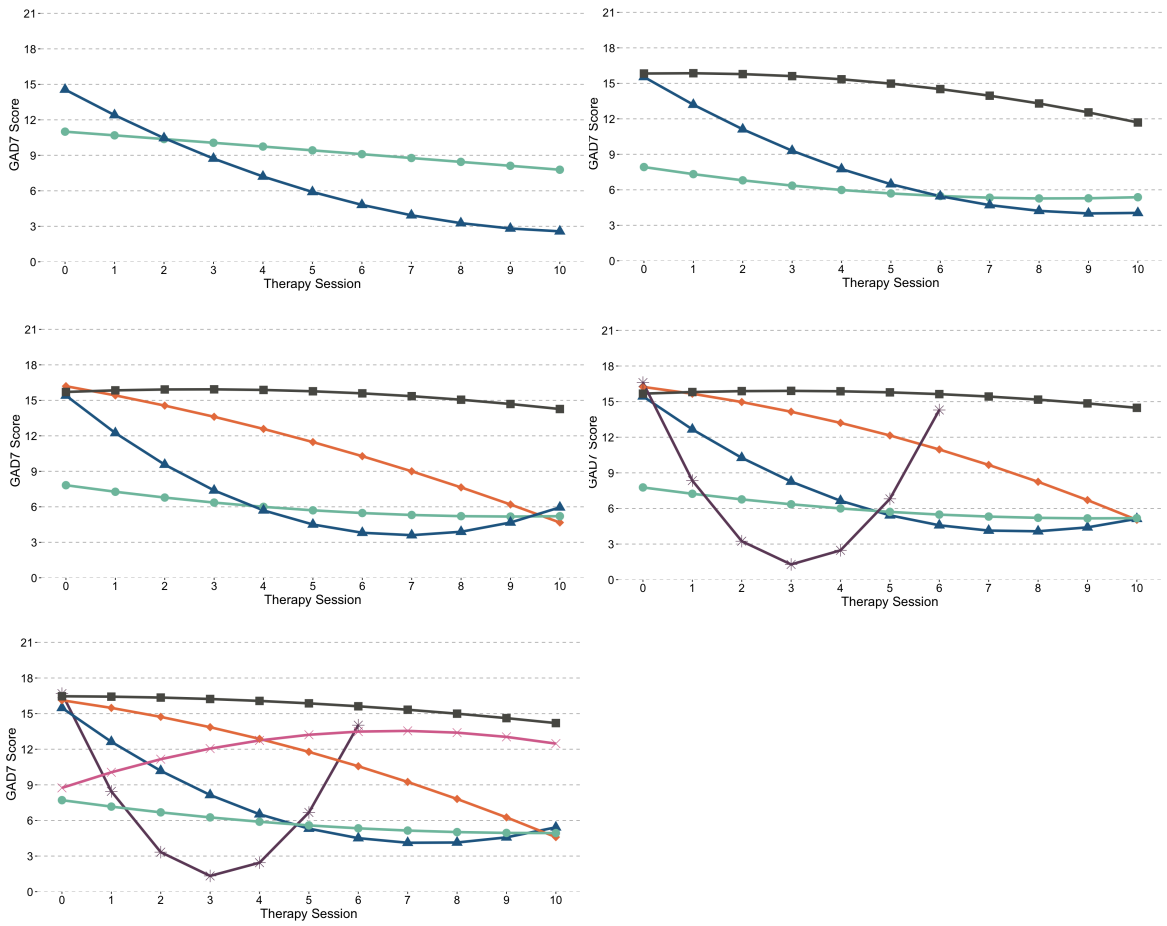
Descriptives of the growth factors for the four-class growth mixture model of depression symptoms

Latent Class	Parameter	Factor	Estimate	SE	Est SE	*p*-value
Moderate-severe plateau	Means	INT	17.165	0.108	158.625	0
		LIN	0.367	0.027	13.362	0
		QUAD	-0.041	0.004	-11.615	0
	Variances	INT	11.751	0.134	87.813	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, gradual improvement	Means	INT	17.957	0.111	161.761	0
		LIN	-0.657	0.08	-8.178	0
		QUAD	-0.055	0.006	-8.522	0
	Variances	INT	11.751	0.134	87.813	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, fast improvement	Means	INT	17.286	0.092	186.886	0
		LIN	-3.83	0.093	-41.182	0
		QUAD	0.271	0.012	22.981	0
	Variances	INT	11.751	0.134	87.813	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Mild, small improvement	Means	INT	8.946	0.049	182.603	0
		LIN	-0.758	0.015	-51.455	0
		QUAD	0.046	0.002	29.611	0
	Variances	INT	11.751	0.134	87.813	0
		LIN	0	0	999	999
		QUAD	0	0	999	999

Supplementary Information 6. Growth mixture model of anxiety symptoms (GAD7)

This section details the model selection for the GMM of anxiety symptoms, including trajectory plots, and describes the selected model. Fit indices are in the main text. All models had classes containing more than 1% of the sample. Information criteria decreased up to six classes and the VLMR LRT *p*-value remained significant, however, six classes were unrealistically high compared with existing studies. The five- and six-class models revealed trajectories outside of the GAD7 range The BIC elbow plot suggested a four-class model, with medium entropy (0.609).

Estimated class trajectories for two- to six-class growth mixture models of anxiety symptoms (GAD7) during internet-enabled cognitive-behavioural therapy (N = 51,667)



Characteristics of patients within each trajectory class of a four-class anxiety growth mixture model (based on most likely trajectory class)

Variable	Moderate-severe plateau N = 7,740	Moderate-severe gradual improvement N = 10,116	Moderate-severe fast improvement N = 9,142	Mild, small improvement N = 24,669
Anxiety Symptoms (GAD7)				
Mean (SD)	15.9 (3.6)	16.8 (2.8)	16.0 (2.7)	7.8 (3.3)
Missing	15	27	17	135
Depression Symptoms (PHQ9)				
Mean (SD)	16.7 (5.4)	16.1 (5.3)	14.0 (5.3)	9.4 (5.0)
Missing	13	30	15	135
Functional Impairment (WSAS)				
Mean (SD)	21.3 (8.3)	19.7 (7.9)	15.7 (7.7)	13.2 (7.3)
Missing	1,855	2,431	2,135	5,766
WSAS Home Management				
Mean (SD)	3.8 (2.2)	3.5 (2.2)	2.8 (2.0)	2.3 (1.9)
Missing	31	68	55	288
WSAS Private Leisure				
Mean (SD)	3.9 (2.3)	3.6 (2.3)	2.8 (2.2)	2.2 (2.0)
Missing	31	68	55	288
WSAS Social Leisure				
Mean (SD)	5.0 (2.3)	4.7 (2.2)	3.7 (2.2)	3.1 (2.1)
Missing	31	68	55	288
WSAS Relationships				
Mean (SD)	4.2 (2.3)	3.9 (2.2)	3.1 (2.2)	2.6 (2.0)
Missing	31	68	55	288
WSAS Work				
Mean (SD)	4.3 (2.4)	4.0 (2.4)	3.3 (2.3)	2.8 (2.1)
Missing	1,855	2,428	2,133	5,761
Agoraphobia Item				
Mean (SD)	3.8 (2.7)	3.5 (2.6)	2.5 (2.4)	1.9 (2.2)
Missing	32	69	57	300
Social Phobia Item				
Mean (SD)	4.4 (2.4)	4.1 (2.4)	3.1 (2.3)	2.7 (2.1)
Missing	32	69	57	300
Specific Phobia Item				
Mean (SD)	3.4 (2.7)	3.0 (2.6)	2.3 (2.5)	2.0 (2.3)
Missing	32	69	57	300
Diagnosis				
Depression	3,233 (41.8%)	3,925 (38.9%)	3,497 (38.3%)	10,504 (42.6%)
GAD	1,822 (23.6%)	2,993 (29.6%)	3,076 (33.7%)	6,220 (25.2%)
Other Anxiety	1,412 (18.3%)	1,881 (18.6%)	1,699 (18.6%)	5,109 (20.7%)
OCD	557 (7.2%)	594 (5.9%)	321 (3.5%)	1,026 (4.2%)
PTSD	544 (7.0%)	492 (4.9%)	303 (3.3%)	759 (3.1%)
Other	167 (2.2%)	215 (2.1%)	242 (2.6%)	1,024 (4.2%)
Missing	5	16	4	27

Prescribed Psychotropic Medication				
No	3,173 (41.9%)	4,468 (45.3%)	4,812 (53.6%)	14,010 (48.3%)
Yes	4,398 (58.1%)	5,400 (54.7%)	4,162 (46.4%)	10,031 (41.7%)
Missing	169	248	168	628
Number of Sessions (incl baseline)				
Mean (SD)	7.4 (3.1)	6.3 (3.4)	5.6 (2.6)	5.6 (3.1)
Missing	0	0	0	0
Case on PHQ9 and/or GAD7				
No	83 (1.1%)	3 (0.0%)	1 (0.0%)	8,430 (34.3%)
Yes	7,645 (98.9%)	10,087 (100.0%)	9,126 (100.0%)	16,118 (65.7%)
Missing	12	26	15	121
Recovered				
No	7,227 (98.5%)	6,347 (66.5%)	2,453 (28.2%)	7,048 (46.2%)
Yes	113 (1.5%)	3,196 (33.5%)	6,239 (71.8%)	8,210 (53.8%)
Missing	400	573	450	9,411
Age (years)				
Mean (SD)	32.3 (11.7)	32.4 (11.5)	35.1 (12.2)	35.3 (12.7)
Missing	117	124	41	216
Gender				
Female	5,954 (77.4%)	7,771 (77.2%)	6,879 (75.4%)	17,403 (70.8%)
Male	1,738 (22.6%)	2,290 (22.8%)	2,241 (24.6%)	7,182 (29.2%)
Missing	48	55	22	84
Ethnicity				
White	5,671 (89.8%)	7,226 (90.0%)	5,726 (90.9%)	16,019 (90.5%)
Minoritised Ethnic Groups	645 (10.2%)	805 (10.0%)	572 (9.1%)	1,677 (9.5%)
Missing	1,424	2,085	2,844	6,973
Disability Reported				
No	3,309 (75.0%)	4,361 (79.8%)	3,372 (84.0%)	10,025 (84.5%)
Yes	1,103 (25.0%)	1,102 (20.2%)	644 (16.0%)	1,845 (15.5%)
Missing	3,328	4,653	5,126	12,799
Employment Status				
Employed	4,550 (61.6%)	6,514 (68.0%)	6,524 (74.7%)	16,925 (72.3%)
Non-worker	1,383 (18.7%)	1,709 (17.8%)	1,334 (15.3%)	4,115 (17.6%)
Unemployed	1,457 (19.7%)	1,362 (14.2%)	873 (10.0%)	2,368 (10.1%)
Missing	350	531	411	1,261

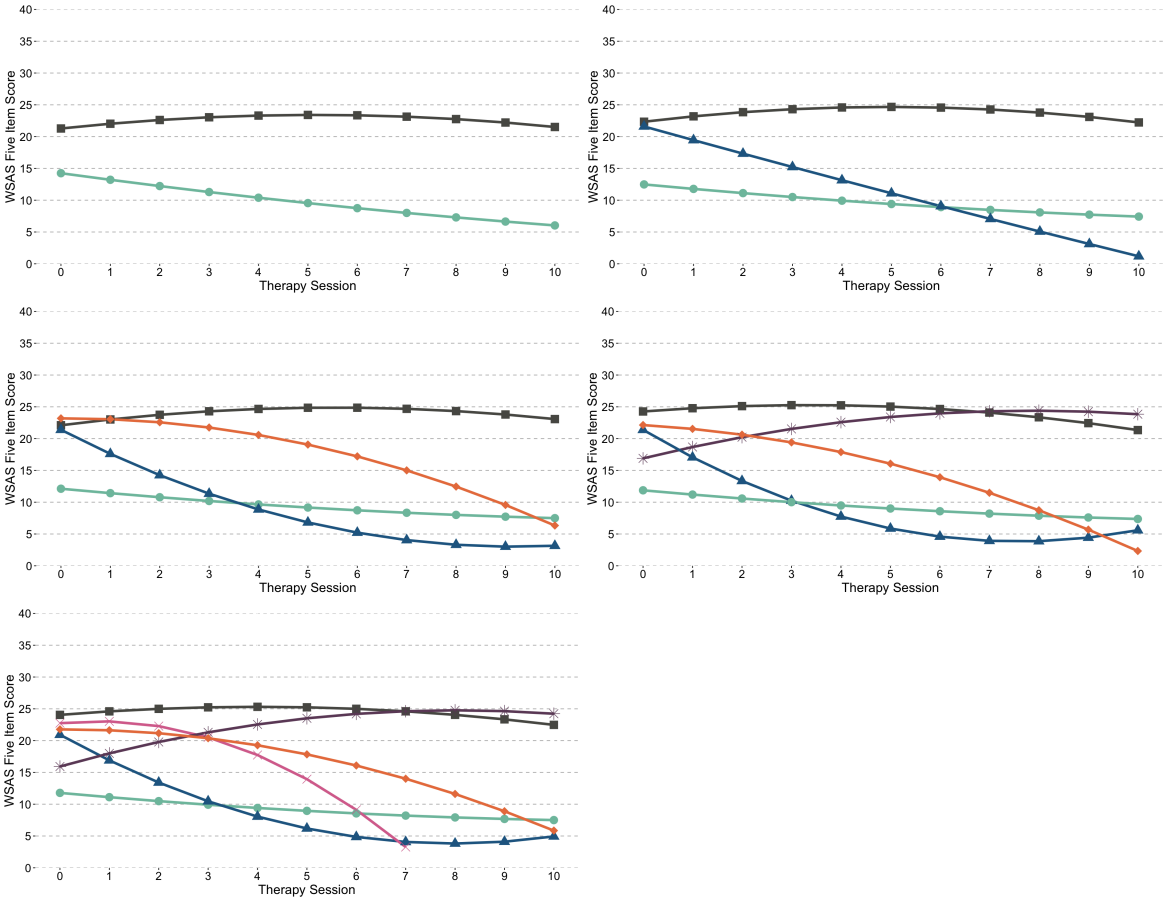
Descriptives of the growth factors for the four-class growth mixture model of anxiety symptoms

Latent Class	Parameter	Factor	Estimate	SE	Est SE	*p*-value
Moderate-severe plateau	Means	INT	15.71	0.065	241.734	0
		LIN	0.167	0.019	8.779	0
		QUAD	-0.031	0.003	-12.309	0
	Variances	INT	6.676	0.071	93.408	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, gradual improvement	Means	INT	16.197	0.058	278.471	0
		LIN	-0.737	0.059	-12.499	0
		QUAD	-0.042	0.005	-8.132	0
	Variances	INT	6.676	0.071	93.408	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, fast improvement	Means	INT	15.405	0.056	275.711	0
		LIN	-3.414	0.065	-52.604	0
		QUAD	0.247	0.008	29.476	0
	Variances	INT	6.676	0.071	93.408	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Mild, small improvement	Means	INT	7.827	0.039	200.284	0
		LIN	-0.587	0.015	-39.647	0
		QUAD	0.033	0.002	20.373	0
	Variances	INT	6.676	0.071	93.408	0
		LIN	0	0	999	999
		QUAD	0	0	999	999

Supplementary Information 7. Growth mixture model of functional Impairment (WSAS)

This section details the model selection for the GMM of functional impairment, including trajectory plots, and describes the selected model. Information criteria decreased up six-classes and the VLMR LRT remained significant. Each model had classes with more than 1% of the sample. The six-class model had a trajectory outside the WSAS range. The BIC elbow plot suggested a four-class model with medium entropy (0.619).

Estimated class trajectories for two- to six-class growth mixture models of functional impairment during internet-enabled cognitive-behavioural therapy (N = 32,168)



Characteristics of patients within each trajectory class of a four-class functional impairment growth mixture model (based on most likely trajectory class)

Variable	Moderate-severe plateau N = 7,097	Moderate-severe gradual improvement N = 1,655	Moderate-severe fast improvement N = 1,865	Mild, small improvement N = 21,551
Anxiety Symptoms (GAD7)				
Mean (SD)	14.7 (4.6)	14.7 (4.5)	13.7 (5.0)	10.8 (5.1)
Missing	19	4	1	114
Depression Symptoms (PHQ9)				
Mean (SD)	16.6 (5.3)	16.5 (5.0)	15.0 (5.4)	10.4 (5.4)
Missing	20	4	0	109
Functional Impairment (WSAS)				
Mean (SD)	23.8 (6.8)	24.6 (5.7)	23.4 (5.4)	12.2 (6.0)
Missing	25	4	1	131
WSAS Home Management				
Mean (SD)	4.4 (2.0)	4.6 (1.7)	4.3 (1.8)	2.1 (1.7)
Missing	25	4	1	131
WSAS Private Leisure				
Mean (SD)	4.3 (2.2)	4.6 (2.0)	4.5 (2.0)	2.0 (1.8)
Missing	25	4	1	131
WSAS Social Leisure				
Mean (SD)	5.5 (2.0)	5.5 (1.8)	5.3 (1.8)	3.0 (1.9)
Missing	25	4	1	131
WSAS Relationships				
Mean (SD)	4.7 (2.1)	4.8 (1.9)	4.6 (2.0)	2.5 (1.9)
Missing	25	4	1	131
WSAS Work				
Mean (SD)	4.9 (2.2)	5.0 (1.9)	4.6 (2.0)	2.6 (1.9)
Missing	25	4	1	131
Agoraphobia Item				
Mean (SD)	3.7 (2.7)	3.3 (2.5)	2.7 (2.5)	2.0 (2.2)
Missing	31	5	3	171
Social Phobia Item				
Mean (SD)	4.6 (2.3)	4.3 (2.3)	3.5 (2.4)	2.6 (2.0)
Missing	31	5	3	171
Specific Phobia Item				
Mean (SD)	3.4 (2.7)	3.0 (2.6)	2.3 (2.5)	2.0 (2.3)
Missing	31	5	3	171
Diagnosis				
Depression	3,392 (47.9%)	835 (50.5%)	930 (49.9%)	7,805 (36.3%)
GAD	1,412 (19.9%)	397 (24.0%)	468 (25.1%)	6,795 (31.6%)
Other Anxiety	1,306 (18.4%)	259 (15.6%)	289 (15.5%)	4,449 (20.7%)
OCD	354 (5.0%)	61 (3.7%)	57 (3.1%)	1,087 (5.0%)
PTSD	396 (5.6%)	79 (4.8%)	73 (3.9%)	644 (3.0%)
Other	227 (3.2%)	24 (1.5%)	47 (2.5%)	750 (3.5%)
Missing	10	0	1	21

Prescribed Psychotropic Medication				
No	2,826 (40.7%)	701 (42.9%)	896 (48.5%)	12,319 (58.3%)
Yes	4,114 (59.3%)	932 (57.1%)	951 (51.5%)	8,808 (41.7%)
Missing	157	22	18	424
Number of Sessions (incl baseline)				
Mean (SD)	6.1 (3.5)	8.4 (2.0)	6.2 (2.4)	5.6 (3.1)
Missing	0	0	0	0
Case on PHQ9 and/or GAD7				
No	251 (3.5%)	40 (2.4%)	121 (6.5%)	5,145 (24.0%)
Yes	6,827 (96.5%)	1,611 (97.6%)	1,744 (93.5%)	16,305 (76.0%)
Missing	19	4	0	101
Recovered				
No	5,870 (90.2%)	583 (37.2%)	372 (22.1%)	7,245 (47.1%)
Yes	640 (9.8%)	983 (62.8%)	1,314 (77.9%)	8,123 (52.9%)
Missing	587	89	179	6,183
Age (years)				
Mean (SD)	32.5 (10.8)	33.3 (10.6)	34.7 (11.1)	34.4 (11.0)
Missing	56	5	6	114
Gender				
Female	5,221 (74.1%)	1,197 (72.7%)	1,316 (70.8%)	15,410 (71.7%)
Male	1,829 (25.9%)	449 (27.3%)	543 (29.2%)	6,089 (28.3%)
Missing	47	9	6	52
Ethnicity				
White	5,162 (90.4%)	1,237 (91.0%)	1,271 (90.5%)	14,277 (92.7%)
Minoritised Ethnic Groups	549 (9.6%)	122 (9.0%)	134 (9.5%)	1,127 (7.3%)
Missing	1,386	296	460	6,147
Disability Reported				
No	3,096 (77.3%)	820 (83.9%)	809 (86.4%)	8,974 (88.7%)
Yes	908 (22.7%)	157 (16.1%)	127 (13.6%)	1,143 (11.3%)
Missing	3,093	678	929	11,434
Employment Status				
Employed	5,049 (74.3%)	1,372 (85.0%)	1,611 (88.5%)	18,505 (89.7%)
Non-worker	588 (8.6%)	104 (6.4%)	101 (5.5%)	1,201 (5.8%)
Unemployed	1,162 (17.1%)	139 (8.6%)	109 (6.0%)	935 (4.5%)
Missing	298	40	44	910

Descriptives of the growth factors for the four-class growth mixture model of functional impairment

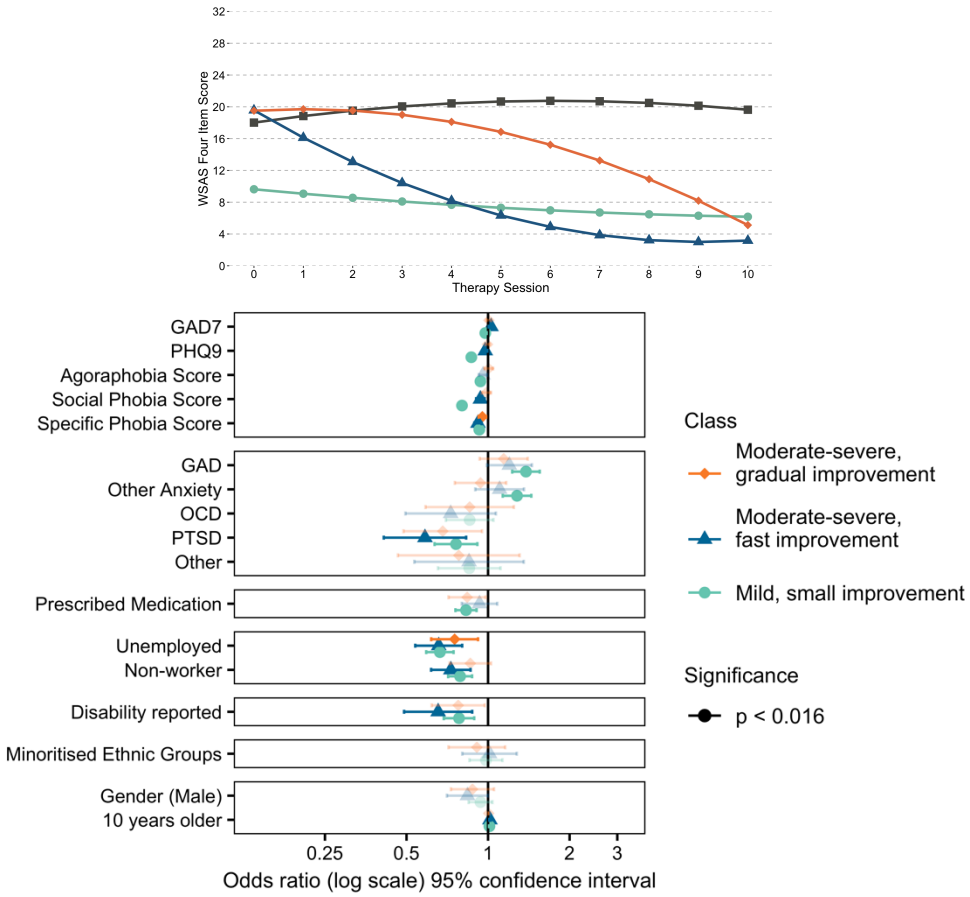
Latent Class	Parameter	Factor	Estimate	SE	Est SE	*p*-value
Moderate-severe plateau	Means	INT	22.1	0.176	125.438	0
		LIN	1.005	0.035	28.505	0
		QUAD	-0.091	0.005	-19.31	0
	Variances	INT	30.522	0.492	62.051	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, gradual improvement	Means	INT	23.187	0.286	81.194	0
		LIN	0.037	0.21	0.174	0.862
		QUAD	-0.172	0.011	-15.309	0
	Variances	INT	30.522	0.492	62.051	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Moderate-severe, fast improvement	Means	INT	21.386	0.259	82.543	0
		LIN	-4.003	0.362	-11.046	0
		QUAD	0.218	0.046	4.689	0
	Variances	INT	30.522	0.492	62.051	0
		LIN	0	0	999	999
		QUAD	0	0	999	999
Mild, small improvement	Means	INT	12.12	0.088	137.418	0
		LIN	-0.72	0.025	-28.486	0
		QUAD	0.026	0.003	8.931	0
	Variances	INT	30.522	0.492	62.051	0
		LIN	0	0	999	999
		QUAD	0	0	999	999

Supplementary Information 8. Growth mixture model of functional impairment (reduced WSAS; 4-item)

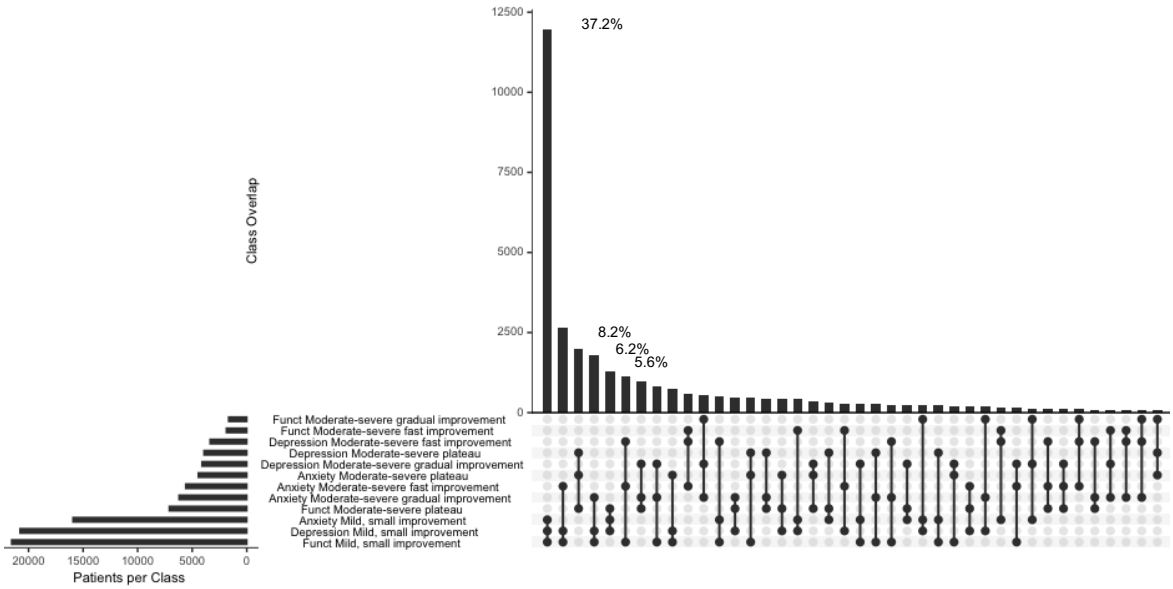
This section details the model selection for the GMM of the 4-item functional impairment score, including fit indices. Trajectory plots were highly similar to those of the 5-item model and are available upon request, as are the descriptives of the model and the participants in each class. All models had classes with more than 1% of the sample. Information criteria decreased up to a six-class model and the VLMR LRT *p*-value did not become non-significant. The six-class model had a trajectory that went outside of the scoring range. The BIC elbow plot suggested a four-class model. Entropy of this model was medium (0.670). A conditional model using this outcome is also presented below.

Fit indices for growth mixture model of functional impairment (reduced WSAS; 4-item) during internet-enabled cognitive-behavioural therapy (N = 19,293)

GMM of Functional Impairment	Parameters	AIC	BIC	Entropy	VLMR LRT *p*-value	Class Proportions
Growth Curve	25	667686	667882	NA	NA	100
Two Class	29	664443	664671	0.627	< 0.001	23.2, 76.8
Three Class	33	661995	662255	0.648	< 0.001	66.4, 22.6, 11.0
Four Class	37	660912	661203	0.670	< 0.001	19.2, 64.5, 7.1, 9.2
Five Class	41	660382	660705	0.684	0.028	8.2, 3.6, 8.5, 17.0, 62.7
Six Class	45	659924	660278	0.680	0.016	15.5, 9.3, 61.0, 8.0, 3.3, 2.8



Supplementary Figure 2. Overlap of class membership for each outcome model (for patients with a trajectory for all three models, n = 32,159)



Supplementary Table 4. Multinomial regression output of four-class growth mixture model of depression symptoms (PHQ9)

Reference class: Moderate-severe plateau.

Class	Baseline Variable	OR	Lower CI	Upper CI	*p*-value	Statistic	df
Moderate-severe gradual improvement	(Intercept)	0.66	0.55	0.80	0.00	-4.24	21658
Moderate-severe gradual improvement	Anxiety Symptoms (GAD7)	1.04	1.03	1.05	0.00	8.93	49216
Moderate-severe gradual improvement	WSAS Home	1.01	0.99	1.03	0.19	1.31	42696
Moderate-severe gradual improvement	WSAS Social Leisure	0.99	0.97	1.01	0.48	-0.71	40590
Moderate-severe gradual improvement	WSAS Private Leisure	1.01	0.99	1.02	0.58	0.55	42133
Moderate-severe gradual improvement	WSAS Relationships	1.00	0.98	1.02	0.94	-0.07	49417
Moderate-severe gradual improvement	Agoraphobia Score	1.00	0.98	1.01	0.67	-0.43	38130
Moderate-severe gradual improvement	Social Phobia Score	0.97	0.96	0.99	0.01	-2.77	38341
Moderate-severe gradual improvement	Specific Phobia Score	0.98	0.96	0.99	0.00	-3.29	45644.
Moderate-severe gradual improvement	Prescribed Medication (Yes)	0.99	0.92	1.07	0.80	-0.26	9539
Moderate-severe gradual improvement	Diagnosis (GAD)	0.89	0.81	0.98	0.01	-2.46	51427
Moderate-severe gradual improvement	Diagnosis (Other Anxiety)	0.84	0.76	0.93	0.00	-3.39	50557
Moderate-severe gradual improvement	Diagnosis (OCD)	0.59	0.50	0.70	0.00	-6.02	51401

Moderate-severe gradual improvement	Diagnosis (PTSD)	0.67	0.58	0.77	0.00	-5.49	49416
Moderate-severe gradual improvement	Diagnosis (Other)	0.82	0.66	1.03	0.08	-1.73	50759
Moderate-severe gradual improvement	Ethnicity (Minoritised Ethnic Groups)	1.05	0.93	1.18	0.41	0.82	1368
Moderate-severe gradual improvement	Disability (Yes)	0.90	0.81	1.00	0.05	-1.96	132
Moderate-severe gradual improvement	Employment (Non-worker)	0.91	0.83	1.00	0.04	-2.03	3738
Moderate-severe gradual improvement	Employment (Unemployed)	0.83	0.75	0.91	0.00	-3.80	2981
Moderate-severe gradual improvement	Age (10 years)	1.04	1.00	1.07	0.03	2.19	8160
Moderate-severe gradual improvement	Gender (Male)	1.04	0.96	1.12	0.38	0.88	28502
Moderate-severe fast improvement	(Intercept)	1.71	1.40	2.08	0.00	5.28	17080
Moderate-severe fast improvement	Anxiety Symptoms (GAD7)	1.02	1.01	1.03	0.00	4.52	47354
Moderate-severe fast improvement	WSAS Home	0.93	0.91	0.95	0.00	-6.73	22806
Moderate-severe fast improvement	WSAS Social Leisure	0.96	0.94	0.98	0.00	-3.84	47401
Moderate-severe fast improvement	WSAS Private Leisure	0.93	0.91	0.95	0.00	-7.29	32724
Moderate-severe fast improvement	WSAS Relationships	0.94	0.92	0.96	0.00	-6.41	41765
Moderate-severe fast improvement	Agoraphobia Score	0.96	0.94	0.98	0.00	-3.95	30160
Moderate-severe fast improvement	Social Phobia Score	0.94	0.92	0.96	0.00	-6.33	36816
Moderate-severe fast improvement	Specific Phobia Score	0.94	0.92	0.95	0.00	-7.56	48295
Moderate-severe fast improvement	Prescribed Medication (Yes)	0.82	0.76	0.89	0.00	-5.04	14228
Moderate-severe fast improvement	Diagnosis (GAD)	0.89	0.81	0.98	0.02	-2.40	49784
Moderate-severe fast improvement	Diagnosis (Other Anxiety)	0.80	0.72	0.90	0.00	-3.82	46004
Moderate-severe fast improvement	Diagnosis (OCD)	0.45	0.36	0.55	0.00	-7.59	50654
Moderate-severe fast improvement	Diagnosis (PTSD)	0.50	0.42	0.60	0.00	-7.77	49616
Moderate-severe fast improvement	Diagnosis (Other)	0.75	0.59	0.96	0.02	-2.30	49159
Moderate-severe fast improvement	Ethnicity (Minoritised Ethnic Groups)	1.09	0.95	1.26	0.22	1.24	220
Moderate-severe fast improvement	Disability (Yes)	0.77	0.67	0.89	0.00	-3.70	68
Moderate-severe fast improvement	Employment (Non-worker)	0.78	0.71	0.87	0.00	-4.66	6354
Moderate-severe fast improvement	Employment (Unemployed)	0.65	0.58	0.72	0.00	-7.69	3201

Moderate-severe fast improvement	Age (10 years)	1.31	1.27	1.35	0.00	16.21	7251
Moderate-severe fast improvement	Gender (Male)	1.02	0.94	1.11	0.62	0.49	40387
Mild, small improvement	(Intercept)	338.35	286.39	399.74	0.00	68.47	18395
Mild, small improvement	Anxiety Symptoms (GAD7)	0.81	0.80	0.81	0.00	-54.32	47137
Mild, small improvement	WSAS Home	0.84	0.82	0.85	0.00	-19.78	28224
Mild, small improvement	WSAS Social Leisure	0.92	0.90	0.94	0.00	-9.01	39834
Mild, small improvement	WSAS Private Leisure	0.86	0.85	0.88	0.00	-17.51	39180
Mild, small improvement	WSAS Relationships	0.93	0.91	0.94	0.00	-8.84	43166
Mild, small improvement	Agoraphobia Score	0.96	0.94	0.97	0.00	-5.56	41324
Mild, small improvement	Social Phobia Score	0.90	0.88	0.91	0.00	-12.54	44712
Mild, small improvement	Specific Phobia Score	1.00	0.98	1.01	0.51	-0.65	22378
Mild, small improvement	Prescribed Medication (Yes)	0.66	0.62	0.70	0.00	-12.60	18418
Mild, small improvement	Diagnosis (GAD)	2.97	2.74	3.22	0.00	26.10	50710
Mild, small improvement	Diagnosis (Other Anxiety)	2.74	2.51	3.00	0.00	22.04	46680
Mild, small improvement	Diagnosis (OCD)	2.28	1.98	2.64	0.00	11.33	50523
Mild, small improvement	Diagnosis (PTSD)	0.91	0.79	1.05	0.19	-1.31	48692
Mild, small improvement	Diagnosis (Other)	1.80	1.48	2.19	0.00	5.86	50711
Mild, small improvement	Ethnicity (Minoritised Ethnic Groups)	0.97	0.86	1.09	0.58	-0.56	324
Mild, small improvement	Disability (Yes)	0.77	0.68	0.86	0.00	-4.70	77
Mild, small improvement	Employment (Non-worker)	0.83	0.76	0.90	0.00	-4.45	4426
Mild, small improvement	Employment (Unemployed)	0.59	0.54	0.65	0.00	-11.23	3393
Mild, small improvement	Age (10 years)	1.18	1.15	1.21	0.00	11.38	8172
Mild, small improvement	Gender (Male)	0.95	0.89	1.02	0.19	-1.31	33139

Supplementary Table 5. Multinomial regression output of four-class growth mixture model of anxiety symptoms (GAD7)

Reference class: Moderate-severe plateau.

Class	Baseline Variable	OR	Lower CI	Upper CI	*p*-value	Statistic	df
Moderate-severe gradual improvement	(Intercept)	1.92	1.64	2.25	0.00	8.07	21277
Moderate-severe gradual improvement	Depression Symptoms (PHQ9)	1.00	1.00	1.01	0.22	1.21	47050
Moderate-severe gradual improvement	WSAS Home	0.99	0.98	1.01	0.48	-0.71	42647
Moderate-severe gradual improvement	WSAS Social Leisure	0.97	0.96	0.99	0.00	-2.86	45048
Moderate-severe gradual improvement	WSAS Private Leisure	0.98	0.96	0.99	0.00	-3.07	46074
Moderate-severe gradual improvement	WSAS Relationships	0.98	0.97	1.00	0.02	-2.28	45337
Moderate-severe gradual improvement	Agoraphobia Score	1.00	0.98	1.02	0.99	-0.02	43341

Moderate-severe gradual improvement	Social Phobia Score	0.99	0.98	1.01	0.33	-0.97	47263
Moderate-severe gradual improvement	Specific Phobia Score	0.97	0.96	0.98	0.00	-4.53	43247
Moderate-severe gradual improvement	Prescribed Medication (Yes)	0.97	0.91	1.03	0.32	-1.00	15353
Moderate-severe gradual improvement	Diagnosis (GAD)	1.26	1.16	1.36	0.00	5.66	49076
Moderate-severe gradual improvement	Diagnosis (Other Anxiety)	1.12	1.02	1.22	0.01	2.45	49315
Moderate-severe gradual improvement	Diagnosis (OCD)	0.85	0.75	0.97	0.01	-2.46	50711
Moderate-severe gradual improvement	Diagnosis (PTSD)	0.79	0.69	0.90	0.00	-3.49	50522
Moderate-severe gradual improvement	Diagnosis (Other)	1.06	0.86	1.30	0.60	0.52	50141
Moderate-severe gradual improvement	Ethnicity (Minoritised Ethnic Groups)	1.02	0.91	1.15	0.71	0.38	299
Moderate-severe gradual improvement	Disability (Yes)	0.90	0.82	1.00	0.05	-2.00	119
Moderate-severe gradual improvement	Employment (Non-worker)	0.90	0.83	0.98	0.01	-2.50	12017
Moderate-severe gradual improvement	Employment (Unemployed)	0.76	0.70	0.83	0.00	-6.04	14573
Moderate-severe gradual improvement	Age (10 years)	1.00	1.00	1.00	0.36	0.92	20301
Moderate-severe gradual improvement	Gender (Male)	1.01	0.94	1.09	0.72	0.36	32699
Moderate-severe fast improvement	(Intercept)	6.06	5.14	7.13	0.00	21.56	30165
Moderate-severe fast improvement	Depression Symptoms (PHQ9)	0.97	0.96	0.98	0.00	-8.18	49561
Moderate-severe fast improvement	WSAS Home	0.96	0.94	0.98	0.00	-4.26	37411
Moderate-severe fast improvement	WSAS Social Leisure	0.93	0.92	0.95	0.00	-7.11	43410
Moderate-severe fast improvement	WSAS Private Leisure	0.93	0.92	0.95	0.00	-7.86	38685
Moderate-severe fast improvement	WSAS Relationships	0.93	0.91	0.95	0.00	-8.30	40089
Moderate-severe fast improvement	Agoraphobia Score	0.96	0.95	0.98	0.00	-4.26	46821
Moderate-severe fast improvement	Social Phobia Score	0.94	0.92	0.95	0.00	-7.06	46599
Moderate-severe fast improvement	Specific Phobia Score	0.94	0.92	0.95	0.00	-8.83	48532
Moderate-severe fast improvement	Prescribed Medication (Yes)	0.84	0.78	0.90	0.00	-5.17	26184
Moderate-severe fast improvement	Diagnosis (GAD)	1.03	0.95	1.12	0.48	0.71	49810
Moderate-severe fast improvement	Diagnosis (Other Anxiety)	1.04	0.95	1.14	0.41	0.82	48412
Moderate-severe fast improvement	Diagnosis (OCD)	0.40	0.34	0.46	0.00	-11.75	51285

Moderate-severe fast improvement	Diagnosis (PTSD)	0.56	0.48	0.65	0.00	-7.35	49390
Moderate-severe fast improvement	Diagnosis (Other)	1.08	0.87	1.34	0.47	0.72	49637
Moderate-severe fast improvement	Ethnicity (Minoritised Ethnic Groups)	1.04	0.92	1.19	0.51	0.66	185
Moderate-severe fast improvement	Disability (Yes)	0.81	0.72	0.91	0.00	-3.54	83
Moderate-severe fast improvement	Employment (Non-worker)	0.81	0.74	0.88	0.00	-4.75	4872
Moderate-severe fast improvement	Employment (Unemployed)	0.69	0.63	0.77	0.00	-7.02	3275
Moderate-severe fast improvement	Age (10 years)	1.02	1.01	1.02	0.00	11.77	13755
Moderate-severe fast improvement	Gender (Male)	1.07	0.99	1.15	0.10	1.67	40003
Mild, small improvement	(Intercept)	302.55	258.97	353.46	0.00	71.98	28627
Mild, small improvement	Depression Symptoms (PHQ9)	0.78	0.77	0.78	0.00	-68.05	48839
Mild, small improvement	WSAS Home	1.05	1.03	1.07	0.00	5.16	46943
Mild, small improvement	WSAS Social Leisure	0.94	0.92	0.96	0.00	-6.84	46450
Mild, small improvement	WSAS Private Leisure	0.92	0.91	0.94	0.00	-9.52	47310
Mild, small improvement	WSAS Relationships	0.93	0.91	0.94	0.00	-8.91	46953
Mild, small improvement	Agoraphobia Score	0.90	0.89	0.92	0.00	-13.13	49708
Mild, small improvement	Social Phobia Score	0.98	0.96	0.99	0.00	-2.93	43681
Mild, small improvement	Specific Phobia Score	0.94	0.92	0.95	0.00	-9.59	46496
Mild, small improvement	Prescribed Medication (Yes)	0.97	0.91	1.03	0.30	-1.04	12840
Mild, small improvement	Diagnosis (GAD)	0.34	0.31	0.37	0.00	-26.79	48435
Mild, small improvement	Diagnosis (Other Anxiety)	0.50	0.45	0.54	0.00	-15.54	50572
Mild, small improvement	Diagnosis (OCD)	0.16	0.14	0.18	0.00	-26.89	51124
Mild, small improvement	Diagnosis (PTSD)	0.39	0.34	0.45	0.00	-12.92	48675
Mild, small improvement	Diagnosis (Other)	0.84	0.69	1.03	0.09	-1.70	47992
Mild, small improvement	Ethnicity (Minoritised Ethnic Groups)	1.13	1.02	1.26	0.02	2.30	1083
Mild, small improvement	Disability (Yes)	0.92	0.83	1.03	0.13	-1.51	101
Mild, small improvement	Employment (Non-worker)	1.03	0.95	1.11	0.51	0.66	8372
Mild, small improvement	Employment (Unemployed)	0.98	0.89	1.07	0.66	-0.43	6637
Mild, small improvement	Age (10 years)	1.01	1.01	1.01	0.00	7.41	15861
Mild, small improvement	Gender (Male)	1.33	1.24	1.43	0.00	7.99	42463

Supplementary Table 6. Multinomial regression output of four-class growth mixture model of functional impairment symptoms (WSAS)

Reference class: Moderate-severe plateau.

Class	Baseline Variable	OR	Lower CI	Upper CI	*p*-value	Statistic	df
Moderate-severe gradual improvement	(Intercept)	0.22	0.17	0.30	0.00	-10.15	29132
Moderate-severe gradual improvement	Anxiety Symptoms (GAD7)	1.01	0.99	1.02	0.49	0.70	30700
Moderate-severe gradual improvement	Depression Symptoms (PHQ9)	1.01	0.99	1.02	0.24	1.19	28310
Moderate-severe gradual improvement	Agoraphobia Score	0.98	0.95	1.00	0.08	-1.78	31294
Moderate-severe gradual improvement	Social Phobia Score	1.00	0.97	1.02	0.76	-0.31	30700
Moderate-severe gradual improvement	Specific Phobia Score	0.97	0.95	0.99	0.01	-2.51	30598
Moderate-severe gradual improvement	Prescribed Medication (Yes)	0.97	0.87	1.09	0.66	-0.44	13483
Moderate-severe gradual improvement	Diagnosis (GAD)	1.13	0.98	1.31	0.09	1.69	31727
Moderate-severe gradual improvement	Diagnosis (Other Anxiety)	0.91	0.77	1.07	0.25	-1.16	31837
Moderate-severe gradual improvement	Diagnosis (OCD)	0.75	0.56	1.00	0.05	-1.99	32024
Moderate-severe gradual improvement	Diagnosis (PTSD)	0.92	0.71	1.19	0.53	-0.63	31686
Moderate-severe gradual improvement	Diagnosis (Other)	0.44	0.29	0.68	0.00	-3.69	31796
Moderate-severe gradual improvement	Ethnicity (Minoritised Ethnic Groups)	0.98	0.79	1.20	0.82	-0.23	755
Moderate-severe gradual improvement	Disability (Yes)	0.79	0.64	0.99	0.04	-2.08	68
Moderate-severe gradual improvement	Employment (Non-worker)	0.71	0.57	0.88	0.00	-3.04	19702
Moderate-severe gradual improvement	Employment (Unemployed)	0.49	0.40	0.59	0.00	-7.17	4628
Moderate-severe gradual improvement	Age (10 years)	1.06	1.01	1.12	0.02	2.33	20591
Moderate-severe gradual improvement	Gender (Male)	1.10	0.97	1.24	0.15	1.45	27550
Moderate-severe fast improvement	(Intercept)	0.59	0.45	0.77	0.00	-3.85	31535
Moderate-severe fast improvement	Anxiety Symptoms (GAD7)	1.00	0.99	1.02	0.69	0.40	31895
Moderate-severe fast improvement	Depression Symptoms (PHQ9)	0.98	0.96	0.99	0.00	-3.81	31710
Moderate-severe fast improvement	Agoraphobia Score	1.00	0.97	1.03	0.94	-0.08	31913
Moderate-severe fast improvement	Social Phobia Score	0.88	0.86	0.90	0.00	-9.23	31706
Moderate-severe fast improvement	Specific Phobia Score	0.92	0.89	0.94	0.00	-7.09	31375

Moderate-severe fast improvement	Prescribed Medication (Yes)	0.88	0.79	0.98	0.02	-2.39	22821
Moderate-severe fast improvement	Diagnosis (GAD)	1.08	0.94	1.24	0.26	1.12	31749
Moderate-severe fast improvement	Diagnosis (Other Anxiety)	0.95	0.81	1.12	0.55	-0.60	31340
Moderate-severe fast improvement	Diagnosis (OCD)	0.59	0.44	0.79	0.00	-3.47	31990
Moderate-severe fast improvement	Diagnosis (PTSD)	0.83	0.64	1.09	0.18	-1.35	31256
Moderate-severe fast improvement	Diagnosis (Other)	0.69	0.49	0.96	0.03	-2.22	31664
Moderate-severe fast improvement	Ethnicity (Minoritised Ethnic Groups)	1.11	0.90	1.37	0.32	1.00	265
Moderate-severe fast improvement	Disability (Yes)	0.71	0.58	0.87	0.00	-3.35	101
Moderate-severe fast improvement	Employment (Non-worker)	0.67	0.54	0.84	0.00	-3.43	12303
Moderate-severe fast improvement	Employment (Unemployed)	0.41	0.33	0.50	0.00	-8.29	6477
Moderate-severe fast improvement	Age (10 years)	1.18	1.12	1.24	0.00	6.60	29625
Moderate-severe fast improvement	Gender (Male)	1.16	1.03	1.30	0.01	2.47	30894
Mild, small improvement	(Intercept)	76.36	64.67	90.16	0.00	51.13	19845
Mild, small improvement	Anxiety Symptoms (GAD7)	0.97	0.97	0.98	0.00	-6.13	30766
Mild, small improvement	Depression Symptoms (PHQ9)	0.87	0.86	0.88	0.00	-35.43	30421
Mild, small improvement	Agoraphobia Score	0.95	0.93	0.96	0.00	-6.76	27421
Mild, small improvement	Social Phobia Score	0.81	0.79	0.82	0.00	-25.45	29225
Mild, small improvement	Specific Phobia Score	0.93	0.92	0.95	0.00	-9.46	25254
Mild, small improvement	Prescribed Medication (Yes)	0.83	0.77	0.88	0.00	-5.69	23373
Mild, small improvement	Diagnosis (GAD)	1.40	1.29	1.52	0.00	7.79	29984
Mild, small improvement	Diagnosis (Other Anxiety)	1.31	1.20	1.44	0.00	5.75	30961
Mild, small improvement	Diagnosis (OCD)	0.86	0.74	1.00	0.05	-1.94	30223
Mild, small improvement	Diagnosis (PTSD)	0.89	0.76	1.04	0.15	-1.43	30024
Mild, small improvement	Diagnosis (Other)	0.84	0.70	1.01	0.06	-1.86	30374
Mild, small improvement	Ethnicity (Minoritised Ethnic Groups)	0.85	0.75	0.96	0.01	-2.52	401
Mild, small improvement	Disability (Yes)	0.72	0.64	0.82	0.00	-5.23	78
Mild, small improvement	Employment (Non-worker)	0.72	0.64	0.82	0.00	-5.24	13373
Mild, small improvement	Employment (Unemployed)	0.44	0.39	0.49	0.00	-14.13	3814
Mild, small improvement	Age (10 years)	1.07	1.04	1.11	0.00	4.40	13263
Mild, small improvement	Gender (Male)	1.06	0.99	1.14	0.11	1.59	23368

References

1. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999 Jan 1;6(1):1–55.
2. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Struct Equ Modeling. 2007 Oct 23;14(4):535–69.
3. Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978 Mar;6(2):461–4.
4. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010; 6:109–38.
5. Muthén BO. Beyond SEM: General latent variable modeling. Behaviormetrika. 2002 Jan;29(1):81–117.
6. Bauer DJ, Curran PJ. The integration of continuous and discrete latent variable models: potential problems and promising opportunities. Psychol Methods. 2004 Mar;9(1):3–29.
7. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Social and Personality Psychology Compass. 2008;2:302–17.
8. van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. Struct Equ Modeling. 2017 May 4;24(3):451–67.
9. Petras H, Masyn K. General Growth Mixture Analysis with Antecedents and Consequences of Change. In: Piquero AR, Weisburd D, editors. Handbook of Quantitative Criminology. Springer, New York, NY; 2010. p. 69–100.
10. Meyer JP, Morin AJS. A person-centered approach to commitment research: Theory, research, and methodology: Person-Centered Commitment Research. J Organ Behav. 2016 May;37(4):584–612.
11. Clark SL, Muthén B. Relating latent class analysis results to variables not included in the analysis. 2009;1–55.
12. Asparouhov T, Muthen B. Using Mplus TECH11 and TECH14 to test the number of latent classes. Mplus Web Notes. 2012;14(22):1–17.
13. Wickrama KK, Lee TK, O'Neal CW, Lorenz FO. Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide. Routledge; 2016. 326.