Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Who does iCBT work for and why: predicting and understanding treatment outcomes in depression

Chi Tak Lee

School of Psychology and Trinity College Institute of Neuroscience
Trinity College Dublin, University of Dublin

A thesis submitted to
Trinity College Dublin, University of Dublin
For the degree of
Doctor of Philosophy

2024

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed: Chi Tak Lee
Date: 31/10/2023

Summary

At any given moment, depression directly affects 5% of the population, and indirectly impacts many more others. Effectively tackling this condition has remained at the forefront of psychiatry research. In the challenging landscape of soaring demands for mental health support and limited resources, digital interventions like internet-delivered cognitive behavioural therapy (iCBT) emerge as a accessible, scalable, and cost-effective solution. While iCBT has shown efficacy, it only works for 30-50% of depressed patients. At this time, we still do not have a clear understanding who it best benefits and why. Past endeavours dedicated to this research area have not been fruitful, mainly because of the inadequacy of the current methodologies to comprehensively address the intricate nature of depression. To this end, the thesis proposes a *big data* revolution to our current research infrastructure, by leveraging large, rich patient datasets with advanced statistical modelling to further elucidate the predictors and mechanisms of iCBT.

Chapter 2 sought to investigate the real-world effectiveness of low-intensity psychological interventions, such as iCBT, for a vulnerable subpopulation at high risk of depression – individuals with physical long-term conditions (LTC), which is a known marker in depression treatment response. We retrospectively analysed a large routine care dataset of patients with and without LTC (N=21,051) experiencing depression and anxiety symptoms, who enrolled in the NHS Talking Therapies service in the UK. We found that both groups improved significantly in these interventions, with slightly different patterns; patients with LTC experienced more improvement in depression, while those without had greater improvement in anxiety. However, these differences were small and unlikely to make a clinical difference. Regardless of LTC status, iCBT was significantly more effective in comparison to guided self-help and group therapy. Our findings represent some of initial real-world empirical support for the application of low-intensity psychological interventions to treat comorbid depression and anxiety in patients with LTC. We advocated for a broader integration of these interventions, in particular iCBT, into the care of patients with LTC.

Chapter 3 expanded the search beyond single markers of iCBT treatment response, and emphasised the importance of a multivariable approach to develop reliable predictive models. Historically, the lack of extensive and diverse data has hindered progress in this area. To address this, we tested the feasibility of a novel, fully internet-based

methodological protocol in the *Precision in Psychiatry* (PIP) study, for upscaling data collection in depression treatment prediction research. We remotely gathered longitudinal, observational data from hundreds of participants initiating iCBT (N=600) or antidepressant medication (N=110) in just ~1 year of active recruitment. Via a web-browser, participants provided extensive self-report and cognitive data prior to treatment. Over the 4-week study duration, the PIP study enjoyed high retention and treatment adherence rates, while witnessing significant clinical improvements not only in depression but also across a range of transdiagnostic symptoms in both treatment arms. Despite participants reported being distracted during online assessments, study schedule compliance remained excellent (~1 day between scheduled date vs. completion date of assessments) and data quality was high (near perfect test-retest reliability of self-report height). We highlight here the potential of internet-based methods for gathering rich, longitudinal data at scale and at speed, crucial for advancing the current state of treatment prediction research in psychiatry.

Chapter 4 harnessed the online dataset gathered in chapter 3 to predict early iCBT treatment response using machine learning (ML). We trained and assessed a range of linear and non-linear ML models with feature sets of varying granularity via cross-validation. The best model, an elastic net regression with 31 variables, explained 14.6% variance in depression improvement following 4 weeks of iCBT. It surpassed a clinical consensus for significance and outperformed a benchmark model of baseline severity, sex, and age. The top performing predictors comprised mainly self-report data; the most important being baseline depression, followed by treatment expectation, and a range of transdiagnostic clinical symptoms, general health and lifestyle, demographics, and environment variables. The inclusion of cognitive variables, while less prominent, underscored the value of a multimodal approach in predicting early iCBT response. Importantly, the model generalised well on hold-out iCBT data (18.5% variance explained), but its predictions were not iCBT-specific, as it also did comparably well on a hold-out antidepressant medication sample (17.7% variance explained). Findings from the PIP study contributed to the evolving field of personalised medicine in psychiatry by using data-driven, algorithmic methods, with important insights for optimising iCBT personalisation based on early treatment responses.

Chapter 5 delved deeper into depression symptoms at baseline via the network theory of psychopathology, which predicts that greater symptom network connectivity is indicative

of increased psychological vulnerability, and in turn treatment resistance. With a substantial dataset (N=40,518), we tested the association between cross-sectional baseline network connectivity and iCBT treatment response, and benchmarked its predictive value against baseline depression symptom severity and variance. At baseline, non-responders exhibited greater level of network connectivity than responders; our power analysis revealed this effect was small however, and disappeared after we controlled for differences in baseline depression sum score variance (not severity) between-groups. Baseline depression severity and variance also had larger effect sizes predicting iCBT treatment response than network connectivity and symptom centrality. We replicated these results in patients completing longer iCBT treatment (N=22,952) and using anxiety symptom networks (N=70,620). Our findings put into question the prognostic utility of cross-sectional network metrics for depression treatment response. Rather, the study encouraged the adoption of personalised networks derived from time-series data to better uncover the therapeutic mechanism of iCBT.

In sum, this thesis concludes with important findings that have substantial implications for both research and clinical practices related to iCBT. More specifically, our work has illuminated key insights into real-world prescription and allocation of iCBT, the prediction of early iCBT treatment response through algorithmic-driven tools, and an improved understanding of iCBT response via group-level depression network symptom analysis. The work in this thesis represents a significant stride towards big data approaches in psychiatry research, which hold the promise of unravelling the complexity and heterogeneity of depression. We believe this step forward has the potential to greatly improve predictive and explanatory insights into personalised depression treatment prediction.

List of Publications and Presentations

This thesis incorporates material already published in the following manuscripts:

- **Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M. (2023).** The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry*, *23*(1), 25.
- **Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M. (2023).** Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression. *Psychological Medicine*, 1-10.
- **Lee, C. T., Harty, S., Adegoke, A., Palacios, J., Gillan, C. M., & Richards, D. (2023).** The Effectiveness of Low-Intensity Psychological Interventions for Comorbid Depression and Anxiety in Patients with Long-Term Conditions: A Real-World Naturalistic Observational Study in IAPT Integrated Care. *International Journal of Behavioral Medicine*, 1-11.

The following are presentations arising from this work:

- **Lee, C. T., Whelan, R., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** Bringing Precision in Psychiatry (PIP) Online: A large-scale internet-based investigation of self-report and cognitive predictors of early response to depression treatment. Invited Symposium at the European Society for Research on Internet Intervention**,** Amsterdam, The Netherlands, 2023.
- **Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M.** Just how *useful* are networks in the real world? Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression. Oral Presentation at the Society of Psychotherapy Research 54th International Annual Meeting, Dublin, Ireland, 2023.
- **Lee, C. T., Whelan, R., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** Bringing Precision in Psychiatry (PIP) online: A large-scale internet-based investigation of self-report and cognitive predictors of early response to depression treatment. Poster Spotlight Presentation at the Computational Psychiatry Conference 2023, Dublin, Ireland. 2023.
- **Lee, C. T., Whelan, R., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** Bringing Precision in Psychiatry (PIP) online: Predicting and understanding treatment outcomes in internet-delivered cognitive behavioural therapy for depression and anxiety. Invited Symposium at the Society of Biological Psychiatry 2023 Annual Meeting**,** San Diego, California, US, 2023.

- **Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** The Precision in Psychiatry (PIP) Study: An internet-based methodology for accelerating research in treatment prediction and personalisation. Poster Presentation at The British Association for Psychopharmacology Summer Meeting, London, UK, 2022.
- **Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M.** Association of depression symptom network connectivity and treatment response to internet-delivered cognitive behavioural therapy Invited Oral Presentation at The British Association for Psychopharmacology Summer Meeting, Virtual, 2021.
- **Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M.** Association of depression symptom network connectivity and treatment response to internet-delivered cognitive behavioural therapy. Poster Presentation at the Treatment Selection Idea Lab (TSIL) - Two Days of Precision Medicine in Mental Health, Virtual, 2021.
- **Lee, C. T., Kelley, S. W., Palacios, J., Richards, D., & Gillan, C. M.** Association of depression symptom network connectivity and treatment response to internet-delivered cognitive behavioural therapy. Poster Presentation at the Neuroscience Ireland Conference, Virtual, 2021.
- **Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** The Precision in Psychiatry (PIP) Study: An internet-based methodology to treatment prediction research. Poster Presentation at the MQ Science Summit, Virtual, 2021.
- **Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., O'Keane, V., Stephan, K. E., & Gillan, C. M.** The Precision in Psychiatry (PIP) Study: An internet-based methodology to treatment prediction research. Poster Presentation at the Society of Biological Psychiatry Annual Meeting, Virtual, 2021.

Acknowledgments

First and foremost, I want to express my sincere gratitude to my supervisor, Dr. Claire Gillan, for her unwavering support, patience, and guidance throughout this journey. Claire, your exceptional intellect, wit, and charisma have been a constant source of inspiration to me, and I have grown so much as a young scholar under your mentorship. You always encouraged and pushed me to reach for the highest standard in everything that I do. In (plenty) moments of self-doubt, you lent me a listening ear and gave me invaluable advice. You convinced me to believe in my capabilities, even when I struggled to do so myself. You have been a true role model to me, and I have no doubt that your influence will continue to shape my journey in the future. I didn't quite believe you when you first said to me, "3 years of PhD is going to absolutely fly by". Now, I've reached the end after what felt like a mere blink of an eye. I'm sure you are only pleased to see me graduating sooner rather than later (so that full professorship can undoubtedly be in your grasp!). I very much look forward to crossing paths once again in our future endeavours.

I would like to thank the Irish Research Council and SilverCloud Health for granting me the valuable opportunity and funding to pursue this PhD. I would like to thank my thesis examiners Dr. Clare Kelly and Dr. Adam Chekroud, and collaborators Dr. Robert Whelan and Dr. Klass Enno Stephan for dedicating their time and expertise to my work. Thank you to Dr. Jacinta Nolan, for her ongoing tutelage. To the patients who contributed their data and experiences to my research, I am profoundly thankful. A special thank you also goes out to my old, beloved colleagues at SC for cheering me on. I am especially grateful to Caroline, Angel, Dan, Siobhan, Simon, Kate, Selin, and James. To Jorge, I couldn't have asked for a better lifelong mentor and a friend in you. You are indeed the best combination of wisdom, craic, and occasional (good) tequila. To Derek, your guidance has been indispensable. I am deeply thankful to you for believing in me at the start of all this. I'll forever remember how you championed my cause, and opened doors to this PhD for me.

I owe a debt of gratitude to the Gillan Gals & Guys and TCIN folks, old and new. A big thank you to Kevin, Owen, Sean, Tricia, Andy, Jonny, and Eoghan, for welcoming me with open arms when I first started, and continued to support me throughout. Thank you to Celine, for being a wonderful flatmate and a constant source of light in my life. Thank you to Robert and Alice, for helping out with the last stages of the thesis. Thank you to

Yihe, Eike (BBG), Tom, Kelly, Julia, Jiv, Sara, Daniel, Declan, Alannah, Clodagh, Anna H, Steph, Anna R, Vanessa, Sol, Judith, Parnian, Mikey, Jade, Qing, Feng, Huiqing, Bolin, Andrew, Gabe, Dasha, Katy, Men, Luisa, Lexi, Maike for stimulating discourse, well-needed emotional support, good food and pints, and endless belly laughs. Your camaraderie have made this academic journey enjoyable and memorable.

To my closest and dearest, Sarah, Hannah, Hugh, Jack, Luke, and Will, to this day I have no doubt none of you can actually tell me what I have been doing for the past few years! But it didn't matter. You were there unconditionally, and I am eternally grateful to have you all. To Corinne, Jamie, Rebecca, Jack, and Diego, may the Psycho-Amigos long live. To Brian, may we continue to laugh so much we cry. To Thomas, thank you for being my rock during the good times and the tough times, and when I needed it the most.

My deepest gratitude goes to my family; Mam and Ah Po, your love and sacrifices kept me going and grounded this whole time. I dedicate this thesis to the both of you. I hope I have made you proud.

Table of Contents

Chapter 1 – General Introduction

1.1 Prevalence and Impact of Depression

With approximately 280 million people impacted worldwide (5% of the population at any given time), depression is a commonly experienced, debilitating condition that negatively affects how an individual feels, thinks, and acts (World Health Organization, 2022). In our daily lives, we often use the term to describe profound feelings of sadness or unease. Many people may recall experiencing moments of feeling 'depressed'; perhaps as an immediate consequence to a significant, traumatic event such as the loss of a loved one, or as feelings that gradually develop over time in response to ongoing stresses, such as chronic pain. However, an important distinction exists, at least medically, between these sorts of common emotional fluctuations and the classification of Major Depressive Disorder (MDD). This shift occurs when several negative emotions and symptoms occur at once, are frequent, persistent, and hinder normal functioning. As per the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-V), an individual is diagnosed with depression when they experience persistent low mood and/or loss of interest and pleasure in activities for a minimum duration of 2 weeks. Additionally, they must also exhibit a combination of other symptoms (5 out of 9) including significant change in weight or appetite, insomnia or hypersomnia, psychomotor agitation, excessive feelings of guilt or worthlessness, difficulty concentrating, or suicidal ideation (American Psychiatric Association, 2013). Finally, the symptoms should cause significant distress or impairment.

When considering the lifetime prevalence of depression, it can vary depending on interregional, sample, and measurement differences. According to a systematic review collating evidence from 63 studies on the epidemiology of depression worldwide, the estimate of lifetime prevalence ranged from 2% to 21%, with the highest rates found in some European countries and the lowest in some Asian countries (Gutiérrez-Rojas et al., 2020). However, it is fair to say depressive symptoms are ubiquitously present throughout the general population, even when a person does not meet the criteria for a formal diagnosis. A study that analysed a national data sample in the US between 2005-2008 revealed over 26% of adults endorsed mild depressive symptoms and approximately 12% endorsed moderate depressive symptoms (Shim et al., 2011). Depression symptoms are also highly comorbid; meaning they are often experienced along with other psychiatric

symptoms (Watson, 2009). A 2014 survey of mental health and wellbeing in England found that 1 in 6 people aged 16+ reported experiencing depression and anxiety symptoms, an increase in prevalence since 1993 (McManus et al., 2016).

Susceptibility to the disorder, however, is dependent on numerous demographic and environmental factors. For instance, the lifetime prevalence of depression tends to vary across different life stages, with the highest observed in adulthood (Kessler et al., 2012). Higher rates of depression have been found in females compared to males (Salk et al., 2017; van de Velde et al., 2010) and those with lower socioeconomic status (Freeman et al., 2016; Lorant et al., 2003) and educational attainment (Bauldry, 2015; Cohen et al., 2020). Individuals at risk of developing depression also typically report experiencing heightened levels of perceived loneliness and social isolation (Lakey & Cronin, 2008; Wickramaratne et al., 2022), childhood adversity (Liu, 2017; Teicher et al., 2022), and increase in stressful events (Stroud et al., 2008). While it is evident that the vulnerability to depression relates to a wide range of individual differences, so does one's resilience; certain individuals can become and remain depressed easily, while others appear to have the ability to recover swiftly from it (Dai & Smith, 2023).

On a population level, depression is the single largest mental health contributor and the second overall leading cause of disability (World Health Organisation, 2017). It is responsible for the highest proportion of disability-adjusted life years (40.5% of the total burden) among all mental health and substance use disorders (Whiteford et al., 2013). Its pervasive effect is inarguably substantial and multifaceted on both a societal and personal level. For an individual, it can lead to lowered work productivity and absenteeism, interpersonal relationship problems, substance misuse, reduced life expectancy and suicide (Lépine & Briley, 2011; Vos et al., 2017). The high disease burden of depression further strains the workplace and healthcare system, and in turn translates into incremental financial burden. In the U.S. alone, direct and indirect costs of depression have been estimated at $326.2 billion in 2018, which has risen ~40% from 2010, attributable to both increases in disease prevalence and rising costs per person (Greenberg et al., 2021). These significant consequences have placed the investigation of effective interventions for depression at the forefront of mental health research, with the aim to develop new and more potent interventions, improve existing ones, and personalise their delivery (Patel et al., 2010).

1.2 Treatments for Depression

Currently, several evidence-based treatment options are available to individuals with depression, with the most common being pharmacological and therapy-based solutions. The National Institute of Clinical Excellence (NICE) recommend antidepressant medication as the first line of treatment (NICE, 2020). Commonly prescribed antidepressants include selective serotonin reuptake inhibitors (SSRIs; e.g., sertraline, citalopram) and serotonin-norepinephrine reuptake inhibitors (SNRIs; e.g., venlafaxine, duloxetine), and in rarer instances, atypical (e.g., mirtazapine, vortioxetine) and tricyclic antidepressants (TCAs; amitriptyline, nortriptyline). At the most basic level, these medications work by increasing the availability of serotonin (and norepinephrine), which are neurotransmitters responsible for mood regulation (Harmer et al., 2017). A network meta-analysis investigating the efficacy of 21 antidepressants revealed that all of them were more efficacious in reducing depressive symptoms than placebo in adults with major depressive disorder, with considerable acceptability and tolerability (Cipriani et al., 2018). In more severe cases, antidepressant polypharmacy (i.e., patients receiving 2 or more antidepressants simultaneously) may optimise treatment effect. This approach may achieve better results due to the engagement of multiple neurotransmitter systems, targeting different types of symptoms experienced by the patient in more severe and complex cases. Empirical evidence highlighting the therapeutic benefits of antidepressant polypharmacy over monotherapy, however, has been quite mixed (Nelson et al., 2004). Moreover, antidepressants often cause unwanted side effects, which could be further exacerbated when taking multiple medications at the same time. Research suggest that side effects are one of the main reasons why patients drop out of treatment prematurely (Kostev et al., 2014; MacGillivray et al., 2003).

For these reasons and more, a considerable proportion of individuals with depression express a strong preference for psychotherapy over pharmacotherapy (70%; McHugh et al., 2013). Cognitive Behavioural Therapy (CBT), among various psychotherapeutic approaches (e.g., psychodynamic, acceptance-based, interpersonal), is the most popular and extensively examined in depression. CBT is a relatively short-term therapy that targets the interconnected cycle of negative thoughts, behaviours, and emotions. It does this by equipping the individual with a toolbox of practical strategies and skills to help modify their unhelpful way of thinking and behaving. Treatment protocols typically include psychoeducation (i.e., increase self-awareness of problem and the CBT model of

therapy), cognitive restructuring (i.e., identifying challenging negative beliefs), and behavioural activation (i.e., strategies to engage in pleasurable and meaningful activities), and problem-solving (Dobson & Dozois, 2021). Despite being the most active area of psychotherapy research with more than 50% of randomised controlled trials (RCT) focusing on it (Cuijpers et al., 2020a), several meta-analyses have demonstrated little to no indication that CBT is more effective than other psychotherapies in reducing depressive symptoms (when compared with usual care, waitlist, or pill placebo) (Barth et al., 2013; Cuijpers et al., 2021b). Another large network meta-analysis of 101 RCTs (11,910 patients) also provided strong evidence that psychotherapy works just as well as antidepressant medication in the short term and can be more effective in the long run, but combining both treatments did better than either intervention alone (Cuijpers et al., 2020c).

All things considered, there is a compelling rationale for prescribing psychotherapy over pharmacological treatments, but the trend in clinical practice is the opposite; the number of individuals with depression receiving psychotherapy has declined over time, while drug prescriptions are rising (Sreeharan et al., 2013). Some posit that this may be due to limited availability of and access to evidence-based psychological treatments (Kazdin & Blase, 2011; McHugh & Barlow, 2010). Individual psychotherapy, which is the most dominant model, is very costly for healthcare systems given the time and expertise involved in its delivery (Kazdin & Blase, 2011). Issues with cost are compounded by major shortages in mental healthcare professionals globally, in particular in low- and middle-income countries (Saraceno et al., 2007; Saxena et al., 2007). While efforts to increase the number of service providers might help, it is unlikely the supply of services will ever meet demand (e.g., 700,000 mental health professionals working for approximately 75 million people in need in the US; Kazdin & Blase, 2011). The geographical distribution of service providers is also skewed. The concentration of mental health professions is highest in densely populated, affluent urban cities, and systemically fails to reach people in smaller, socioeconomically disadvantaged and rural areas (Kazdin & Blase, 2011).

1.2.1 Internet-Delivered Cognitive Behavioural Therapy (iCBT)

To remedy this situation, for the last two decades, mental health care for depression has undergone a technological transformation. Internet-delivered CBT (iCBT) has become a fast-growing, scalable, and affordable alternative to conventional psychotherapy

(Andersson et al., 2019c; Kumar et al., 2017). iCBT is delivered via online software platforms accessible on computers, smartphone, and/or tablets. These online programs typically provide therapeutic content in rich multimedia formats like text, video, audio, along with homework assignments, teaching core cognitive and behavioural skills to help patients manage their depression symptoms. The platform often administers assessment tools in the form of self-report questionnaires measuring emotional wellbeing and psychiatric symptoms to monitor treatment progress (Andersson et al., 2019c). Depending on the type of program and clinical presentation, intervention durations can range from a single session (Schleider et al., 2020) to several weeks (i.e., ~8-12 weeks for a typical course) (Furukawa et al., 2021). iCBT can be unguided (i.e., self-led) or guided (i.e., clinician-led). Guided iCBT typically involves a trained clinician to provide guidance and feedback through regular online and/or telephone review sessions to the patient as they advance through the treatment. While self-guided iCBT is a more scalable option, research indicates that patients tend to achieve more favourable therapeutic outcomes and retention rates from online interventions when clinician guidance is involved (Karyotaki et al., 2021). In fact, research shows that delivery formats of CBT does not impact the effectiveness of the intervention (i.e., individual, group, online, telephone) (Cuijpers et al., 2019), as long as there is human support (Cuijpers et al., 2023). Treatment adherence rates between guided iCBT and face-to-face CBT for depression are also comparable, in particular when treatment duration is specified to the patients (van Ballegooijen et al., 2014).

But there are several advantages iCBT has over face-to-face CBT. First, by delivering the intervention online, this means treatments can be made more widely accessible (as long as there is internet connection) from the comfort and convenience of one's own home, thus eliminating lengthy and burdensome travel for appointments and reducing some of the systematic inequities in access. Second, and perhaps most importantly, iCBT is cost-effective as it dramatically reduces human resource needs and premises costs. It is associated with much reduced wait times, as there is no need for patients to wait for in-person therapy sessions to be scheduled, and thus can flexibly and quickly access iCBT online when they feel most motivated to engage in therapy. Indeed, research shows that the amount of therapist time required to deliver in-person CBT is 7.8 times that required in iCBT (Andrews et al., 2018). Third, iCBT may allow clients to therapeutically benefit in a stigma-free way, as they do not need to engage in face to face interactions

(Andersson et al., 2019c; Andersson & Titov, 2014). For psychotherapy research, iCBT may also hold key advantages over face to face therapy (Andersson et al., 2019c). By streamlining diagnostic/screening procedures and psychiatric assessments online, recruitment can be carried out much faster from more diverse samples. One study showed that a web-based intervention in English and Spanish for smoking cessation reached more than 4,000 smokers from 74 countries in just 21 months, illustrating the extensive reach of online interventions (Muñoz et al., 2006). Another notable advantage for research is that digital interventions record objective and granular treatment engagement data as a matter of routine. Both clinicians and researchers can harness this data to monitor treatment progress and to generate therapeutic insights useful for advancing research agenda. A final advantage is that the digital interventions are delivered in a fully systematised manner, removing variability associated with mental health practitioners that add noise into research studies aiming to understand their efficacy.

1.2.2 Heterogeneity in Depression Treatment Response

It is believed that the quality of treatments for depression have significantly improved since the 1980s, and have become increasingly accessible for those in need (Ormel et al., 2022). It is rather surprising however, that despite this, the overall population prevalence of depression has not decreased in proportion to these improvements. This has been coined the *treatment-prevalence paradox* (Ormel et al., 2022). Numerous systematic reviews and meta-analyses have widely established the effectiveness of iCBT in treating depression (Cuijpers et al., 2023) with large effect sizes (Spek et al., 2007). However, as is the case with antidepressant medication (Muñoz et al., 2006; Trivedi et al., 2006) and face-to-face CBT (Cuijpers et al., 2021a; Hofmann, 2012), the efficacy of iCBT varies across individuals, with only roughly half of the patients respond to it (i.e., ~50% symptom reduction) and only a third achieve some sort of meaningful clinical remission (i.e., below an established cut-off for depression) (Andersson et al., 2019a). A big reason for this variability might be that we fail to match the right treatment to the right patient; there are currently no evidence-based tools to assist in the prescription process, forcing clinicians to operate using a combination basis of resource-constraints (availability, cost) and ‘trial-and-error’ when it comes to allocating treatments to patients. With each treatment course typically taking weeks at a time, this process can be both time-consuming and costly. Patients who do not respond to initial treatments often require a series of subsequent interventions, sometimes without improvement, leading to escalating

clinical risks and costs (Al-Harbi, 2012; Trivedi et al., 2006; Warden et al., 2007). The top priority for psychiatry research is to resolve this; leveraging the advancements of big data analytics and technology, to prescribe, predict, and explain iCBT treatment response, improve outcomes, and provide it to the right patient the first time around.

1.3 To Prescribe iCBT – Subgroup Analysis

Determining whether iCBT is the most suitable treatment for each individual remains a complex challenge, but prescribing it to subgroups with shared patient characteristics most likely to benefit from the intervention, can help push this agenda forward. For instance, evidence from NICE supports the use of iCBT to manage mild to moderate depressive symptomatology (NICE, 2020); various healthcare service providers such as Talking Therapies in the UK have incorporated iCBT in their service delivery pathways as a low-intensity treatment for less severe depression (NICE, 2020). Studies found that iCBT also works for all ages of the population including children and adolescents (Vigerland et al., 2016), adults (Andersson et al., 2019b), and the elderly (Dear et al., 2015), with older adults found to benefit just as much from the intervention as their younger counterparts (Mewton et al., 2013; Pabst et al., 2020). Other subpopulations may be more or less likely to benefit from iCBT, for example patients with perinatal depression (Chen et al., 2023; Lee et al., 2016), depression associated with bereavement (Wagner et al., 2020; Zuelke et al., 2021), or with chronic and long-term physical conditions (Beatty & Lambert, 2013; Charova et al., 2015). These complex cases may present new challenges for digital interventions, with physical and situational factors potentially interacting with and undermining the effectiveness of iCBT. With these challenges, however, may come signal for smart prescribing.

1.3.1 Interplay between Physical and Mental Health

Physical and mental health are intrinsically linked and affect each other in many ways. The prevalence of chronic, long-term physical conditions (LTC) are ever increasing, with studies showing they account for 60% of all deaths globally (World Health Organization & Public Health Agency of Canada, 2005). Individuals suffering from LTC such as diabetes, chronic pain, and heart diseases, are particularly vulnerable to experiencing depressive symptoms; the distress, pain, and limitations associated with LTC can significantly contribute to feelings of sadness and hopelessness. Indeed, estimates showed that around 20% of individuals with LTC will develop depression, a risk that is ~2-3

times higher than that of the general population (Katon et al., 2010; Moussavi et al., 2007). In addition to treating the LTC, addressing the accompanying mental health comorbidity is also crucial, as depression can negatively impact and complicate the management and prognosis of these chronic conditions, and vice versa (McManus et al., 2016; Naylor et al., 2012). For instance, patients with LTC who are also struggling with depressive symptoms may find it difficult to self-manage and adhere to treatments effectively, which can impede recovery and lead to further functional impairment (DiMatteo et al., 2000). Conversely, certain medications prescribed for the LTC may have side effects that can worsen depressive symptoms. The longer depression persists and goes untreated, the greater the negative impact it has on an individual's quality of life, encompassing various domains including social, financial, and interpersonal (Beatty & Lambert, 2013; McManus et al., 2016). In the worst case, untreated depression in those with LTC can even lead to elevated risk of mortality (Dossa et al., 2011; Meijer et al., 2011; van Dooren et al., 2013). This reinforces the urgency of addressing the mental well-being of individuals with LTC as a major component of their overall healthcare management.

1.3.2 iCBT for Patients with LTC

There is a clear need for timely, evidence-based interventions that can efficiently address comorbid depression in LTC patients. CBT, of all psychological interventions, has proven particularly effective in managing depression in this cohort; the primary goal is to alleviate depressive symptoms in an effort to lift the burden they may have on a person's self-management (Wroe et al., 2018). According to Lorig & Holman (2003), self-management of chronic diseases requires core cognitive-behavioural skills involving problem-solving, decision-making, perspective taking, and behavioural activation, which are essential components taught in CBT. CBT can help patients adjust to their LTC by altering specific beliefs and attitudes and teaching management strategies (Halford & Brown, 2009). By tailoring the intervention, it may also tackle disease-specific distress associated with particular conditions, that are otherwise distinct from depression but marked by emotional strain linked to disease management. One example is diabetic distress, which is characterised by unique emotional issues directly related to the burden of living with diabetes such as worry, frustration, concern, and aspects of burnout (Marathe et al., 2017). Despite evidence underscoring the effectiveness of CBT for treating comorbid depression in LTC patients, mental health problems are still

underrecognized and undertreated in this cohort. For instance, depression remains untreated in 50% of people with diabetes mellitus (Egede & Ellis, 2010). Perhaps more so than the general population, patients with LTC face significant barriers when it comes to accessing appropriate mental health treatments, such as limited physical capacity and mobility to attend in-person appointments and augmented medical costs for both their physical and mental health needs (Beatty & Lambert, 2013; Naylor et al., 2012; van Beugen et al., 2014). In this regard, the advantage of iCBT over traditional face-to-face CBT is even more pronounced, due to its low-cost, remote nature and extensive reach that allow patients to receive treatment in their own home.

Having said that, evidence in support of iCBT for treating comorbid depression symptoms in people with LTCs is rather limited (Beatty & Lambert, 2013). Firstly, many clinical trials examining iCBT tend to exclude patients with coexisting physical illnesses and their associated effects. Of the ones that do exist, effects have been shown to be small to moderate (van Beugen et al., 2014), just slightly below previously reported for the general population (Cuijpers et al., 2008). But there are also inconsistencies and inadequacies in the evidence; many studies are statistically underpowered and have been classed as having fair to poor methodological quality (Beatty & Lambert, 2013; Charova et al., 2015). Most studies also use a waitlist as a comparator (Cuijpers et al., 2008; Mehta et al., 2018), which tends to overestimate effects and fails to provide insight into the value of iCBT over other forms of treatment. In the few studies that compared iCBT to other CBT-based treatments, there was a lack of evidence strongly supporting one intervention over another (Adhikary et al., 2023). In addition, there is emerging evidence that patients engage in iCBT in a different way inside and outside of clinical trial settings. Specifically, patients are more active and involved in iCBT during RCT, compared to iCBT use in real-world settings, which can inflate their adherence and treatment completion rates, and thereby efficacy estimates (Baumel et al., 2019; Fleming et al., 2018). For this reason, while acknowledging that RCTs are a crucial tool for assessing efficacy, it is important to complement evidence from RCTs with data concerning real world clinical use to assess if iCBT can serve as a low-cost, effective treatment for this subgroup of depression sufferers (Kumar et al., 2017).

Presently, mental and physical health care services often operate independently with minimal coordination, which is not only inconvenient for the patient and costly to the healthcare system, but also likely to produce suboptimal outcomes. In the UK, as part of

their expansion agenda, the Improving Access to Psychological Therapies (IAPT) services have been working to integrate mental health treatments for patients with LTC who experience comorbid depression. The IAPT-LTC pathway seeks to deliver the same IAPT standards of psychological care to patients with LTC within the stepped-care model, with timely referral, assessment, and treatment allocation (National Collaborating Centre for Mental Health, 2018). The IAPT stepped-care model adopts a framework of psychological care that addresses the mental health needs of individuals within the constraints of limited resources. It recognises and considers the severity of the patient's presentation as well as their treatment preference, so that it can prescribe appropriate interventions accordingly. Among the NICE recommended evidence-based therapies, IAPT employs iCBT at step 2 and 3 for treating mild to moderate depressive symptomatology, along with other low-intensity CBT-based interventions such as guided self-help programs (GSH) and psychoeducational group therapy (PGT). In this specific setting, Palacios and colleagues (2023) compared all three of these low-intensity psychological interventions using real-world routine care data gathered from N=21,215 patients. Using propensity score matching to retrospectively control for treatment allocation bias, they observed the largest average treatment effect for iCBT in comparison with GSH and PGT. This is encouraging work underscoring the use of iCBT in a naturalistic, routine care setting, but this study did not examine patients with LTC and comorbid depression separately. Therefore, **chapter 2** of this thesis sought to conduct a large-scale investigation on how effective iCBT is compared to other low-intensity interventions in treating comorbid depressive symptoms in patients with LTC, relative to those who do not have any physical illnesses, in a major real-world healthcare service.

1.4 Markers of iCBT Treatment Response

Efforts in identifying particular subgroups, such as those with physical and psychological comorbidities, who may benefit most from iCBT over other treatments, have laid the foundation for the emerging movement of *personalised medicine* in psychiatry. Personalised medicine refers to an approach to medicine wherein patients are matched to the treatment most likely to benefit them. An aspirational goal for psychiatry, it is hoped that personalisation can improve average response rates, decrease illness durations, and reduce relapse (Cohen & DeRubeis, 2018; Cuijpers et al., 2012). Moreover, if we can better understand which patients might not improve with iCBT before they start treatment, this can help service providers to allocate alternative interventions that may

yield better recovery instead. This endeavour, however, hasn't proved to be easy; for many years, researchers sought (and are still seeking) for specific patient characteristics that may reliably predict depression treatment response based on theoretical groundings (Chekroud et al., 2021). Of these, biomarkers such as gene expression (Hodgson et al., 2012; Mariani et al., 2021), structural/functional neuroimaging scans (Fonseka et al., 2018; Kang & Cho, 2020) and electroencephalogram signals (Watts et al., 2022) have shown promise, but robust associations are lacking. Some looked into environmental markers that have well-established associations with therapeutic outcomes, such as childhood trauma (Kessler et al., 1997; Williams et al., 2016) and stress levels (Hicks et al., 2022; Mazure et al., 2000), which are often overlooked during patient screening and referral processes. The most consistently identified and most predictive factors have been derived from sociodemographic and clinical information (e.g., social support, psychiatric chronicity) (Maj et al., 2020). Unlike the above, these are commonly collected variables in routine care, but the scope of data depends on what clinicians consider as clinically relevant, which limits the extent of information that is available for prognosis in depression treatment (Chekroud et al., 2021).

In the context of iCBT, there is even less agreement on the specific factors that determine whether someone will respond successfully to the intervention (Andersson, 2016; Andersson et al., 2019c; Andersson & Hedman, 2013). One small-sample study (N=73) found that larger pre-treatment right rostral anterior cingulate cortex (ACV) volume predicted depression symptomatology improvement post-iCBT after controlling for demographic and clinical variables (Webb et al., 2018). This is consistent with prior work showing that increased rostral ACV activity predicted depression improvement in pharmacological and psychological treatments (Fu et al., 2013; Nouretdinov et al., 2011). However, collecting biomarkers such as neuroimaging data is resource-intensive and time-consuming, which makes integration into clinical practice challenging. Some have turned to cognitive measures, such as cognitive flexibility and functioning, as proxies for brain-based measurements. Yet, evidence supporting their predictive utility remains inconclusive (Lindner et al., 2016; Silfvernagel et al., 2012). Various demographics and clinical characteristics have been assessed as well, but it is common to find inconsistencies in the predictive value of variables like gender, age, education, marital status, employment, psychiatric history, and general functioning (Button et al., 2012; Edmonds et al., 2018; El Alaoui et al., 2016; Hadjistavropoulos et al., 2016). Perhaps the

single most robust predictor of treatment response is the severity of depression symptoms at baseline, but even so, the direction in which it moderates response is also mixed. Some studies suggest higher baseline severity predicts greater improvement (Button et al., 2012; Edmonds et al., 2018; El Alaoui et al., 2016; Hadjistavropoulos et al., 2016), while others indicate no effect (in psychotherapy more general; Furukawa et al., 2017, 2018; Weitz et al., 2015). How predictive baseline severity is may also dependent on the specific outcome measure due to an intrinsic mathematical coupling of the two: lower baseline severity may predict remission as it is closer to that state, while higher severity predicts greater change because there is more room to improve. Finally, intervention-related factors, such as treatment credibility, expectation of success, and engagement (e.g., frequency of logins, completion rates, clinician contact), have all been individually but inconsistently linked to outcomes in iCBT (Edmonds et al., 2018; El Alaoui et al., 2016; Enrique et al., 2019; Hadjistavropoulos et al., 2016).

Converging findings from different studies can be complicated, given the differences in program features, research settings, statistical methodologies, and sample populations being studied (Beatty & Binnion, 2016). To-date, it is notable that none of these predictors has been adopted for iCBT treatment allocation in clinical practice, because not a single characteristic provides a prediction accurate or robust enough to be clinically meaningful (Chekroud et al., 2021; Gillan & Rutledge, 2021). It is not a problem unique to iCBT; a meta review of 199 reviews conducted on antidepressant medication response revealed significant methodological discrepancies and effect size heterogeneity, making it challenging to establish consistent and widely acceptable predictors (Perlman et al., 2019). This is not surprising; each candidate predictor variable accounts for only a small amount of variance in treatment response. Historically, conventional treatment prediction research has focused on hypotheses testing and post-hoc analyses with respect to a single modality or a limited range of variables. This is useful in identifying potentially important predictors, but it is also limited given the significant heterogeneity that exists in mental health symptom expression, causes, contexts and treatments (Chekroud et al., 2021; Simon & Perlis, 2010). This complexity makes depression challenging to predict and understand from a small set of variables, especially in small-sample studies that are underpowered to reliably estimate what are known to be small effects. To ensure prognostic models produce methodologically sound and valid predictions to be brought forward for clinical implementation, we need to gather a large amount of data from lots of

individuals, ideally followed through time. To this end, a shift towards a data-driven approach that can harness the potential of large, rich datasets, is gaining momentum. It is believed that big data, coupled with advanced statistical techniques can unlock the potential for predictive and explanatory insights in individualised depression treatment prediction.

1.5 To Predict iCBT Response – Machine Learning

Aligned with the data-driven paradigm underpinning precision psychiatry, researchers are shifting their emphasis from hypotheses testing and confirming, to interrogating large and complex datasets and validating insights based on the ability of the model to predict future events in new, unseen datasets (Chekroud et al., 2021). Machine Learning (ML) is fast becoming the most common approach to this. ML algorithms are powerful supervised learning tools that can consider the relationship between variables and outcomes by iteratively and contemporaneously analysing complex, non-linear relationships between multiple variables and the outcome. The individual, weak effects of each variable are aggregated in a way to account for maximal variance in treatment response (Chekroud et al., 2021; Rost et al., 2023). This way, they can produce single best prediction values for each individual, and identify robust and generalisable predictors of treatment response (Cohen & DeRubeis, 2018; Gillan & Whelan, 2017).

1.5.1 Machine Learning Prediction in Depression Treatment Response

There have been considerate efforts to advance this field in recent years, particularly in the area of antidepressant response prediction (Chekroud et al., 2021). One of the earliest examples is a study by Chekroud and colleagues (2016) that tested if self-report data routinely gathered in trial datasets could predict remission to antidepressant medication. They developed a model with just 25 clinical and demographic variables based on the STAR*D dataset (N=1949). The model was able to achieve a modest accuracy of ~60% when tested on external datasets of other medication groups including escitalopram-placebo (N=151) and escitalopram-bupropion (N=134), but was not able to predict remission better than chance (i.e., ~51%) in the venlafaxine-mirtazapine group (N=140). They identified baseline depression severity as the top predictor, and beyond that, depression item/subscale scores (e.g., psychomotor agitation, energy, and sadness) also contributed to the prediction. The implications of this study were important in a number of ways. First, it demonstrated the predictive power of more granular information beyond

sum scores of self-report scales. Second, given the differential predictive performances across medications, it provided early evidence that model predictions may be treatment specific. While one may argue that an accuracy of ~60% may not be of clinical use, this study provided a firm basis for future work, suggesting that including richer baseline datasets, spanning multiple modalities, may enhance prediction (Lee et al., 2018).

A subsequent study (N=280) led by Iniesta and colleagues (2018) expanded upon this work, and showed that models developed using a combination of clinical information (e.g., depression symptoms, stressful life events, and medication status) and genetic markers (e.g., single nucleotide polymorphisms and copy number variants) showed the best performance at predicting remission. Their model predictions were also drug-specific. In external validation, the escitalopram-trained-model yielded better performance predicting escitalopram response with an area under the curve (AUC) of 0.77 compared to nortriptyline (AUC=0.57), and the nortriptyline-model yielded better predictions for nortriptyline (AUC=0.77) than escitalopram (AUC=0.62). The added predictive value and cost-benefit of biomarkers over and beyond self-report data, however, has been debated. One study showed that models that used genetic information without clinical information reported prediction not better than chance (Maciukiewicz et al., 2018). Another study found that model predictions generated from a combination of clinical and biological data (e.g., somatic health measures, inflammatory and metabolic markers) fared better than that from models including only clinical or biological variables in isolation. The largest difference in performance was observed between the combination model and the biological model, while the smallest difference was observed between the combination model and the clinical model (Dinga et al., 2018). Altogether, these results suggest that even though adding biomarkers to prediction models can lead to increases in performance, their additional value on top of clinical data is small. On top of that, considering how costly and practically challenging the collection of biomarker data can be, it may not be appropriate to integrate them in an algorithm-supported treatment allocation process.

Aside from a few gold-standard studies, the majority of research in the field of treatment prediction in psychiatry suffers from significant methodological issues. Recent reviews (Chekroud et al., 2016; Ermers et al., 2020; Lee et al., 2018; Sajjadian et al., 2021) found that most ML prediction studies did not collect data for the purpose of developing a predictive model. Instead, studies most commonly retrospectively reanalysed large,

interventional clinical trial data with a narrow range of variables, which are those typically collected in routine care. Moreover, studies typically did not assess out-of-sample performance (i.e., external validation), essential to ensure that predictions are not specific to the data they were trained on. Without external validation, there may be overfitting - an overestimation of predictive accuracy, which may be further compounded by issues of sample size. As a result, there are growing concerns that the field is overhyped and reliant on publications reporting large effects from small samples, which may not be statistically, methodologically, and clinically sound.

Outside of medication-prediction, even less progress has been made in psychotherapy. A recent scoping review of ML studies attempting to predict response to psychotherapy revealed only 44 studies (mostly proof-of-concepts) have been carried out in the area, of which only 14 had N>200 and only 3 externally validated their models (Aafjes-van Doorn et al., 2021). Psychotherapy trials are notoriously expensive to run and therefore rarely have big enough sample sizes sufficient for ML predictions. To combat this, Buckman and colleagues (2021) collated individual patient data from 6 pre-existing clinical trial datasets (N=1722 for model development, N=918 for model testing), and built 9 models from symptom-level/sum-score data from depression and anxiety symptoms, social support, alcohol support, and life events. They found all models predicted depression severity (not change) at 3-4 months post-treatment better (~14-17% variance explained) than a null baseline depression model (-0.01%), with no clear advantage in using individual items over sum scores. While results are encouraging, the large amount of unexplained variance in the outcome may be due to the model only including a small number of variables not comprehensive enough to capture the biopsychosocial complexity of depression. The overall performance may also have been impacted by the variability of the samples from pooling and retrospectively re-analysing data from different RCTs. Despite RCTs being considered the gold-standard to treatment evaluation, they lack ecological validity as they typically enforce strict eligibility criteria, such as excluding patients with multiple comorbidities. Consequently, these trials tend to involve more homogenous samples, thereby limiting their ability to inform treatment outcomes in real-world care settings (Webb et al., 2020). In this sense, some argue that these predictive models may need to be developed initially with large observational datasets (Aafjes-van Doorn et al., 2021).

1.5.2 Moving Towards a Big Data Approach

To move the field forward from the initial proof-of-concept stage, we need larger datasets with more diverse, richer variables to develop our models. Crucially, studies must include external validation of these models, to ensure that predictions generalise (Ermers et al., 2020). There is a need to invest in methods that can obtain these high quality and large quantity datasets (Aafjes-van Doorn et al., 2021; Ermers et al., 2020). One avenue is through multi-site collaboration via research consortia, where researchers from different countries, academic, and industrial settings team up to collectively advance common goals in treatment prediction research. Few projects like this exist already, including the EMBARC clinical trial (Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care; Trivedi et al., 2016), the PReDICT study (Predictors of Response in Depression to Individual and Combined Treatments; Dunlop et al., 2012), and the iSPOT-D clinical trial (International Study to Predict Optimized Treatment for Depression; Williams et al., 2011). While we cannot overlook the success of amassing large-scale datasets via multi-site research collaborations, it is also extremely time, resource, and cost intensive, with recruitment periods often spanning over years, or even beyond a decade (e.g., 5-11 years in the studies outlined above). To complement this approach, there is a growing need for developing more practical ways to collect the large, rich, and longitudinal data required to progress this field, beyond what traditional, laboratory-based, single-site research can achieve (Gillan & Rutledge, 2021).

In **chapter 3** of this thesis, we describe a novel, online-based methodology that can be used to capture a wide range of interindividual and intraindividual data as patients undergo depression treatment. Researchers have begun to take advantage of technological advancements in longitudinally tracking extensive and diverse data from patients (Gillan & Rutledge, 2021; Rutledge et al., 2019). Internet-based testing facilitates rich data collection, not only comprehensive behavioural and clinical self-report assessments, but also computerised cognitive tasks that can efficiently tap into brain functioning traditionally linked to depression treatment outcomes, as a proxy for brain-based measurements otherwise unfeasible to administer (Gonda et al., 2015; Lam et al., 2014; Saragoussi et al., 2017). Although there are concerns about the reliability of online data collected remotely without direct supervision, it has demonstrated that such data can be trustworthy (Crump et al., 2013; Germine et al., 2012; Shapiro et al., 2013). The larger samples recruited via online testing can also help offset potential increased noise in the

data (Gillan & Rutledge, 2021). Another exciting aspect of applying internet-based methods to depression treatment prediction research is its compatibility with digital interventions like iCBT. Both online research and iCBT complement each other by allowing remote collection of standardised, longitudinal, large-scale data quickly and efficiently in a way that can be streamlined and integrated. Machine learning outputs may also be incorporated into the intervention to enhance its delivery, for example, offering just-in-time risk alerts and intervention to the clinician and patient respectively (Nahum-Shani et al., 2017). In this sense, both digitised research and intervention lend themselves well to the application of ML predictive models. **Chapter 3** elucidates the potential of this synthesis by outlining in detail the methodological protocol of the *Precision in Psychiatry* (PIP) study, which employed a fully internet-based method to recruit, assess, and follow through time, mental health sufferers engaging in iCBT and also receiving antidepressant medication.

1.5.3 Machine Learning Prediction in iCBT Treatment Response

Research predicting treatment outcomes in iCBT using machine learning is increasingly gaining traction, with initial exploratory, small-scale studies showing promising results in obsessive-compulsive disorder (OCD) (N=61; Lenhard et al., 2018), social anxiety (N=26; Månsson et al., 2015), and body dysmorphia (N=88; Flygare et al., 2020). However, as described above, it is widely acknowledged that much larger sample sizes are required to build reliable prediction models, or to establish superiority of a prediction model over another. The high accuracies (between 75-92%) reported by these studies (without external validation) are likely artificially inflated due to overfitting, considering large-scale studies with thorough model development and validation procedures often show lower but more realistic predictive accuracies (Isacsson et al., 2023).

One important difference between the ML approaches in predicting iCBT response in depression is whether predictions are solely generated based on pre-treatment/baseline data, or data gathered during the course of the treatment (e.g., routine outcome measures or ecological momentary assessments; EMA) (Chekroud et al., 2021). Studies suggest that including change data during treatment can improve model accuracy when predicting psychotherapy outcomes (Bone et al., 2021; Isacsson et al., 2023; Li et al., 2023). Pearson and colleagues (2019) trained several ML algorithms including an elastic net regression, random forest, and an ensemble of the two to predict post-treatment depression score after 8-weeks of iCBT (N=283), using predictors of psychopathology, demographics,

treatment, and environmental context (e.g., census data). They found that the ensemble model generated the highest predictive performance (predictive R^2=0.25), and significantly outperformed the benchmark model of linear regression (R^2=+0.08). Notably, key predictors included pre-treatment variables such as baseline depression severity (total and item scores), comorbid psychopathology, disability, and treatment credibility, but also variables pertaining to specific usage of intervention modules and therapist access which were collected during treatment. However, this study had a relatively low sample size, and the variance explained in the post-iCBT depression score may be confounded by the inclusion of pre-iCBT depression score in the model. They also did not independently validate their findings with an external dataset. Another study conducted by van Breda and colleagues (2018) directly tested the added value of incorporating EMA data in their ML model predictions. They trained 3 ML models (random forest, k-nearest neighbours, and general linear model with likelihood boosting) on 80% of their data (N=182) and independently tested on the rest (N=45). The goal was to predict treatment success 3 months after patients underwent blended-therapy (i.e., combining face-to-face CBT and iCBT) with only self-report information at baseline (demographics, treatment-related, psychotic and health symptoms), and to assess whether adding EMA data of mood ratings (measured once a day) improved model performance. Treatment success was defined as (i) having ≥50% improvement post-iCBT and scoring ≤9 on the Patient Health Questionnaire (PHQ-9) and (ii) scoring ≥5 pre-treatment and ≤4 post-iCBT on the PHQ-9. They were able to demonstrate an average AUC of 0.78 (over 20 iterations) in predicting treatment success in hold-out data, with the most important predictors being depression-related items as well as demographic and treatment-related factors. Interestingly, they found adding EMA data to the model did not result in better predictions. Findings here may need to be interpreted with caution; EMAs of daily mood are not synonymous to standard measurement of depression (Armey et al., 2015), and the apparent increased noise in the EMA data may require bigger datasets to improve prediction performance.

Incorporating symptom change data into ML prediction models likely improves predictions of iCBT treatment response, simply because training data is gathered more proximally to the outcome. However, relying on clinical change data means that models cannot be used to guide and individualise treatment choice prior to one starting treatment. One study showed that 6 weeks' worth of patient-rated symptom scores are needed to

accurately predict iCBT treatment failure above an established benchmark for clinical acceptance (>65% accuracy) (Forsell et al., 2020). This may not be a useful model in practice, considering the amount of time and resource lost only to reveal iCBT could be ineffective for the patient. Being able to identify pre-treatment characteristics that predict treatment response is therefore crucial, allowing for the most cost-efficient and clinically meaningful patient stratification at the time of treatment prescription (Jankowsky et al., 2022; Koutsouleris et al., 2016; Rost et al., 2023).

To this end, two large-scale studies highlighted the potential of a baseline approach. Hornstein and colleagues (2021) led a study where they trained three ML models (random forest, support vector machine, and naïve Bayes) and a reference logistic regression to predict responder status (i.e., ≥ 5-point reduction on the PHQ-9) post-iCBT (N=970) with 14 pre-treatment self-report measures spanning demographics, psychiatric symptoms and chronicity, and treatment-related variables. The winning model, random forest, outperformed the logistic regression significantly during cross-validation, and yielded a moderate out-of-sample AUC of 0.60 and 60% balanced accuracy (N=279). Baseline depression items were the most important predictors, as randomly shuffling them would have decreased accuracy by 4.6%. There remains considerable room for improving the model's prediction, however, attributable to the limited range of only self-report predictors included in the study. Incorporating a wider array of predictors from various sources may be beneficial. To illustrate this, Wallert and colleagues (2022) leveraged baseline multimodal predictors spanning demographics, clinical, process (e.g., time of day completing assessment), and genetic data from N=894 to predict depression remission status post-iCBT treatment. Of all models (i.e., logistic regression, random forest, eXtreme gradient boosting machine, and a meta-learner combining all algorithms) trained on 60% of the sample (N=537), random forest was the only model that significantly outperformed a null-information model, and generated the best prediction performance (i.e., 65.6% accuracy) when tested on the remaining 40% as the independent validation sample (N=357). This final model comprised 45 predictors, and made use of all 4 different variable types which all independently contributed to predicting iCBT remission. While the majority of the retained predictors was clinical self-report data, findings here underscored the predictive utility for integrating multi-modal data for the routine prediction of iCBT remission in depressed patients.

One major caveat of this study, however, is the lack of a treatment comparison group to determine whether the model prediction was specific to iCBT response, or generalisable to other depression treatments as well. To our knowledge, while previous studies have investigated the drug-specificity of ML models predicting antidepressant treatment response (Chekroud et al., 2016; Iniesta et al., 2018), no study has directly tested the treatment specificity of ML-generated treatment predictions in psychotherapy. To address this gap, **chapter 4** of this thesis applied ML to predict depression treatment response for patients initiating iCBT, and tested the model in a hold-out iCBT sample but also in those initiating antidepressant medication. We trained and tested various ML models using a wide array of multi-modal pre-treatment data. The data encompassed comprehensive self-report that evaluated various aspects of life (including physical health comorbidity, diet, exercise), as well as cognitive measures linked to brain functions associated with depressive and highly comorbid symptoms, such as anxiety and compulsivity (Browning et al., 2015; Gillan et al., 2016; Koenen et al., 2009; Rouault et al., 2018; Seow et al., 2021). We coupled this prediction work with an investigation of the explainability of the best model. This is important for improving treatment in a theory-based manner and ensuring ML applications in clinical practice do not produce unintended consequences due to bias. There has been considerable research in this area, with the consensus that decision trees and regression models are generally more interpretable as they provide clear rules and coefficients to explain predictions. In contrast, tree-based or ensemble methods like random forest or neural networks may be less interpretable due to their increased level of processing layers and data transformation (Vieira et al., 2017, 2022). Relatedly, we explored how varying levels of data granularity (e.g., individual symptom scores vs. sum scores) may affect prediction performance across different types of ML algorithms. It may be the case that a symptom-level approach may be more sensitive to picking up predictive signals compared to the use of sum-scores, which may obscure meaningful treatment effects. In doing so, we introduce a major area of research regarding dynamical systems perspectives on depression, the putative importance of symptoms, and their interactions (Borsboom, 2017).

1.6 Critiques of Our Current Conceptualisation of Depression

The current systems of psychiatric disorder classifications, such as the Diagnostic Statistical Manual of Mental Disorders-Fifth Edition (DSM-V) and the International Classification of Diseases (ICD-11) are useful for facilitating a common language

towards treatment options and decisions. They reflect a traditional assumption that mental disorders arise from a single common cause, similar to how physical diseases work (Fried & Nesse, 2015). For example, just as lung cancer can cause a multitude of symptoms in a person (e.g., difficulties breathing, coughing up blood), the different depression symptoms one experiences are seen as the outcome of 'having depression' (i.e., a latent variable that is the common cause of the symptoms). In this view, by addressing the root cause that is the condition itself, this can in turn alleviate the associated symptoms.

However, when we apply psychometric assumptions underlying the common cause model to psychopathology, they do not necessarily hold up. Depression severity and diagnosis are typically determined by calculating a sum of all individual symptom scores. This approach assumes these symptoms are independent of each other and contribute equally to depression (Borsboom, 2008). In reality, depression symptoms can affect each other over time (e.g., having sleep problems the previous night can lead to elevated levels of fatigue and difficulty in concentration the next day), which underscores the importance of considering symptom-level information beyond just their sum total (Fried & Nesse, 2015). Depression symptom patterns are also highly comorbid across disorders, and heterogeneous within the disorder itself. Individuals with different risk factors (Fried et al., 2014), comorbidities (Lux & Kendler, 2010), and levels of impairment (Fried & Nesse, 2014) experience varying combinations of depression symptoms profiles, which ultimately violates the interchangeability of symptoms under the common cause model. In fact, Zimmerman and colleagues (2015) suggest there are 227 different combinations of depression symptoms that can meet the diagnosis criteria of a major depressive episode (MDE). Another study by Fried and colleagues (2016) found 1030 unique DSM symptom profiles in N=3703 depressed patients. Depression symptoms are also not unique to the diagnosis, with many common symptoms shared among other psychiatric conditions such as anxiety disorders, obsessive compulsive disorders, and post-traumatic stress disorder (Forbes et al., 2023; Thaipisuttikul et al., 2014). The severity of these symptoms also seem to exist on a spectrum rather than in discrete categories (Haslam, 2003; Markon et al., 2011). These considerations pose a significant challenge for the binary classification system frequently adopted in psychopathology. All in all, traditional unidimensional, categorical conceptualisation models may be too simplistic to adequately capture the neurobiological and psychosocial complexity of depression, and this may critically influence or impede efforts to understand, predict and treat it.

1.7 To Understand iCBT Response – Network Theory of Psychopathology

As an alternative view, network theory posits that mental disorders, such as depression, arise from causal interactions between symptoms that actively influence each other over time. By centring the focus on symptoms as agents that constitute (and not passively indicate) depression, the network approach speaks to what many consider the dynamic, co-evolving nature of depressive symptomatology in the real world (Borsboom, 2017; Bringmann et al., 2014; Cramer et al., 2010). According to this theory, a depressive episode occurs when a sufficient number of related symptoms are triggered by an external event (e.g., multiple job rejections) and persist for a period of time. Therefore, it is not just the severity of these individual symptoms that matter, but how they may mutually reinforce each other over time, creating a positive feedback loop of co-deteriorating symptoms that can worsen one's depressive state. Recovery from depression can happen when these symptoms subside, or when the causal relations between them break, often due to treatment.

By applying network analysis to study depression, one major advantage is that it can easily visually represent and mathematically describe the complex interplay between depression symptoms. We can use these observations to generate unique, testable predictions and hypotheses about the underlying theory (e.g., cyclic feedback loop in panic disorder; Cramer et al., 2010). In a group-level network (i.e., cross-sectional network), symptoms are called 'nodes', and the relationships between them are called 'edges' (McNally, 2020). Nodes typically represent various psychological/behavioural constructs. For example, this can be an individual item, a subscale, or the sum score of a psychometric scale. Some nodes are more important than others, and this is indicated by its centrality. Common centrality estimates include strength (i.e., sum of the absolute strength of all edges in and out of a node), closeness (i.e., sum of the shortest paths from one node to other nodes), and betweenness (i.e., number of shortest paths passing through a node). The higher the centrality a node has, the more influence it exerts onto other nodes in the same network. Edges can depict the associations between symptoms in many ways, with the most common being partial correlations (i.e., the unique relationship between two specific nodes after controlling for the rest of the nodes within the same network). Depending on the source of data, they can differ in sign (positive/negative edge), strength (thickness), and if based on time-series data, direction of influence (based on Granger causality) (Borsboom, 2017).

1.7.1 Network Characteristics and Treatment Response

Network theory has gained prominence in psychiatry research in recent years (Robinaugh et al., 2020), but there remains an ongoing debate about how best to use and interpret symptom networks (Borsboom, 2017; Forbes et al., 2017; Wichers et al., 2017). Under this framework, there are several key predictions it makes with regards to individual differences in network characteristics that may contribute to heterogeneous treatment outcomes. One of these key predictions is that symptoms propagate and activate each other more easily in a network that is more densely connected. Therefore, the tighter the network, the less psychologically resilient it is, as it may react more vigorously to external perturbations and take longer to recover from them. This means that individuals with a tightly connected network of symptoms are predicted to have greater overall depression vulnerability, experience more severe depression, and have poorer prospects for recovery during treatment (Cramer et al., 2016; Pe et al., 2015; van Borkulo et al., 2015). Preliminary support for this came from a simulation study by Cramer and colleagues (2016), where they found that agents with densely connected networks tend to remain in a depressive state, even in the absence of any triggering events. If this holds true empirically, individual differences in network connectivity could provide valuable prognostic information. This information can then be used as a tool to facilitate treatment personalisation and generate insights into understanding differential treatment responses.

A host of prior studies have attempted to test this idea, using cross-sectional networks, which analyse one observation from each patient at a single timepoint, and often use partial correlations to assess the relationships between nodes (Epskamp et al., 2018a; Epskamp & Fried, 2018).These studies have largely taken the approach of dividing samples into groups that have different clinical outcomes and comparing their baseline network properties. For example, assessing whether baseline symptom network connectivity is higher in those with worse post-treatment outcomes in depression. van Borkulo and colleagues (2015) were the first to statistically test this, by comparing the baseline connectivity differences between patients who achieved remission (N=262) and those with persistent depression (N=253), defined at a two-year follow-up. They found initial supporting evidence in line with the network theory, in that patients with persistent depression showed tighter baseline connectivity compared to those who remitted. This finding, however, did not hold up in further sensitivity analyses involving different analytical decisions in network estimation procedures (i.e., tuning of the hyperparameter

in regularisation). Since, other research has replicated this result (N=566/174; McElroy et al., 2019), but null findings have also emerged. In an adolescent sample, baseline connectivity was found to be more elevated in relatively poor responders (N=232) vs. good responders (N=233) to depression treatment, but this was only trending towards significance (Schweren et al., 2018). Another study collapsed individual patient data from 6 RCTs and examined depression and anxiety symptoms together. Results also did not reveal connectivity differences at baseline between future remitters of depression (N=956) and those who went on to have persistent symptoms (N=1466) (O'Driscoll et al., 2021).

Further adding to the ambiguity are studies that examine changes in network connectivity during the course of treatment as an indicative marker of clinical change. According to network theory, if a treatment is successful, we should anticipate a decrease in network connectivity from baseline to post-treatment, due to symptom alleviation and the network becoming 'looser' as a result. Contrary to this, the majority of studies instead observed an increase in network connectivity after treatment with no clear explanations (Beard et al., 2016; Berlim et al., 2020; Blanco et al., 2020; Bos et al., 2018; Curtiss et al., 2021). To tease this out further, McElroy and colleagues (2019) classified depressed patient groups into those who improved (n = 556), those who remained unchanged (n = 2277), and those who deteriorated (n = 174) after treatment. While those who improved had the sparsest network at baseline compared to others, all groups showed a significant increase in network connectivity post-treatment, with the biggest increase shown in the improved group. These findings altogether challenge the prevailing narrative that greater connectivity is a function of greater psychological vulnerability.

1.7.2 Gaps in the Network Literature on Treatment Response

Overall, the current state of the research is too ambiguous to make any definitive conclusions about the prognostic value of cross-sectional networks for depression treatment response. In light of these inconsistencies, the following section highlights several considerations that need to be addressed.

The first limitation refers to fundamental issues regarding what network connectivity actually means when applied to the study of psychopathology. According to network theory, a tightly connected network of symptoms is closely linked to the severity of the psychological presentation; once symptom(s) are activated, they positively reinforce each

other and together sustain a depressive state via a downward spiralling of simultaneous deterioration. A study by Pe and colleagues (2015) illustrates this, where compared to healthy controls, they found that participants with depression had greater network connectivity of negative emotions but not positive emotions. On the other hand, van Borkulo and colleagues (2015) found that baseline connectivity differences between treatment responders and non-responders remained even after controlling for the higher average depression severity observed in non-responders. They argued that network estimates rely on covariance of symptoms rather than their mean levels (whether it is higher or lower severity). While the mean levels of symptoms themselves do not confound network connectivity, they may still indirectly affect network connectivity estimations if they are linked to symptom variance. Variance directly relates to the strength of the association between two symptoms and how much they (can) co-vary with each other, where increased variance may lead to increased connection strength (Terluin et al., 2016). It has been suggested that severity may confound network connectivity estimates, due to its association with variance in skewed data (Fried et al., 2016b; Terluin et al., 2016). As a result, if sub-grouping patients based on severity level indirectly leads to differential range restriction of items (e.g., imbalance in item variance between-groups due to severity differences), this may artificially inflate or diminish observed connectivity between symptoms within a network (Linn, 1968; Terluin et al., 2016). While prior research has attempted to control for this by matching comparison groups based on baseline severity (McElroy et al., 2019; van Borkulo et al., 2015), they failed to adequately consider the role of variance as a separate confounding construct. Notably, as far as we know, no study has tested whether psychometric properties such as severity (mean) and/or variance distinctly drive the potential connectivity differences between responders and non-responders.

These potential confounds related to mean and variance expose major issues with how the field has thus far attempted to test network theory. Due to issues of data availability, the vast majority of studies use cross-sectional data, which means inferences are made on a group-level (Fisher et al., 2018). This is a significant gap, because network theory is conceptually based on the dynamic properties of symptoms, their evolution and interaction over time – not static correlations across individuals. With studies often tied to cross-sectional data, statistical comparisons are often made using two group-level networks, constructed from two halves of a sample that differ in some respect (e.g.,

severity, prognosis). A common test to use in this regard is the Network Comparison Test, which is a two-tailed resampling-based permutation test that evaluates differences between two cross-sectional networks (van Borkulo et al., 2022). Insights generated from group-level comparisons, however, cannot be applied to individual patients. These group-level comparisons also make it impossible to adequately tease out potentially confounding effects of factors like symptom severity and variance.

Another major caveat not unique in depression but across the psychopathology network literature is the inclusion of small sample sizes that are underpowered to detect true effects (Forbes et al., 2017). Network estimation procedures rely on sampling variations, and these variations are strongly influenced by sample sizes. This raises serious concerns about the stability of networks, and in turn, replicability of findings (Borsboom et al., 2017; Bringmann et al., 2022), which is a hot topic that concerns the generalisability of network research in psychopathology. A recent systematic review conducted by Schumacher and colleagues (2022) found 56 studies explored the use of symptom network characteristics in mental health treatment, and revealed a median study sample size of N=151. Often, variations in methodological and analytical procedures can profoundly affect the significance of findings and result in different conclusions. Specifically, de Vos and colleagues (2017) has shown that mathematical procedures used in network estimation can significantly influence whether depression network connectivity appears to be greater in depressed vs. healthy individuals. A large sample size is therefore necessary to ensure reliable networks capable of withstanding slight deviations in procedural choices. It is worth noting, however, the current field has no clear rules or guidelines for determining the appropriate sample size needed to robustly estimate cross-sectional networks with a pre-determined set of parameters (number of edges and nodes). While methods for estimating the stability and robustness of estimated network parameters exist (Epskamp, Borsboom, et al., 2018), only around 50% of the related network studies use them (Schumacher et al., 2022). Having said that, large-scale studies in this area of network research, albeit rare, do exist (Beard et al., 2016; Esfahlani et al., 2017; Lorimer et al., 2020; O'Driscoll et al., 2021). However, these studies often include unequal sample sizes when comparing between-groups, thus warranting caution in the interpretation of findings.

Lastly, there is an inherent lack of understanding of just how big these effects really are, which is crucial to properly evaluate the prognostic utility of network characteristics for

depression treatment response. In addition to looking at the connectivity of the network as a whole (i.e., global network strength), research in this area has delved into the local connectivity of each node in the network. Another related prediction of network theory is that central symptom nodes, due to their increased ability to influence other symptom nodes, are thought to be preferential treatment targets (Borsboom & Cramer, 2013; Boschloo et al., 2015; Fried et al., 2016). In cross-sectional network research, studies consistently highlight 'sadness' as the most central symptom in depression (Contreras et al., 2019). Taking that cross-sectional insight to the individual level, network theory posits that a reduction in the severity of sadness should propagate accordingly to the other symptoms in the network. Identifying these influential symptoms could potentially enhance and personalise depression treatments; if we can prioritise targeting central symptoms over more peripheral ones, we can more effectively and efficiently drive clinical improvement. In some cases, there are mixed evidence (beyond depression studies) that central symptoms predict treatment response better than non-central symptoms. For instance, in cross-sectional research, studies have linked baseline elevations in central symptoms with non-response (Elliott et al., 2020; Hagan et al., 2021; Esfahlani et al., 2018). Some also showed that changes in the severity of central symptoms corresponded to overall changes in other symptoms during the course of treatment (Papini et al., 2019; Robinaugh et al., 2016; Rodebaugh et al., 2018). But findings concerning the centrality hypothesis are not entirely consistent. Spiller and colleagues (2020) tested whether baseline centrality indices such as strength, predictability, and expected influence in cross-sectional networks were predictive of therapeutic change in N=710 treatment-seeking patients with PTSD. Of all indices, only expected influence (i.e., sum of signed associations between one node and the rest of the network) predicted how strongly changes in symptoms associated with changes in the rest of the symptoms. This effect disappeared when the outlier symptom 'amnesia' in PTSD was removed from analyses. Both mean symptom severity and infrequency of symptom endorsement, two non-network metrics, turned out to be better predictors of response than expected influence.

In sum, the evidence linking greater network connectivity to treatment non-response is incomplete. This is primarily due to the use of small study samples that render network estimations unreliable, while further limiting investigations into the potentially confounding effects of symptom severity and variance in connectivity differences

between responders and non-responders. Importantly, without considering network connectivity with other metrics, we also do not know for certain how useful network connectivity is in predicting treatment response, in comparison to other readily available information about responders and non-responders. In order to build a convincing case for the prognostic value of network connectivity (and node centrality) for predicting treatment response, we need to significantly scale-up network research such that robust, contextualised evidence can be generated. **Chapter 5** therefore leverages the scalability of iCBT to fill in these research gaps; by analysing data from N>40,000 individuals, our study sought to comprehensively test for baseline connectivity differences between iCBT responders and non-responders, address the above limitations, and shed light on understanding why some patients benefit more from iCBT than others.

1.8 Theory Aims and Objectives

Amid a global mental health crisis underpinned by surging demand for scarce resources, we have a treatment that is widely accessible, scalable, and effective to help patients suffering from depression – iCBT. Yet, despite its growing adoption in clinical practice, we do not know who it works best for and why. This heterogeneity in treatment response may be due to the intricate dynamics of depression, which we are currently ill-equipped to unravel with current methodologies and small samples. In this thesis, I sought to remedy this via a *big data* approach to delve deeper into the predictors and mechanisms of iCBT for depression. Specifically, I assessed the real-world effectiveness of iCBT for vulnerable subpopulations prone to depression to inform who we should *prescribe* the intervention to, I leveraged algorithmic, data-driven tools to *predict*, at baseline, treatment response to early stages of iCBT, and lastly, I applied cross-sectional network analysis to depression symptoms to *understand* why iCBT works for some, not others.

Chapter 2 – The effectiveness of low-intensity psychological interventions for comorbid depression and anxiety in patients with long-term conditions: A real-world naturalistic observational study in IAPT integrated care

Chapter 2 evaluated the effectiveness of low-intensity psychological interventions, such as internet-delivered cognitive behavioural therapy (iCBT), guided self-help (GSH), and psychoeducational group therapy (PGT) for comorbid depression and anxiety in individuals with long-term conditions (LTC). We conducted a retrospective analysis of patients enrolled in Talking Therapies, which is an IAPT service in the NHS, UK, from

2016 to 2020. The study included N=21,051 patients, of which N=4,024 reported having at least one long-term condition. We first defined and reported each IAPT outcome variable (i.e., recovery, reliable improvement, and reliable recovery) for the LTC and non-LTC cohort in each intervention group. We then investigated whether low-intensity psychological interventions (as a whole) were overall more effective in improving clinical outcomes in patients with or without LTC. Lastly, we tested which, if any, of the low-intensity interventions were linked to comparatively greater changes in clinical outcomes, and whether there is a differential pattern of intervention effectiveness for the LTC and non-LTC patients. In all analyses, we controlled for baseline severity, age, and gender as potential covariates and modelled all potential interactions.

Chapter 3 – The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation

Chapter 3 tested the feasibility of an internet-based methodology for accelerating the acquisition of large datasets needed for treatment prediction and personalisation research in psychiatry. We remotely gathered longitudinal, observational self-report and cognitive data via a web-browser from patient cohorts receiving iCBT (N=600) or antidepressant medication treatment (N=110) for 4 weeks. Here, we described our fully internet-based study design protocol, patient characteristics of our online sample, their treatment adherence, study adherence and compliance. We evaluated the quality of our online data, and examined the tolerability of our protocol using qualitative feedback from patients on study design and implementation. Finally, we detailed the advantages and disadvantages of this method and offered suggestions and guidelines for future studies.

Chapter 4 – Machine learning prediction of depression symptom change following internet-delivered cognitive behavioural therapy for depression

Chapter 4 leveraged the large-scale online samples gathered in chapter 3, and sought to use machine learning techniques to predict early depression treatment response for patients initiating iCBT using a wide range of self-report and cognitive data gathered before treatment. We tested a combination of linear and non-linear machine learning models (elastic net regression, random forest, eXtreme gradient boosting machine learning in linear and tree form) and evaluated the predictive utility of feature sets with varying levels of granularity. Each type of algorithm was also compared to a benchmark of their own, comprising only baseline depression severity, sex, and age. In total, we

trained 20 models using repeated nested cross-validation and evaluated their performance in predicting depression symptomatology change. The best model was then tested for significance against its benchmark via permutation-based significance testing. We repeated the cross-validation analyses 100 times for the best model, and brought forward the averaged best model for external validation on 1) a hold-out iCBT sample to assess generalisability and 2) an antidepressant medication sample to assess treatment-specificity of the model's predictions. Finally, we examined the top ranked predictors of the best model using Shapley additive explanation values (SHAP).

Chapter 5 – Estimating the prognostic value of cross-sectional network connectivity for treatment response in depression

Chapter 5 tested for network connectivity differences between responders and non-responders in a sample of N>40,000 patients seeking iCBT for depression at baseline, and separately for both cohorts, pre-post treatment. Leveraging this big sample, we carried out power analysis on subsamples of our data to contextualise prior research. We then further split our sample into smaller, independent subsets, and estimated cross-sectional networks for each of these subsets. Using these subsamples, we conducted novel parametric analyses to tease out the effects of baseline severity and variance in connectivity differences observed between responder and non-responder groups. Finally, we benchmarked the magnitude of these effects by comparing the predictive utility of network metrics such as connectivity and centrality measures to other baseline differences such as sum/item score means and variances for prognosis information.

Chapter 2 – The Effectiveness of Low-Intensity Psychological Interventions for Comorbid Depression And Anxiety in Patients with Long-Term Conditions: A Real-World Naturalistic Observational Study in IAPT Integrated Care

2.1 Introduction

Common psychological disorders, such as depression and anxiety, are two to three times more likely to occur in people with physical long-term conditions (LTCs), for instance diabetes, cardiovascular disease, arthritis, and asthma, when compared with the general population (Guthrie et al., 2016) . The co-occurrence of these psychological disorders can significantly impact the prognosis for the LTC, as well as the individual's broader quality of life across different domains (McManus et al., 2016; Naylor et al., 2012). For example, depression and anxiety were significantly associated with poorer health-related (e.g., physical functioning, difficulties with medical care) and work-related outcomes (e.g., more sick leave and work interference) in patients with diabetes (Das-Munshi et al., 2007). LTC patients with comorbid depression have also been shown to exhibit less effective self-care in the form of poorer treatment adherence for their physical condition (i.e., three times the odds of treatment noncompliance compared to non-depressed patients) (DiMatteo et al., 2000). This is not surprising, as the interplay between physical and mental illness has been suggested to raise healthcare costs by at least 45% for each LTC patient experiencing comorbid mental health problems (Naylor et al., 2012), a considerable increase for a population found to disproportionately live in deprived areas with reduced access to all kinds of resources (Hoang et al., 2013). If left untreated, depression and anxiety can further exacerbate the complications surrounding the LTCs, such that several studies have posited a link between these psychological comorbidities and increased use of urgent care (Dickens et al., 2012; Guthrie et al., 2016) and risk of mortality (Dossa et al., 2011; Watkins et al., 2013).

Considering the magnitude of complications that arise when depression and anxiety accompany a LTC, it is imperative to effectively address these psychological comorbidities. One way to manage depression and anxiety symptoms in people with LTCs is through cognitive behaviour therapy (CBT) (NICE, 2009). The fundamental premise of CBT is that thoughts, physical symptoms, mood, and behaviours are all interrelated, and the way people make sense of their environment affects their feelings

and behaviours. CBT has been delivered through evidence-based protocols to people with LTCs; the goal is to alleviate their depression and anxiety symptoms so as to reduce the burden these may have on one's self-management (Wroe et al., 2018). However, many significant barriers remain for people with LTCs that limit their capacity to benefit from CBT, including difficulty in accessing treatments due to physical limitations imposed by their LTC, increased healthcare costs in managing multiple illnesses, as well as disjointed, insufficient service delivery in meeting both their physical and mental health needs (Naylor et al., 2012).

In the UK, the Improving Access to Psychological Therapies (IAPT) are working towards an integrated stepped-care model to extend their access to people with LTCs for treating comorbid depression and anxiety symptoms (National Collaborating Centre for Mental Health, 2018) In line with current NICE recommendations NICE, 2009; 2011), IAPT delivers CBT-based low-intensity psychological interventions to patients with LTCs. These include guided self-help programs (GSH) which combines written self-help materials with telephone support, psychoeducational group therapy (PGT) which delivers psychoeducation in a group-based format, and internet-based cognitive behavioural therapy (iCBT) which is a clinician-guided intervention delivered online. While these interventions typically follow a standardised treatment protocol and may not tailor to the specific needs of individuals presenting with a particular LTC, they require less intensive resources for implementation (i.e., less clinician time and involvement, lowered costs), and can therefore help improve treatment access which is a proven obstacle for LTC patients (National Collaborating Centre for Mental Health, 2018).

The effectiveness of these low-intensity psychological interventions has been widely established in the general patient population. A large-scale naturalistic cohort study of IAPT patients found these interventions to be effective in improving depression, anxiety and impaired functioning, revealing iCBT to have a greater average treatment effect when compared to GSH and PGT (Palacios et al., 2023) These interventions also enjoy comparable effectiveness to traditional face-to-face CBT, with several systematic reviews and meta-analyses suggesting no association between varying delivery formats of CBT and outcomes (Cuijpers et al., 2019, 2023; Weitz et al., 2018). In this regard, however, evidence for their effectiveness in people with LTC is lacking. Firstly, most clinical trials investigating psychological interventions disregard patients with an accompanying physical illness and its impact. Of the existing studies, systematic reviews found

preliminary, favourable evidence for the use of low-intensity psychological interventions for patients in LTC, but they also highlighted crucial limitations such that studies are generally underpowered (Charova et al., 2015; Ould Brahim et al., 2021) and have poor to fair methodological quality (Jackson et al., 2019; Mehta et al., 2018; Ould Brahim et al., 2021). The present study sought to fill this gap, by comparing the effectiveness of low-intensity psychological interventions provided through IAPT for over 4,000 LTC patients and over 17,000 non-LTC patients, in a real-world setting. The study additionally examined whether the LTC cohort showed differential response to iCBT, guided self-help, and psychoeducational group therapy when compared to non-LTC patients.

2.2 Methods

2.2.1 Study Design and Setting

The study adopted a retrospective, observational design examining four years of routine data of patients within Talking Therapies, an IAPT service in Berkshire Healthcare NHS Foundation Trust. Talking Therapies employs a stepped-care model of psychological care for patients by matching treatment intensity to their needs. Each patient has an initial appointment with a clinician, who assess the severity of their symptoms so to determine an appropriate care pathway in conjunction with the patient. If deemed suitable by the clinician in agreement with the patient, iCBT, GSH, or PGT are offered as low-intensity interventions as treatment options.

2.2.2 Participants

The study examined patients experiencing depression and/or anxiety symptoms who were enrolled in iCBT, PGT, or GSH at step 2 or 3 within Talking Therapies from April 1st 2016 to March 31st 2020. In line with the IAPT reporting criteria for establishing universally measurable outcomes (National Collaborating Centre for Mental Health, 2018), a course of treatment is defined as attendance at two or more treatment sessions (in-person or via telephone contact), or receiving two or more online reviews (i.e., patients were not required to have completed the entire course of treatment). Therefore, the study excluded data from patients who did not complete a course of treatment, as well as those below 18 years old at the initial assessment appointment. According to the IAPT-LTC service implementation guide, an LTC is defined as a range of long-term physical health conditions including, but not limited to, cardiovascular disease, chronic obstructive pulmonary disease, diabetes, chronic pain, and musculoskeletal disorders (National

Collaborating Centre for Mental Health, 2018). Patients indicated 'Yes' or 'No' for their LTC status, and those who did not provide this information were further excluded from analyses in the study. As a result, 21,051 patients were included in the study, of which 4024 had at least one LTC, defined as (19.1%). It is noteworthy to highlight that information regarding the specific type of LTC(s) each patient had in this study was not available to the authors for analyses.

2.2.3 IAPT Low-Intensity Interventions

Internet-Delivered Cognitive Behavioural Therapy (ICBT). The iCBT programs consist of seven online modules following evidence-based CBT principles for the treatment of depression and/or anxiety (Richards et al., 2020). The intervention comprises common cognitive, emotional and behavioural components (e.g., behavioural activation, mood and lifestyle monitoring) and additionally tailors content to the patient's mental health diagnosis and presentations (e.g., *Challenging Core Beliefs* for depression; *Worry Tree* for anxiety). Programme content includes several forms of rich media content (videos, quizzes, animations, audio) to facilitate the delivery of the intervention. Each patient is assigned a single, trained clinician who guides treatment progress and responds to queries through weekly/bi-weekly reviews. The recommended duration of guided iCBT is 6-8 weeks, after which the patient can still access the program for up to 12 months.

Guided Self-Help (GSH). GSH begins with a face-to-face consultation between the patient and their clinician for treatment planning. The treatment plan incorporates CBT-based strategies using written self-help materials, which include information on the patient's specific condition and CBT-based techniques for self-management (e.g., behavioural activation, cognitive restructuring), along with related homework exercises (Baguley et al., 2010). Clinician support is provided through 4-6 telephone calls typically scheduled every two weeks, each lasting 20-25 minutes.

Psychoeducational Group Therapy (PGT). Typically facilitated by two clinicians, PGT is a group-based course delivering CBT psychoeducation for managing depression and anxiety symptoms. PGT seeks to normalise patients' difficulties within a group setting. Patients are encouraged to share their experiences and discuss relevance of taught materials amongst peers to increase awareness of individual issues in a collective manner. Patients are also tasked with small homework exercises, taking 15-20 minutes daily. The

recommended duration of PGT consists of four weekly sessions, each lasting approximately 90 minutes with up to 15 patients in attendance at once.

2.2.4 Outcome Measures

This study included depression and anxiety as the primary outcomes and functional impairment as the secondary outcome.

Patient Health Questionnaire-9 (PHQ9). This is a nine-item self-report of depression symptoms with a total score ranging from 0-27, where higher scores indicate more severe level of depression. The PHQ-9 distinguishes well between depressed and non-depressed individuals using the clinical cut-off total score ≥ 10 with good reliability and validity (Kroenke et al., 2001).

Generalized Anxiety Disorder (GAD-7). This is a seven-item self-report of anxiety symptoms with a total score ranging from 0-27, where higher scores indicate more severe level of anxiety. The GAD-7 uses a cut-off point of ≥ 8 and has good convergent validity with anxiety scales and good reliability (Spitzer et al., 2006).

Work and Social Adjustment (WSAS). This is a five-item self-report of functional impairment, examining the experiential impact of a disorder across different life domains from the perspective of the patients. The WSAS has a total score ranging from 0-40, where higher scores indicate poorer adjustment. The measure has also demonstrated good reliability and sensitivity (Zahra et al., 2014).

2.2.5 Data Analysis

The study analysed four years of IAPT patient data from April 1st 2016 to March 31st 2020. Baseline demographics and symptom severity were compared between LTC and non-LTC cohorts across three low-intensity interventions. Patients' depression, anxiety, and functional impairment scores in their last treatment session before discharge were treated as their 'post-treatment' score. Outcome variables of Recovery, Reliable Improvement, and Reliable Recovery defined according to the IAPT reporting criteria (National Collaborating Centre for Mental Health, 2018) were compared between LTC and non-LTC cohorts. IAPT Reliable Change Indices (RCI) of six and four points were used as cutoffs to measure reliable change on the PHQ-9 and GAD-7, respectively. Recovery is determined if patients transitioned from being at caseness pre-treatment to non-caseness post-treatment, where caseness is defined as scores ≥10 on the PHQ-9 and ≥ 8 on the GAD-7. Reliable Improvement is determined when there is a decrease in the

PHQ-9 or the GAD-7 that is greater than the RCI and with no increase on either measure larger than the RCI. Finally, Reliable Recovery is determined when patients achieved both Recovery and Reliable Improvement.

To investigate whether the interventions were overall more effective in improving clinical outcomes in patients with or without LTC, Analyses of Covariance (ANCOVAs) were used, with each clinical outcome serving as dependent variables, time (pre-post treatment) and LTC status (LTC or non-LTC) as independent variables, and baseline severity of clinical scores as a covariate. To determine which, if any, of the low-intensity interventions were associated with comparatively greater changes in clinical outcomes, and whether there is a differential pattern of intervention effectiveness for LTC and non-LTC patients, the above analyses were repeated with the addition of intervention (iCBT, PGT, GSH) as an independent variable in the models. Additionally, control analyses on intervention-specific effects were conducted with the LTC cohort specifically. Considering imbalances in age and gender ratios within each cohort across the interventions (see **Table 2.1**), the models controlled for potential effects of age and gender by including them as covariates. All potential interactions between independent variables were modelled, and Tukey adjusted pairwise comparisons were conducted to decompose significant interaction effects. Data analyses were carried out using R statistical package Version 4.1.1.

Table 2.1. Sample characteristics of LTC and non-LTC cohorts across intervention groups (iCBT, PGT, GSH).

LTC Status Group Comparisons				
Characteristics	**LTC Status**		**$\chi 2$ / t (df)**	**p**
	LTC (n=4024)	Non-LTC (n=17,027)		
Gender (N, %)			2.10 (1)	0.147
Female	2620 (65.10)	11,291 (66.31)		
Male	1404 (34.90)	5736 (33.69)		
Age				
Mean, SD (Range)	46.05, 15.88 (18-80)	35.79, 12.96 (18-80)	-43.14 (21,049)	< 0.001***
Baseline PHQ-9				
Mean, SD	14.46 (6.03)	13.71 (5.85)	-7.31 (21,049)	< 0.001***
Baseline GAD-7				
Mean, SD	12.82 (5.17)	12.98 (4.86)	1.89 (21,049)	0.059
Baseline WSAS				
Mean, SD	17.83 (9.70)	16.80 (8.88)	-6.52 (21,049)	< 0.001***
Treatment Duration (Days)[a]				
Mean, SD	90.47 (58.25)	89.68 (53.63)	-0.75 (17,431)	0.452

Number of Appointments				
Mean, SD	5.14 (2.64)	5.10 (2.41)	-0.91 (21,049)	0.361
IAPT Clinical Outcomes				
Caseness (N, %)	3636 (90.36)	15,516 (91.13)	2.34 (1)	0.126
Recovery (N, %)	1984 (54.57)	8618 (55.54)	1.10 (1)	0.295
Reliable Improvement (N, %)	2364 (65.02)	10,334 (66.60)	3.24 (1)	0.072
Reliable Recovery (N, %)	1803 (49.59)	7848 (50.58)	1.12 (1)	0.290

[a]Instances where treatment duration is 0 due to administrative errors were removed from analyses

2.3 Results

Of the 21,051 patients that were included in the study, 12,746 received GSH, 6,857 received iCBT, and 1,448 received PGT. Among the 4,024 of the patients that had at least one LTC, 2,620 patients underwent GSH, 1,065 patients underwent iCBT and 339 patients underwent PGT. **Table 2.1** outlines the demographic and clinical characteristics of both samples. The gender ratio for both LTC patients (65.1% females) and non-LTC patients (66.3% females) were comparable, $\chi 2_{GENDER}$ (1) = 2.10, p = 0.15. However, the LTC cohort (M = 46.05, SD = 15.88) were on average older than those without a LTC (M = 35.79, SD = 12.96), t(21049) = -43.14, $p < 0.001$. In relation to pre-treatment symptom severity, the LTC cohort (M = 14.46, SD = 6.03) had significantly higher levels of depression than their non-LTC counterparts (M = 13.71, SD = 5.85), t(21049) = -7.31, $p < 0.001$. Similarly, the LTC patients also had significantly higher pre-treatment functional impairment (M = 17.83, SD = 9.70) than non-LTC patients (M = 16.80, SD = 8.88), t(21049) = -6.52, $p < 0.001$. Anxiety scores were not significantly higher in the non-LTC cohort (M = 12.98, SD = 4.86) than in the LTC cohort (M = 12.82, SD = 5.17), though there was a trend towards significance, t(21049) = 1.89, p = 0.059. The two cohorts did not differ in the mean number of treatment days (p = 0.452) nor in the number of clinician-guided appointments or reviews (p = 0.361) (see **Supplementary Materials 8.1.1 and 8.1.2** for additional demographics and treatment characteristics).

2.3.1 Clinically Significant Changes

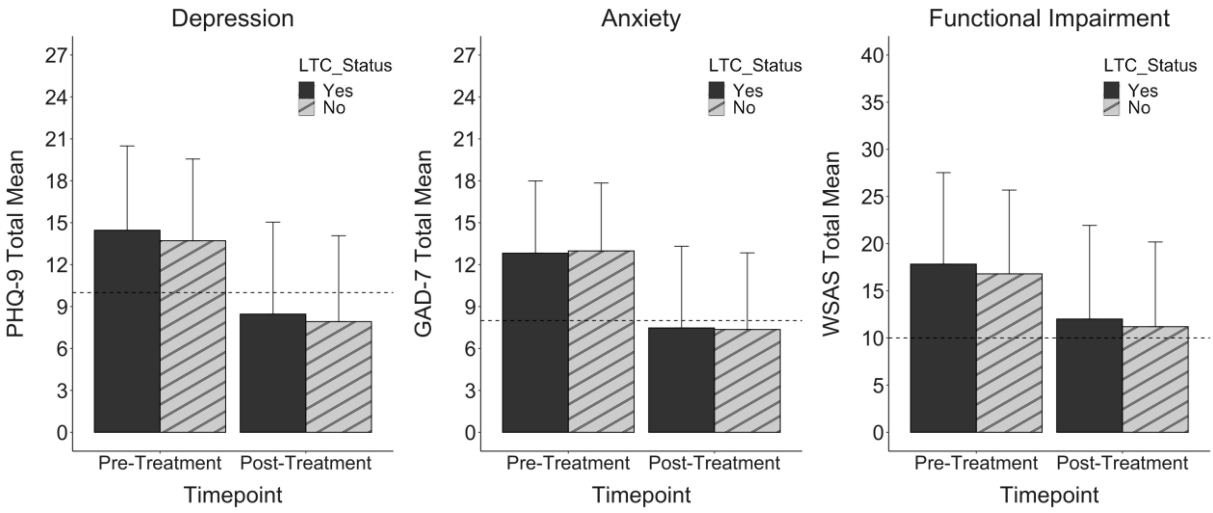
For both the LTC and non-LTC cohorts, around 90% of patients met the criteria for caseness at pre-treatment (LTC n = 3,636, 90.36%, non-LTC n = 15,516, 91.13%, $\chi 2_{CASENESS}$ (1) = 2.34, p = 0.126). Over half of both cohorts achieved Recovery (LTC n = 1,984, 54.57%, non-LTC n = 8,618, 55.54%, $\chi 2_{RECOVERY}$ (1) = 1.10, p = 0.295), and around two-thirds of both cohorts achieved Reliable Improvement (LTC n = 2,364, 65.02%, non-LTC n = 10,334, 66.60%, $\chi 2_{RELIABLE\ IMPROVEMENT}$ (1) = 3.24, p = 0.072). The proportion of patients who achieved Reliable Recovery in the LTC cohort (N = 1,803,

49.59%) and the non-LTC cohort (N = 7848, 50.58%), which is approximately half for both samples, were also comparable, $\chi 2_{\text{RELIABLE RECOVERY}}$ (1) = 1.12, p = 0.290 (see **Table 2.1**).

2.3.2 Overall Intervention Effectiveness for LTC vs. Non-LTC Patients

A comparison of intervention effects between the LTC and non-LTC cohort was conducted using two-way ANCOVAs with baseline severity as a covariate (see **Figure 2.1**). For depression, there was a general reduction in PHQ-9 scores from pre- to post-treatment, F(1, 42097) = 13551.39, $p < 0.001$. A significant LTC status by time interaction was also evident, F(1, 42097) = 4.31, p = 0.038, driven by the LTC cohort experiencing a greater reduction in symptoms (adj. pre M = 14.00, SE = 0.06; adj. post M = 8.00, SE = 0.06, d = 1.47) than the non-LTC cohort (adj. pre M = 13.82, SE = 0.03; adj. post M = 8.00, SE = 0.03, d = 1.42). For anxiety, GAD-7 scores overall reduced from pre- to post-treatment, F(1, 42097) = 13645.64, $p < 0.001$. While there was also a significant time by LTC status interaction, F(1, 42097) = 8.70, p = 0.003, in contrast to depression, reductions in anxiety were also greater for the non-LTC cohort (adj. pre M = 12.96, SE = 0.03; adj. post M = 7.34, SE = 0.03, d = 1.48) compared to the LTC cohort (adj. pre M = 12.91, SE = 0.06; adj. post M = 7.57, SE = 0.06, d = 1.41). In terms of functional impairment, WSAS scores also reduced from pre- to post-treatment, F(1, 42097) = 5602.91, $p < 0.001$, but having an LTC did not influence this, F(1, 42097) = 2.18, p = 0.140. **Table 2.2** provides an overview of the adjusted means for all three outcome measures at pre- and post-treatment for LTC and non-LTC patients.

Figure 2.1. Observed means of LTC and non-LTC service users at pre-treatment and post-treatment for each outcome measure.



Note. The dashed line indicates the threshold for caseness for each measure and the error bars indicate standard deviation of the mean. LTC patients showed slightly greater reductions in depression symptoms relative to non-LTC patients, while non-LTC patients showed slightly greater reductions in anxiety symptoms when compared to their LTC counterpart. There were no group-differences in the extent to which functional impairment improved post-treatment.

Table 2.2. Adjusted means of each outcome measure for LTC and non-LTC service users at pre-treatment (pre) and post-treatment (post).

Measures	**LTC**	**Non-LTC**	**F (df)**	**p**
	Pre v. Post M (SE)	Pre v. Post M (SE)		
PHQ-9	14.00 (0.06) v. 8.00 (0.06)	13.82 (0.03) v. 8.03 (0.03)	4.31 (1, 42097)	0.038*
GAD-7	12.91 (0.06) v. 7.57 (0.06)	12.96 (0.03) v. 7.33 (0.03)	8.70 (1, 42097)	< 0.001***
WSAS	17.21 (0.10) v. 11.39 (0.10)	16.94 (0.05) v. 11.35 (0.05)	2.18 (1, 42097)	0.140

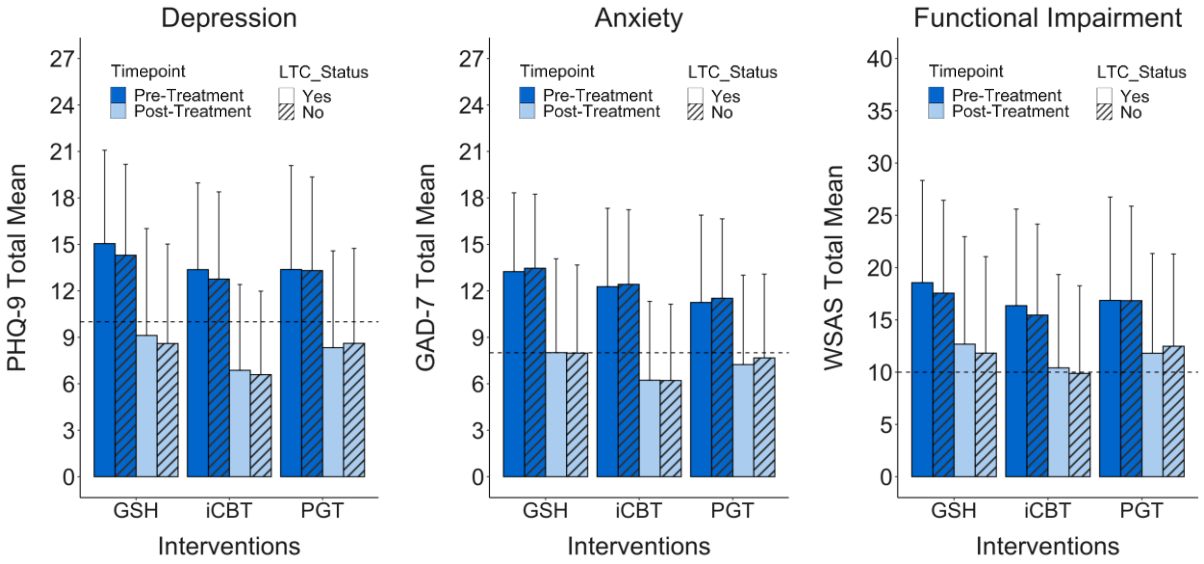
Note. GAD-7, generalised anxiety disorder-7 item questionnaire; LTC, long-term condition; PHQ-9, patient health questionnaire-9 item. WSAS, work and social adjustment scale; SE, standard error.

2.3.3 Intervention-Specific Effectiveness for LTC vs. Non-LTC Patients

Three-way ANCOVAs were employed to investigate differences in the effectiveness of specific low-intensity psychological interventions (iCBT, GSH, and PGT) for LTC vs non-LTC patients, with baseline severity as a covariate (**See Figure 2.2**). For depression, a significant time by intervention interaction indicated decreases in PHQ-9 scores according to intervention type, $F(2, 42089) = 27.52$, $p < 0.001$. Post hoc tests indicated that iCBT was associated with the greatest improvements in PHQ-9 score (adj. pre M = 13.66, SE = 0.07; adj. post M = 7.33, SE = 0.07, $d = 1.55$), compared to GSH (adj. pre M = 14.06, SE = 0.04; adj. post M = 8.24, SE = 0.04, $d = 1.43$), and PGT (adj. pre M =

13.73, SE = 0.13; adj. post M = 8.86, SE = 0.13, d = 1.20), while those in GSH improved more than those in PGT (all $p < 0.001$). However, there was no three-way interaction, $F(2, 42089) = 0.14$, $p = 0.87$, suggesting that the effectiveness of these treatments is consistent across LTC and non-LTC patients. Likewise, for anxiety, there was no significant three-way interaction between time, intervention, and LTC status, $F(2, 42089) = 0.65$, $p = 0.52$. As with depression, there was a significant interaction between time and intervention, $F(2, 42089) = 72.46$, $p < 0.001$, where iCBT was associated with the greatest improvement in GAD-7 scores (adj. pre M = 12.79, SE = 0.06; adj. post M = 6.66, SE = 0.06, d = 1.63), followed by GSH (adj. pre M = 13.06, SE = 0.04; adj. post M = 7.70, SE = 0.04, d = 1.42) and PGT (adj. pre M = 12.53, SE = 0.12; adj. post M = 8.60, SE = 0.12, d = 1.04). Similarly for functional impairment, there was no three-way interaction between time, intervention, and LTC status, $F(2, 42089) = 0.60$, $p = 0.55$, but a significant time by intervention interaction suggested that the interventions were associated with different degrees of improvement, $F(2, 42089) = 7.55$, $p < 0.001$. While patients in both GSH (adj. pre M = 17.30, SE = 0.07; adj. post M = 11.50, SE = 0.07, d = 0.94) and iCBT (adj. pre M = 16.70, SE = 0.10; adj. post M = 10.90, SE = 0.10, d = 0.94) improved more in functioning than those in PGT (adj. pre M = 17.00, SE = 0.19; adj. post M = 12.30, SE = 0.19, d = 0.77) (all $p < 0.001$), there were no group differences between patients in GSH and iCBT ($p = 0.816$). **Table 2.3** provides an overview of the adjusted means for each clinical outcome at pre- and post-treatment for all patients across each of the interventions.

Figure 2.2. Pre-treatment to post-treatment observed means in depression, anxiety, and functional impairment exhibited by LTC and non-LTC patients across GSH, iCBT, and PGT interventions.



Note. The dashed line indicates the threshold for caseness for each measure and the error bars indicate standard deviation of the mean. For depression and anxiety symptoms, iCBT was associated with the greatest improvements, followed by GSH, and then PGT. For functional impairment, both iCBT and GSH were associated with greater improvement than PGT. For all clinical outcomes, the effectiveness of interventions is consistent across LTC and non-LTC patients.

Table 2.3. Adjusted means of each outcome measure for patients across intervention groups at pre-treatment (pre) and post-treatment (post).

Measures	GSH	iCBT	PGT	F (df)	p
	Pre v. Post M (SE)	Pre v. Post M (SE)	Pre v. Post M (SE)		
PHQ-9	14.06 (0.04) v. 8.24 (0.04)	13.66 (0.07) v. 7.33 (0.07)	13.73 (0.13) v. 8.86 (0.13)	27.52 (2, 42089)	< 0.001***
GAD-7	13.06 (0.04) v. 7.70 (0.04)	12.79 (0.06) v. 6.67 (0.06)	12.53 (0.12) v. 8.60 (0.12)	72.46 (2, 42089)	< 0.001***
WSAS	17.27 (0.07) v. 11.47 (0.07)	16.71 (0.10) v. 10.95 (0.10)	16.96 (0.19) v. 12.25 (0.19)	7.55 (2, 42089)	< 0.001***

Note. GAD-7, generalised anxiety disorder-7 item questionnaire; LTC, long-term condition; PHQ-9, patient health questionnaire-9 item. WSAS, work and social adjustment scale; SE, standard error.

2.3.4 Control Analyses

To explore whether age and gender explained any of the observed effects, these variables were added as covariates to the models for each outcome measure. There was a significant main effect of age in each model (all $p < 0.001$), indicating that older age was generally associated with less severe symptoms. However, each of the observed time by intervention interactions remained significant, suggesting that age and gender did not explain the observed effects in the models. Additional analyses investigating intervention-

specific effectiveness in only the LTC cohort revealed similar patterns as the above analyses (see **Supplementary Materials 8.1.3**).

2.4 Discussion

This study investigated the effectiveness of low-intensity psychological interventions for treating comorbid depression and anxiety in LTC patients through retrospective analyses of large-scale patient data from a real-world mental health service. As these low-cost, scalable interventions become more mainstream in mental healthcare services, it is important to establish their therapeutic benefit in new cohorts, tested in real-world contexts. The study found that LTC patients presented with varying demographics and clinical characteristics compared to their non-LTC counterparts (i.e., LTC patients were older, more depressed, and more functionally impaired). Accounting for gender, age, and symptom severity differences, statistically significant differences were observed in the way LTC and non-LTC patients improved from low-intensity psychological interventions, where LTC patients yielded greater reduction in depression symptoms and non-LTC patients yielded greater reduction in anxiety symptoms. However, it is important to note that the overall effect sizes of these treatments were large for both cohorts (all Cohen's $d > 1.4$), with marginal differences between them (i.e., Cohen's d differences < 0.08) that amounted to <1 score difference on both depression and anxiety measures post-treatment. The aim of CBT-based interventions is to address specific mental health issues. In line with this, the low-intensity psychological interventions employed here did exactly what they were purposed to do, which is to treat the symptoms of depression and anxiety regardless of LTC status. This observation of a general improvement in depression, anxiety, and functional impairment irrespective of LTC status constitutes some of the first evidence from large-scale real-world patient data to support the use of these interventions to treat comorbid anxiety and depression in LTC patients.

Consistent with the above, there was no effect of LTC status when investigating the differential effectiveness of three low-intensity psychological interventions for LTC and non-LTC patients. While all interventions were shown beneficial, iCBT was found to be the most effective in reducing both depression and anxiety symptoms, relative to PGT and GSH, for both cohorts. A similar pattern of results were observed when the LTC cohort was examined specifically. Overall, the finding that iCBT is generally more effective than GSH and PGT for reducing depression and anxiety in general patients enrolled in IAPT has been reported elsewhere (Palacios et al., 2023). We thus focus on discussing the

relevance and implication of our results for LTC patients. iCBT employs the same principles and components as traditional face-to-face CBT, with many additional benefits that may prove particularly beneficial for the LTC cohort. Murray (2008) states that the combination of online psychoeducation with clinician support allows patients to interpret and internalise available information. This can lead to change in and interaction between knowledge motivation, emotional state, and self-efficacy for improved health behaviours, and in turn improve LTC-related outcomes (e.g. understanding risks may alter anxiety which can lead to enhanced motivation to improve health) (Murray, 2008). Furthermore, by integrating technological advances into the treatment approach, iCBT bypasses the need for in-person attendance by presenting an online solution that is stigma-free and easily accessible, which can facilitate LTC patients' access to healthcare due to their limited physical capacity (van Beugen et al., 2014). In relation to GSH, LTC patients may also find its remote service delivery to be beneficial in terms of increased flexibility and accessibility. While clinician support in GSH may not be as frequent or readily available as that in iCBT, research has suggested adding an element of guided support performs better than no support (Fischer et al., 2020; Karyotaki et al., 2021). In this regard, while one may argue PGT involves the most intensive form of clinician support as it is delivered and facilitated in a face-to-face format, it requires LTC patients to attend in-person, thus hindering the real-world accessibility of this intervention to those with physical limitations within this cohort.

Furthermore, our findings revealed that gender and age did not influence the effectiveness of each intervention, but rather, less severe clinical symptoms were observed in older patients. This is in line with a broader body of research highlighting the association between aging and decreased susceptibility to depression and anxiety (Jorm, 2000) which may be partially due to an accumulation of buffering psychosocial protective factors across the lifespan (Blazer, 2010). Within the context of chronic illness, younger persons with LTCs may potentially experience greater disruption to identity and routine life events than their older counterparts, which may lead to elevated levels of psychological difficulties (Piazza et al., 2007; Wilson & Stock, 2019). According to Lorig and Holman (2003), the use of CBT constitutes a vital component of LTC management in addressing the emotional difficulties accompanying the condition (Lorig & Holman, 2003). The clinically significant improvements in depression and anxiety exhibited by real-world LTC patients in the present study thus builds upon existing research; prior studies and

trials have shown the effectiveness of CBT-based interventions for treating comorbid depression and anxiety across a variety of LTCs (Fiest et al., 2016; Li et al., 2017; Reavell et al., 2018; Zhang et al., 2020). Furthermore, improvements in functioning among LTC patients was also observed post-treatment, which is in line with previous studies highlighting the utility of psychological interventions for improving the impact LTCs can have on one's quality of life, routine functioning, and wellbeing (Anderson & Ozakinci, 2018). All in all, CBT assists LTC patients in gaining a better understanding of their illness by increasing their awareness of negative thoughts and unhelpful behaviours, while allowing them to become active participants in their own wellbeing journey through the development of adequate self-management mechanisms (Anderson & Ozakinci, 2018).

2.4.1 Strengths and Limitations

Several strengths and limitations of the present study must be considered. The study entailed a real-world analysis of a large, naturalistic patient sample enrolled in routine care, thus underscoring the ecological validity of our findings. While the generalisability of our results beyond the service within which this study has been conducted may be limited, patients were recruited from a representative mental health service in Berkshire NHS Trust, UK, of which outcome data is suggested to be comparable to the nationwide data. Nevertheless, given the nature of our retrospective, observational analyses, there may exist biases in treatment allocation that could potentially affect our results (e.g., selection bias in decision-making by clinicians and patients). There were also a lack of differences in the overall and intervention-specific treatment effects between LTC and non-LTC patients, thus warranting caution in over-interpreting the specificity of the order of treatment effects for the LTC cohort. Furthermore, both LTC and non-LTC cohorts in our study experienced, on average, moderate depression and/or anxiety symptomatology for which the low-intensity psychological interventions were intended. Thus, we cannot comment on the effectiveness of these low-intensity interventions for LTC patients who experience more severe depression and/or anxiety and may require more intensive care. Finally, a considerable weakness of the study refers to the lack of information regarding the specific chronic illness of each LTC patient (e.g., type and duration of the LTC, number of comorbidities). This was not available to the authors for analyses in the study, and as such, this study could only shed light on the effectiveness of low-intensity psychological interventions on LTC as a whole, whereas treatment effects may differ

dependent on the type of chronic illness. Our study also lacked data from disorder-specific measures. It may be that accounting for pre-treatment disorder-specific distress levels would influence the intervention effects.

2.4.2 Implications

Comorbid depression and anxiety in patients with LTC pose grave implications to the prognosis of the patient, while increasing the burden on healthcare systems to meet the demands of mental health services. Low-intensity psychological interventions are available to help alleviate this burden. Findings of this study underscore the effectiveness of these interventions, in particular iCBT, in treating comorbid depression and anxiety in LTC patients. This has potential implications for informing decision-making regarding treatment allocation for this particular cohort. Our study further supports the delivery of mental health services via integrated care, and contributes to the literature highlighting many positive effects brought about by the model, including patient satisfaction, improved access, and increased perceived quality of care (Baxter et al., 2018) as well as improved control and reporting of clinical outcomes (Katon et al., 2010b). Beyond the IAPT programme in the UK, the successful implementation of integrated care models has also been observed in other countries including Canada, the US, and Australia (Baxter et al., 2018).

2.4.3 Future Research

Stemming from our approach, future research should expand the scope by examining real-world data from multiple, geographically different IAPT services to further increase the generalisability of our findings. Comparing different services across the UK would also help identify and control for confounding covariates. While our results demonstrated no clinically significant differences in the way low-intensity psychological interventions work for LTC and non-LTC patients, recent evidence has suggested that generic CBT-based interventions might fail to acknowledge the role of LTC in mental illness. However, tailored interventions, that consider the complex interactions between the physical and mental health conditions may make treatments more meaningful and relevant for the LTC cohort (Hind et al., 2010). Future research could thus examine the extent to which tailored versus standardised low-intensity CBT-based interventions in treating comorbid depression and anxiety in LTC patients, so to verify whether the personalised version of the interventions result in greater improvements. Future research

should also consider including disorder-specific measures, such as the diabetes-distress scale (Fisher et al., 2012), as these are likely more capable of capturing the particular psychological distress associated with each chronic illness. These measures would also likely be more sensitive to tailored interventions, and significant improvement on disease-specific outcomes could in turn provide justification for tailoring interventions to specific cohorts (van Beugen et al., 2014). Another point of interest for future research regarding iCBT is its cost-effectiveness. The fact that iCBT can reduce significant public and individual costs can be hugely beneficial (Kumar et al., 2017). Evidence-based research on this key concept of health economics with particular consideration of the burdens associated with LTC is warranted to inform the wider implementation of iCBT in healthcare systems at both national and international levels.

2.5 Conclusion

With long-term conditions exerting a life-changing impact on an individual's wellbeing and functioning, it is vital to have interventions in place that are effective in mitigating the psychological comorbidities accompanying these physical illnesses. The current study is the first to highlight the effectiveness of low-intensity CBT interventions for treating comorbid, clinically significant depression and anxiety symptoms in people with LTCs in a naturalistic routine care setting. The treatment effects demonstrated in this study support the implementation of low-intensity psychological interventions, particularly iCBT, as effective treatments for LTC patients. Further extension of iCBT as a treatment for depression and anxiety in people with LTC will alleviate the burden on healthcare systems by meeting the increasing demands for mental health services amongst this population.

Chapter 3 – The Precision in Psychiatry (PIP) Study: Testing an Internet-Based Methodology for Accelerating Research in Treatment Prediction and Personalisation

3.1 Introduction

A range of evidence-based treatments for depression exist, including pharmacotherapy, psychological therapies, and neurostimulation. These treatments work on average, but not all patients benefit. In fact, clinical trial data suggests that only 50% of patients respond to the initial treatment they receive, with just 30% achieving remission (Andersson et al., 2019c; Trivedi et al., 2006). Many patients must try multiple, sequential and/or parallel treatments on a trial-and-error basis, each taking weeks or months for potential therapeutic effects to unfold, without guarantee of success (Rush et al., 2006; Warden et al., 2007). This leads to sustained human suffering, accumulation of side-effects, and substantial economic costs (Al-Harbi, 2012; Crown et al., 2002).

One potential approach to reducing trial-and-error in psychological treatment is to develop data-informed tools that can assist mental health practitioners in prescribing the best treatment for each individual patient (Cohen & DeRubeis, 2018). This type of 'precision medicine' approach is not new; for more than two decades, researchers have studied the potential predictive power of a wide range of factors including socio-demographics, clinical characteristics, as well as biomarkers derived from genetic, biochemical, and neuroimaging data (McMahon, 2014; Perlman et al., 2019). While numerous factors have been observed to have an association with treatment response in individual studies, effect sizes are mostly too small to have real-world clinical value (Perlis, 2016; Perlman et al., 2019).

A solution to this problem may lie in the development of multivariable models that are informed by data from complementary domains, such as cognitive, (neuro)physiological and molecular data (Gillan & Whelan, 2017; Hawgood et al., 2015; Kessler, 2018). Machine learning is one such method that can iteratively and contemporaneously analyse multiple variables and their interaction, aggregating small individual effects into single predictive values (Chekroud et al., 2021). Using machine learning to optimise treatment approaches is promising, but to-date the published work in depression has suffered from quality issues. A recent review on predicting treatment outcomes in depression

highlighted that out of 54 published studies, just 8 met basic quality control standards of including a large sample size (i.e., > 100 participants) and an adequate validation method (Sajjadian et al., 2021).For those studies that have large sample sizes, data tend to come from clinical trials that have access to only a small number of variables per patient (Ermers et al., 2020; Lee et al., 2018; Sajjadian et al., 2021). This was the case for a model developed by Chekroud and colleagues (2016) that identified 25 self-report demographics and clinical measures which predicted treatment response to antidepressants with 60% accuracy (49% sensitivity and 71% specificity) in their held-out test dataset. Subsequently, Iniesta and colleagues (2018) trained an algorithm using a combination of clinical and molecular genetic data, which achieved high predictive performance (0.77 in the area under the receiver operating characteristic curve) in external data (69% sensitivity and 71% specificity). This suggests that incorporating different data modalities may be an important next step for improving the performance of these models.

One way to acquire these datasets is through multisite collaboration via research consortia such as the EMBARC (Trivedi et al., 2016), PReDICT (Dunlop et al., 2012), and iSPOT-D studies (Williams et al., 2011). These large randomized controlled trials are gold-standard but are costly, time-consuming, resource-intensive, and due to the involvement of many sites, are logistically complex. Therefore, there is a growing need to find alternative methodologies that can complement these approaches, providing us with larger datasets, more rapidly, and in more diverse populations.

To address these gaps, the Precision in Psychiatry (PIP) study used a novel internet-based protocol to recruit, comprehensively assess, and follow through time, mental health patients about to initiate internet-delivered cognitive behavioural therapy or antidepressant medication. Here, we tested the feasibility of this internet-based methodology in collecting large-scale patient data of various types, at home and in a flexible manner. We outline in detail the design of this study, the patient demographic and clinical characteristics, pre-post clinical changes, study attrition, schedule compliance, treatment adherence, data quality, and qualitative patient-perspectives gathered from exit surveys. In examining these facets of the study, we aim to provide guidance for the design of future internet-based studies by highlighting which factors favourably influence recruitment and data collection. We discuss the benefits and limitations of this methodology and make suggestions for future studies adopting a similar approach.

3.2 Methods

3.2.1 Participant Identification and Recruitment

Internet-Delivered Cognitive Behavioural Therapy. Participants receiving clinician-guided iCBT on the SilverCloud Health platform were digitally recruited from two sites: (i) a National Health Service (NHS) mental health service 'Talking Therapies' based near Reading, West London, United Kingdom (Berkshire Foundation Trust) and (ii) a mental health charity based in Dublin, Ireland (Aware Ireland) that provides free education programs, and information services for the public impacted by mood-related conditions. A key difference across sites was that at Talking Therapies, patients have an initial consultation with a clinician who assesses the patient's needs before deciding whether to offer them an iCBT program via SilverCloud. In contrast, individuals recruited via the Aware charity are self-referring. At both sites, the iCBT intervention includes clinician support via the platform. At Berkshire, clinicians are made up of specially trained psychology graduates called Psychological Wellbeing Practitioners (PWPs). At Aware, graduate volunteers provide the clinical support. All supporters have been trained in using the platform. Potential participants at each site received an automated 'Welcome' email upon registering for SilverCloud, which contained an invitation to participate in this study via a web-link.

Antidepressant Medication. Individuals initiating antidepressant medication were recruited internationally using a combination of online (Google Ads, social media platforms, mental health charities) and in-print advertisement campaigns (pharmacies, general practitioners, counselling clinics, newsletters). Participants in the antidepressant arm were asked to provide details on the type and dosage of the antidepressant medication treatment they were prescribed, and to upload a photograph of their prescription for verification purposes. Participants were not required to be medication-free prior to starting the study. They were eligible to participate if they were about to experience a change in pharmacotherapy; initiating, switching, or adding medication.

3.2.2 Screening and Study Entry Requirements

Screening. In both treatment arms, participants read the information sheet online and provided electronic consent. Participants were notified that their participation was entirely voluntary and would not impact on their care in any way, and that their clinician would not be notified about their participation. They were also informed that they were free to

terminate or alter their treatment at any time during the study, and that this would not affect their ability to continue to participate and receive payment. After providing informed consent, participants in both arms were directed to a screening survey used to determine their eligibility. Participants provided their age, English language fluency, email address, listed medications they were taking, confirmed computer access, and told us where they heard about the study. Participants indicated whether they had already started treatment, or if they were planning to start in the future and provided an approximate treatment start date. As stated above, participants in the antidepressant arm also provided a photo of their prescription, which was manually checked for drug name, dose, and date prior to their admission. All participants completed the Work and Social Adjustment Scale (WSAS) (Mundt et al., 2002) which was used to determine eligibility.

Inclusion/Exclusion Criteria. Participants in both arms were excluded if they were not between 18–70 years of age, were not fluent in English, or reported that they did not have computer access. Participants were also required to score a 10 or above on the WSAS (Mundt et al., 2002) which is a transdiagnostic measure capturing impairments in daily functioning arising from mental health problems. In the antidepressant arm, participants were required to have recently started (< 2 days ago) or be planning to start/change treatment soon (< 2 days from now). If they indicated that they were planning to start their treatment in > 2 days after the study sign-up date, they were contacted via email and advised to reapply for the study closer to their treatment start date. In the iCBT arm, participants were invited to our study via automated email directly following their registration on the SilverCloud platform. Given the self-paced nature of iCBT (i.e., users undertake the treatment at their own pace, and can freely choose the order of intervention modules and content they complete in), we also asked participants to indicate when they planned to start treatment. Participants who indicated that they had already started iCBT > 2 days prior to signing up were not included in the study, and those who indicated that they planned to start in > 2 days were contacted via email, and a treatment start date and study schedule agreed upon with the research team manually. In those cases, patients had technically registered on the platform, but were not planning to engage in the modules immediately for various personal reasons.

3.2.3 Study Schedule

If participants were deemed eligible to take part in the study, they were sent an individualised study schedule and a web-link for completing the baseline assessment.

While we endeavoured to have participants complete the baseline on the same day they initiated treatment, we took a pragmatic and flexible approach, allowing participants a window of 4 days from their treatment start date in which to complete the baseline assessment. Four participants in the antidepressant arm completed their baseline assessments 5 days after their treatment start date due in part to administrative issues, and we chose to retain their data for analysis. In the iCBT arm, there were no participants outside of this criterion. Weekly check-in assessments and the final assessment for each participant were approximately scheduled at a 7-day interval following their treatment start date and were provided to participants 1 day before they were due with the instruction to complete them on the following day. **Figure 3.1** shows an overview of the study design and the assessments involved at each timepoint.

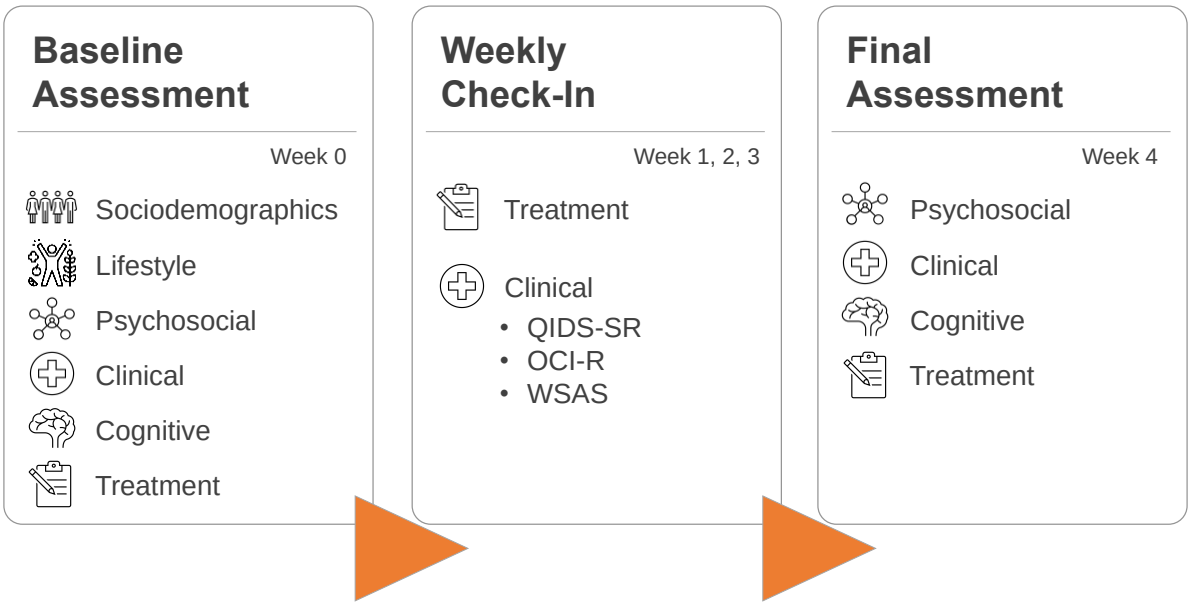


Figure 3.1. An overview of study design. Participants who gave informed consent and met our inclusion / exclusion criteria were invited to complete the baseline assessment, comprising cognitive tests, and a variety of self-report questions concerning participants' treatment, clinical symptoms, psychosocial factors, lifestyle, and socio-demographics. Participants were sent an invitation for a weekly check-in assessment on a scheduled basis for 3 consecutive weeks, which tracked any changes in clinical symptoms and treatment adherence. Participants completed the study with a fifth and final assessment after 4 weeks of treatment, which was an abbreviated version of the baseline assessment.

3.2.4 Outcome Measure

The primary outcome measure for this study is percent change in depression symptom severity on the Quick Inventory of Depressive Symptomatology – Self-Report (QIDS-SR) (Rush et al., 2003). For certain types of planned machine learning analysis that require

categorical outcomes, we binarize this as 'early response', defined as ≥ 30% pre-post improvement. We selected this because (i) participants are not expected to achieve 'response' (i.e., ≥ 50% pre-post improvement in QIDS-SR) or 'remission' (i.e., QIDS-SR score of ≤ 5) in a 4-week timeframe and (ii) prior work has shown that this threshold of early response is a strong indicator of 8-week clinical outcomes (Nierenberg et al., 2000).

3.2.5 Baseline Assessment

The baseline assessment took approximately 1.5–2 hours to complete and required a mouse and a keyboard. Six categories of data were gathered, spanning (i) clinical data, (ii) treatment data, (iii) cognitive test data, (iv) socio-demographics, (v) psychosocial factors and (vi) lifestyle factors (see **Supplementary 8.2.1** for a full outline of variables collected in the study).

Clinical Data. To assess whether treatment has a transdiagnostic effect on mental health, we considered the WSAS as a secondary outcome, measuring general impairment in psychosocial functioning due to mental health problems (Mundt et al., 2002). To assess specific clinical changes, we administered a range of clinical self-report scales assessing obsessive–compulsive disorder measured by the Obsessive–Compulsive Inventory – Revised (OCI-R) (Foa et al., 2002), depression measured by the Self-Rating Depression Scale (SDS) (Zung, 1965), trait anxiety measured by the trait portion of the State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1983), alcohol addiction measured by the Alcohol Use Disorder Identification Test (AUDIT) (Saunders et al., 1993), apathy measured by the Apathy Evaluation Scale (AES) (Marin et al., 1991), eating disorders measured by the Eating Attitude Test (EAT-26) (Garner et al., 1982), impulsivity measured by the Barratt Impulsivity Scale (BIS-10) (Patton et al., 1995), schizotypy measured by the Short Scales for Measuring Schizotypy (SSMS) (Mason et al., 2005), and social anxiety measured by the Liebowitz Social Anxiety Scale (LSAS) (Liebowitz, 1987). These instruments allow for the estimation of 3 transdiagnostic dimensions (anxious-depression, compulsivity, and social withdrawal) based on factor loadings identified in a prior study (Gillan et al., 2016). These transdiagnostic dimensions have been shown to map onto certain aspects of cognition better than standard questionnaires, such as model-based planning (Gillan et al., 2016; Seow et al., 2021) and metacognitive bias (Rouault et al., 2018; Seow & Gillan, 2020). In addition to self-report symptoms, we assessed our participants' history and chronicity of mental health problems. More specifically, we assessed the number of mental health episodes they have experienced,

what age they were when they experienced their first mental health episode, the duration of their current mental health episode, the number of psychiatric diagnoses they had, and the number of close family members with psychiatric diagnoses. As previously mentioned, Chekroud and colleagues' study (2016) developed a predictive model that achieved ~ 60% accuracy in predicting antidepressant response in a re-analysis of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) dataset. We further included 8 miscellaneous items from this study in order to recapitulate their model as a benchmark against which to compare our own (see **Supplementary Materials 8.2.1**).

Treatment Variables. Treatment variables included history of medication and/or psychological treatments for mental health, concurrent medication and/or psychological treatments for mental health, as well as participants' expectations about the mental health treatment they were about to initiate. For participants in the iCBT treatment arm, we examined objective engagement data for each participant, which was provided by SilverCloud.

Cognitive Test Data. Participants completed 4 browser-based gamified cognitive tasks in randomised order, interspersed with blocks of self-report assessments as outlined in the previous section. These were implemented in JavaScript and Python, hosted on a server at Trinity College Dublin and were accessible through any commonly used web-browser. Participants completed a two-step decision making task (Daw et al., 2011; Decker et al., 2016) which estimates various reinforcement learning parameters, including separate estimates of model-based and model-free learning, choice perseveration, and learning rate. Prior studies have shown that model-based planning is linked to compulsivity in the general population (Gillan et al., 2016) and compulsive disorders like obsessive–compulsive disorders (OCD) (Gillan & Robbins, 2014) which benefit from antidepressant medication (albeit at higher doses) (Pittenger & Bloch, 2014) and are commonly co-morbid with anxiety and depression (Torres et al., 2006). The second task in our battery is an aversive learning task that manipulates environmental volatility (Behrens et al., 2007) to assess the extent to which participants adjust their learning rate appropriately as volatility increases. A reduced sensitivity to volatility has been previously linked to trait anxiety and the functioning of the noradrenergic system (Browning et al., 2015). The third task we included measures metacognitive bias and sensitivity in the context of perceptual decision making. Individuals who score high on a transdiagnostic dimension of anxious-depression symptoms have lower confidence in their decision-making, while

those high in compulsivity have over-confidence (Rouault et al., 2018; Seow et al., 2021). Our final cognitive assessment was abstract reasoning using a computerised adaptive task based on Raven's Standard Progressive Matrices (Raven, 2000). Reasoning deficits are associated with risk for various mental health conditions (Koenen et al., 2009).

Socio-Demographics. In addition to age, which they reported at study intake, participants self-reported their sex, country of residence, marital status, education level, subjective social status, and employment status.

Psychosocial Variables. Perceived social support was assessed using the Multidimensional Scale of Perceived Social Support (MSPSS) (Zimet et al., 1988), perceived stress was assessed using the Perceived Stress Scale (PSS) (Cohen et al., 1983), experience of stressful life events in the past 12 months was assessed using the Social Readjustment Rating Scale (SRRS) (Holmes & Rahe, 1967), and childhood traumatic experiences were measured by the Childhood Trauma Questionnaire (CTQ) (Pennebaker & Susman, 1988).

Physical Health and Lifestyle. This included exercising habits, smoking habits, dietary quality, current and prior recreational drug use, height, and weight. Physical health comorbidities were measured by the Cumulative Illness Rating Scale (CIRS) (Linn et al., 1968), and somatic symptoms were measured by 5 items pertaining to stomach, back, limbs, head, and chest pain in the Patient Health Questionnaire-15 (PHQ-15) (Kroenke et al., 2002).

3.2.6 Weekly Check-ins

Weekly check-in assessments were sent to participants in each week of the study. They could be completed using a computer, tablet, or smartphone and took approximately 10-15 minutes to complete. Participants had 4 days to complete these assessments or were otherwise excluded from further participation. They completed 3 standardised questionnaires each week, including the QIDS-SR for depression symptoms, the WSAS for impairment symptoms, and the OCI-R for OCD symptoms. In addition, participants also answered questions about treatment adherence, side effects and dosage changes (for those in the antidepressant arm), whether they initiated any other mental health treatments, and other extra relevant information they wished to inform the study co-ordinators after they have begun participation in the study.

3.2.7 Final Assessment

Participants were asked to complete a detailed final assessment after 4 weeks of treatment. This was almost identical to the baseline assessment, comprising 4 gamified cognitive tasks and self-report questionnaires administrated in a randomised order. Self-report variables gathered during the baseline assessment that were not expected to change (e.g., childhood trauma, age, education etc.) were not re-collected (see **eTable 8.2.2.1** in **Supplementary Materials 8.2.2** for schedule of assessments). Contingent on completion of the final assessment, a proportion of participants were invited to complete a short feedback survey on their experience of the study and to provide suggestions for future studies with similar scope and design.

3.2.8 Quality Control

Participants completed their assessments in an at-home environment where traditional experimental control is absent. To understand how this might affect data quality, we included questions to help us identify bad quality data. At the end of both the baseline and final assessments, participants were asked if they were distracted during the session and if so, by what. They were also asked if they had consumed any substances (e.g., alcohol/drugs) 5 hours prior to participation. Participants were assured their continued participation would not be affected by their response. In addition, we included a 'catch question' that was embedded in both the OCI-R and WSAS questionnaires at baseline and in the WSAS questionnaire at all subsequent timepoints. These 6 catch questions asked participants to select a specific answer option if they were paying attention.

3.2.9 Clinical Interventions

Internet-Delivered Cognitive Behavioural Therapy (iCBT). SilverCloud provides low-intensity, clinician-guided iCBT intervention programs for a range of common mental health problems (e.g., 'Space from Depression', 'Space from Anxiety', 'Life Skills', 'Space from Stress'). The programs partially overlap in terms of content, but also have unique components. All follow evidence-based cognitive behavioural therapy (CBT) principles (Richards et al., 2014, 2020). Each module takes approximately 1 hour to complete, and while users can self-pace, they are generally recommended to complete at least 1 module per week. The intervention comprises cognitive, emotional, and behavioural components (e.g., behavioural activation, self-monitoring, activity scheduling, mood, and lifestyle monitoring). Each module incorporates introductory

quizzes and videos, interactive activities, informational content, as well as homework assignments and summaries. Personal stories and accounts from other users are also included into the presentation of the content. The interventions additionally provide tailored content and modules dependent on the user's clinical presentation (e.g., 'Challenging Core Beliefs' module for depressive symptoms; 'Worry Tree' activity for managing symptoms of anxiety). Although the programs are clinician-guided, users are welcome to engage with the modules and content at their own pace and in the order they opt. A clinician, typically an Assistant Psychologist or a Psychological Wellbeing Practitioner (Clark, 2018), trained in the delivery of SilverCloud iCBT programs, is assigned to a user once they have registered and guides their progress through the intervention. During treatment, the clinician reviews the user's progress while leaving feedback and responding to queries. Typically, 6-8 weekly/fortnightly review sessions are offered across the supported period of the intervention (up to 12 weeks), however, this depends on the user's specific needs.

Antidepressant Medication. Participants in the antidepressant group ($N = 92$) were initiating a range of antidepressant medications. Most (86%) were taking selective serotonin reuptake inhibitors (SSRIs), but 13% were taking serotonin-norepinephrine reuptake inhibitors (SNRIs), 7% taking atypical antidepressants, and 2% were taking tricyclic antidepressants (TCAs). Due to polypharmacy, these numbers do not add to 100%. 8% of participants were taking more than 1 antidepressant medication and 5% were taking another non-antidepressant medication. The most common antidepressant medications were Sertraline (40%), Escitalopram (19%), and Fluoxetine (15%) (see **eTable 8.2.2.2**, **eTable 8.2.2.3**, and **eTable 8.2.2.4** in **Supplementary Materials 8.2.2**). Most participants (90%) experienced side effects from their treatment, with the most common including sleep-related problems such as day-time sleepiness (59%) and night-time sleep disturbances (55%), gastrointestinal symptoms (52%), migraines and headaches (36%), and sexual problems (36%).

3.2.10 Compensation

Participants in both arms were paid €60 in an accelerating payment schedule through PayPal or digital gift cards. Participants received €10 for completing the baseline assessment, €20 euros after the third weekly check-in, and €30 upon completion of the final assessment. The feedback survey was optional and compensated with an additional €10.

3.2.11 Data Analysis

In this paper, data are reported on participants who have fully completed the study in the iCBT and antidepressant arms, recruited from 4th February 2019 to 20th July 2021 ($N = 594$). Where appropriate, participants' recruitment trajectory, socio-demographic and clinical characteristics, treatment, and study compliance data (e.g., retention rates) were compared between study arms using chi-square, t-tests, and repeated measures analyses of variance (ANOVA) (for comprehensive results see 'Between-Group Comparisons' in **Supplementary Materials 8.2.2**). The significance of results (p-value) is reported for context and comparison, however, this study is largely descriptive in nature rather than focusing on hypothesis testing. As such, no correction for multiple comparisons was conducted. To assess data quality, we reported on the numbers of participants who were inattentive, distracted or intoxicated; and to assess the impact of this on data quality, we compared response consistency (i.e., the correlation of similar self-reported items) across the groups who were and were not flagged by these criteria. Finally, a qualitative content analysis was conducted on 4 open-ended free-text questions from the online feedback survey. This method allows researchers to quantify concepts in the data by counting the number of times these concepts appeared, thus providing descriptive statistics fit for the quantitative reporting of this data (Hsieh & Shannon, 2005).

3.3 Results

3.3.1 Participant Recruitment and Retention

Detailed information regarding recruitment and retention are presented in **Figure 3.2**. Recruitment for the antidepressant arm began in February 2019 and the iCBT arm began in March 2020.[1] For the iCBT arm, once both sites were active (Aware and Berkshire), we reached a peak recruitment rate of 59 per month, with a mean of 47 (SD = 10.14) (estimated for 12 months, August 2020—July 2021). In the antidepressant arm, active paid and unpaid recruitment via multiple sources spanned February 2019 to March 2020 (13 months), with a peak of 15 per month and a mean of 7 (SD = 3.67). The arm remained open for participants after that time, but active advertising efforts were halted (**Figure 3.3**). At the time of article preparation, screening data from N = 1811 were assessed for eligibility across the iCBT (N = 1507) and antidepressant (N = 304) arms. Of those

[1] The first informal participant of the antidepressant arm was recruited in December 2018, however, as this was during the pilot recruitment phase, this participant and their recruitment date were therefore not considered in the study.

eligible participants, 63% of participants completed the baseline assessment (N = 710), comprising N = 600 in the iCBT arm and N = 110 in the antidepressant arm. For both groups, retention of baseline completers to weekly check-in 3 was excellent at ≥ 92%, only dropping to 84% for the final assessment. For study completers in the antidepressant arm, most were referred from Google Ads (39%), followed by advertisements through pharmacies (25%), social media campaigns (11%), and general practitioners (8%). For the iCBT group, all were referred from SilverCloud and most came from Talking Therapies in the United Kingdom (83%) and the remaining through Aware Ireland (17%).

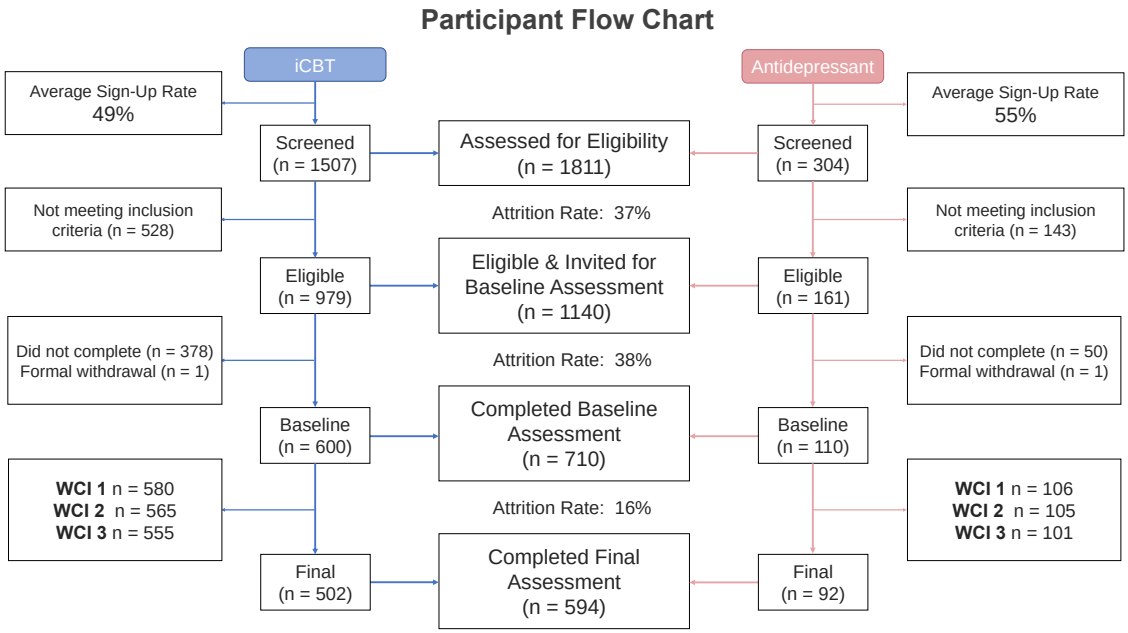


Figure 3.2. Participant flow chart (CONSORT chart). Once they completed the assessment at each study timepoint, participants were progressed onto the next stage of the study. Participants were progressed if they completed the assessments fully at each study stage. If due to technical errors participants were not able to complete specific components of their assessments, it was deemed appropriate to progress them onto the next stage of the study or be financially compensated.

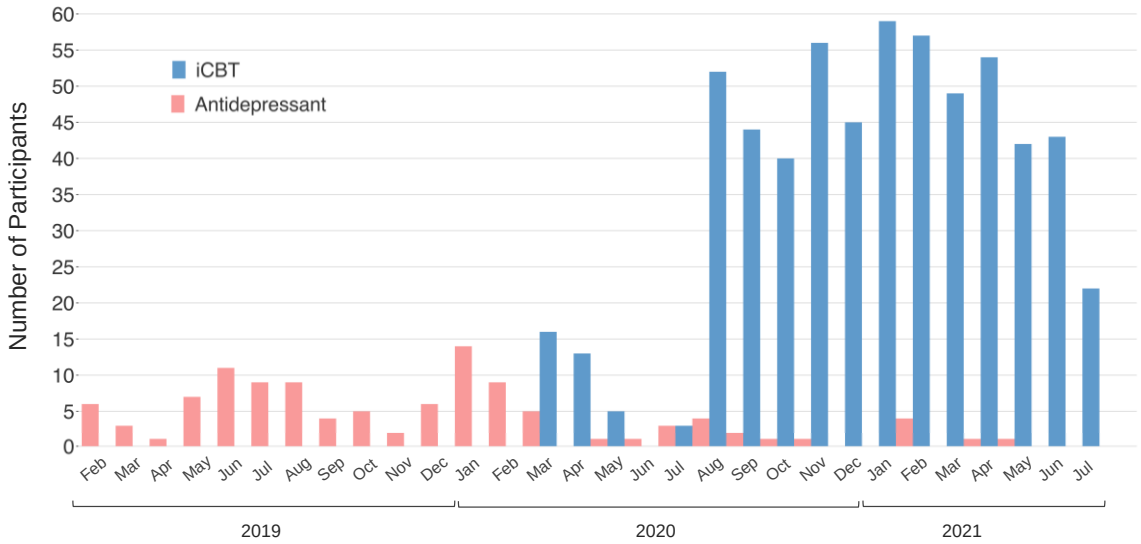


Figure 3.3. Recruitment Rates. Number of participants recruited from each arm from February 2019 to July 2021. The antidepressant arm launched first, initiating recruitment in February 2019. Paid recruitment efforts were focused on a 13-month period from that date to March 2020, when the iCBT arm commenced. The iCBT arm was initiated in March 2020 via Aware Ireland, and in August 2020 recruitment began through Talking Therapies, Berkshire, South London, U.K.

3.3.2 Patient Demographic and Clinical Characteristics

Baseline characteristics of the study completers in both treatment arms are presented in detail in **Tables 3.1** and **3.2** (see **eTable 8.2.2.5** and **eTable 8.2.2.6** in **Supplementary Materials 8.2.2** for baseline characteristics of baseline completers). Participants in both arms were primarily young in their mid- to late-twenties, white, female, employed, third level educated, came from the United Kingdom and Ireland, and subjectively rated themselves on average in the middle of social class status. Most participants reported having 1 or more mental health diagnoses, the most common being depression and/or generalised anxiety. Most participants reported not having a family member with mental health illness, but they themselves have had ≥ 2 lifetime mental health episodes which first began in their adolescence/adulthood. Most participants reported not having engaged in mental health treatment before, and in their self-report of expectations about treatment efficacy on a scale from 0–9 ("I don't expect to feel any better" to "I expect to feel completely better"), the average patient rated a 5.

Table 3.1 Baseline demographic characteristics.

Sample Characteristics	**iCBT**			**Antidepressant**			**t / X^2 (df)**	***p***
	N	%	Median (SD)	N	%	Median (SD)		
Sex	502			92			4.21 (3)	0.24
Female	391	77.89		65	70.65			
Male	107	21.31		25	27.17			

Other	4	0.80		2	2.17			
Country	499			94			86.93 (2)	<0.001
UK	407	81.56		38	41.30			
Ireland	84	16.83		38	41.30			
Other	8	1.60		16	17.39			
Age	501		29 (11.10)	91		26 (9.98)	-1.78 (590)	0.08
Marital Status	502			94			1.39 (5)	0.93
Single	191	38.05		39	42.39			
In a Relationship	150	29.88		28	30.43			
Married	128	25.50		19	20.65			
Divorced	18	3.59		3	3.26			
Separated	14	2.79		3	3.26			
Widowed	1	0.20		0	0.00			
Education Level	502			94			3.64 (2)	0.16
<Third Level	122	24.30		14	15.22			
Some/Complete Third Level	268	53.39		55	59.78			
>Third Level	112	22.31		23	25.00			
Employment Status	502			94			12.81 (2)	0.002
Employed	346	68.92		46	50.00			
Unemployed	150	29.88		45	48.91			
Retired	6	1.20		1	1.09			
Subjective Social Status[a]	502		4 (1.68)	92		4 (2.06)	1.03 (592)	0.30

Outliers were not excluded in the descriptive analyses of demographic characteristics. [a]Subjective Social Status is measured by the MacArthur Scale of Subjective Social Status (i.e., the SES ladder) (Adler et al., 2000). The scale has a range of 0-10, where the higher the score, the higher the rating of subjective social status.

Table 3.2. Baseline clinical characteristics.

Sample Characteristics	**iCBT**			**Antidepressant**			***t* / X^2 (df)**	***p***
	N	%	Median (SD)	N	%	Median (SD)		
No. of Current Diagnosis	502			92			21.67 (2)	<0.001
None	155	30.88		8	8.70			
One	183	36.45		37	40.22			
>One	164	32.67		47	51.09			
Types of Diagnoses[a]	502			92			9.57 (5)	0.09
None	155	30.88		8	8.70			
Depression	245	48.80		70	76.09			
GAD	209	41.63		48	52.17			
Panic Disorder	25	4.98		4	4.35			
PTSD	20	3.98		13	14.13			
OCD	23	4.58		4	4.35			
Others	41	8.17		13	14.13			
Family with Mental Disorders	502			92			1.80 (3)	0.62
None	207	41.24		32	34.78			
One	156	31.08		30	32.61			
Two	81	16.14		16	17.39			
≥Three	58	11.55		14	15.22			
No. of Lifetime Episodes	494			91			11.09 (2)	0.004
<2	53	10.73		7	7.69			

2-5	246	49.80		31	34.07			
>5	195	39.47		53	58.24			
Age of onset (years)	492			92			8.68 (2)	0.01
Childhood (1-12)	86	17.48		23	24.47			
Teenage (13-17)	211	42.89		47	50.00			
Adulthood (18-70)	195	39.63		22	23.40			
Current episode length (days)	457		199 (2557)	84		190 (2463)	0.05 (539)	0.96
History of Past Treatment	502			92			5.95 (3)	0.11
Never Before	224	44.62		30	31.91			
Psychotherapy & Medication	115	22.91		28	29.79			
Medication only	82	16.33		14	14.89			
Psychotherapy only	81	16.14		20	21.28			
Treatment Expectation (0-9)	502		5 (2.04)	92		5 (1.89)	-1.95 (592)	0.05

Outliers were not excluded in the descriptive analyses of clinical characteristics. [a]Types of Diagnoses: The total number of diagnoses type exceeds the sample size of baseline completers (i.e., participants have the option to pick more than one diagnosis).

In terms of clinical severity at baseline, participants in the antidepressant arm had a mean QIDS-SR score of 16.51 (SD = 4.17) and a mean WSAS score of 22.73 (SD = 6.84), indicating severe depression (Rush et al., 2003) and functioning (Mundt et al., 2002), respectively. Participants in the iCBT arm had a somewhat lower QIDS-SR score of 13.86 (SD = 4.28) at baseline, corresponding to moderate depression severity, and a mean WSAS score of 19.02 (SD = 6.65), also falling in the moderate range (**Figure 3.4**). Clinical severity of other symptoms assessed at baseline are presented in **Table 3.3** (and **eFigure 8.2.2.1** in **Supplementary Materials 8.2.2**).

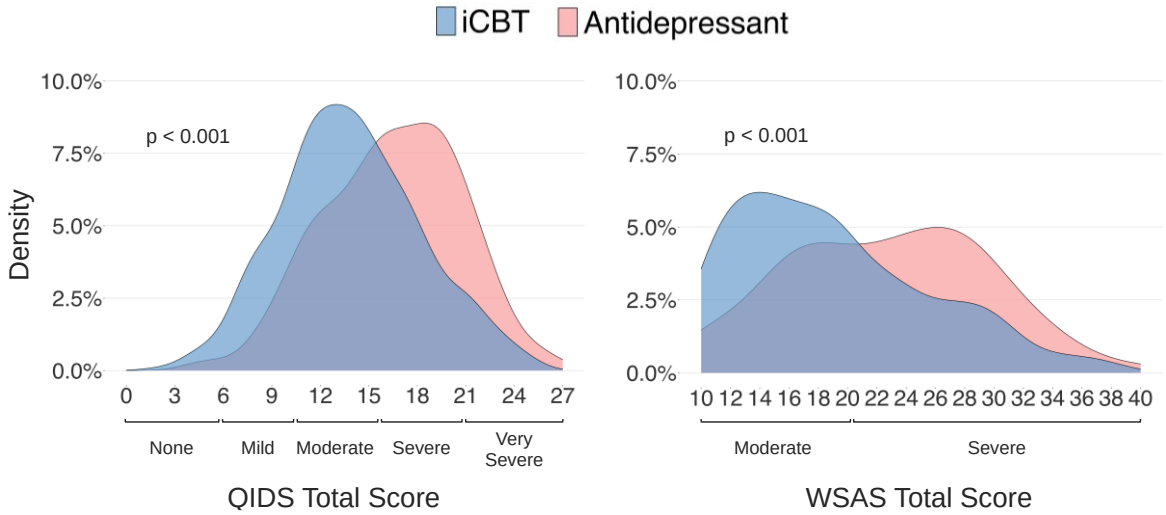


Figure 3.4. Baseline clinical symptom score distribution of depression (QIDS) and impairment (WSAS) for participants in the iCBT and antidepressant arm.

Table 3.3. Baseline clinical symptom scores across treatment arms.

Clinical Symptoms	iCBT		Antidepressant		t(df)	*p*
	Mean	SD	Mean	SD		
Depression (QIDS-SR)	13.86	4.28	16.51	4.17	5.47 (592)	<0.001
Impairment (WSAS)[a]	19.02	6.65	22.73	6.84	4.61 (580)	<0.001
Apathy (AES)	41.89	8.94	43.31	9.46	1.49 (592)	0.14
Alcohol Use (AUDIT)[a]	5.82	5.57	7.74	7.78	2.82 (589)	0.005
Impulsivity (BIS)	67.50	11.15	68.21	10.88	0.56 (592)	0.57
Eating Disorder (EAT)	12.15	10.83	13.99	12.38	1.46 (592)	0.14
Social Anxiety (LSAS)	38.45	16.62	40.89	17.14	1.29 (592)	0.20
OCD (OCI-R)	23.45	12.64	25.38	12.87	1.34 (592)	0.18
Schizotypy (SSMS)	19.09	6.87	22.10	7.50	3.81 (592)	<0.001
Depression (SDS)	54.17	8.09	58.55	7.91	4.80 (592)	<0.001
Trait Anxiety (STAI)	61.13	8.78	65.08	8.66	3.97 (592)	<0.001

[a]At baseline, N = 3 were missing AUDIT symptom score and N = 12 were missing WSAS symptom score.

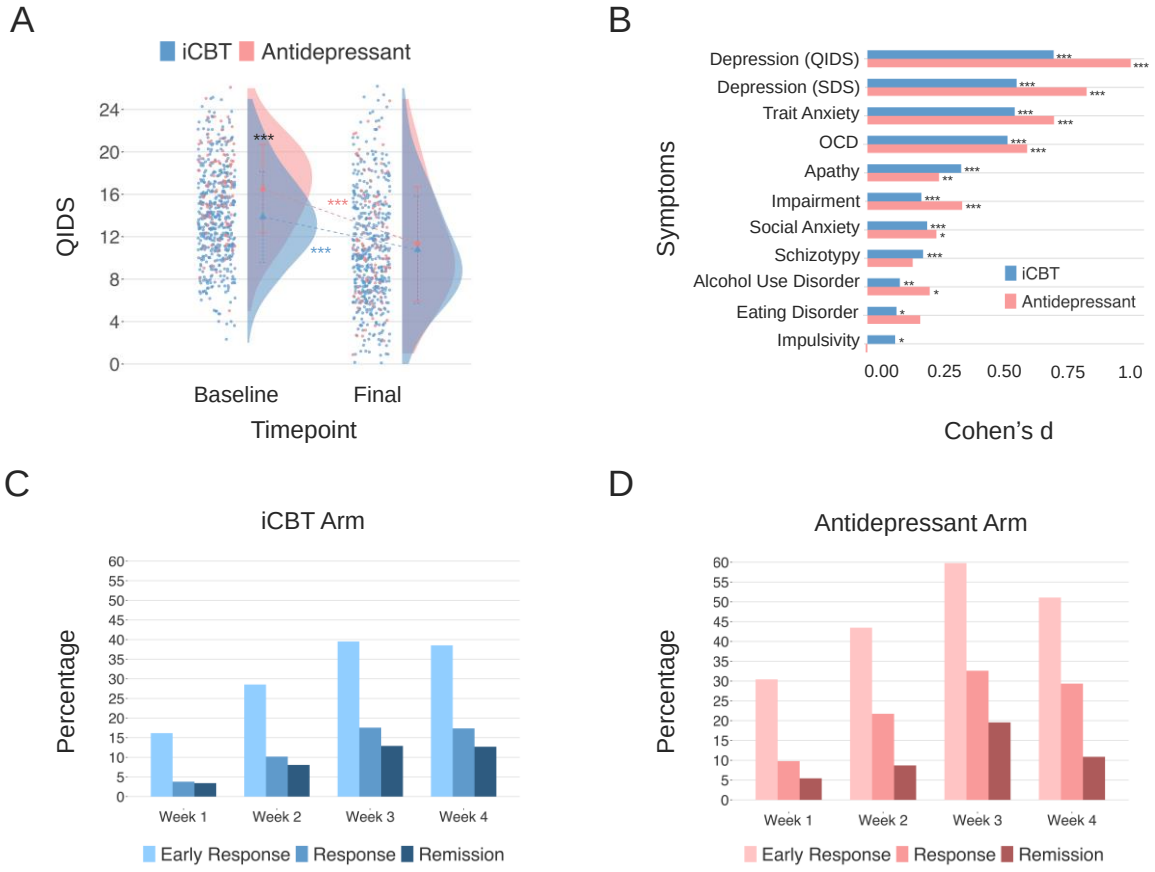
3.3.3 Treatment Adherence and Concurrent Treatments

For both groups, the treatment adherence was high by week 3, where over ≥97% of the iCBT group reported still undergoing treatment (i.e., 98% at WCI 1 and WCI 2 and 97% at WCI 3) and ≥98% of the antidepressant group reported still taking antidepressant medication (i.e., 99% at WCI 1 and WCI 2 and 98% at WCI 3). In the antidepressant arm, 4 participants altered the dosage of their medication during the study participation (N=2 took less than prescribed and N=2 took more than prescribed). For the iCBT group, 29% were receiving another treatment during the study (N=148), of which 39 (8% of total iCBT arm) were taking concurrent medication and 109 (22%) were receiving a concurrent form of psychotherapy. For the antidepressant group, 9% (N=8) were taking at least one other medication for a mental health problem at the time of study intake and 36% (N=33) were receiving some form of psychotherapy. Thus, there were partial overlaps in the treatments received across the two study arms.

3.3.4 Pre-Post 4-Week Clinical Changes

Participants in both arms experienced significant improvement in depression symptoms after 4 weeks of treatment. Participants in the antidepressant arm experienced a significantly larger percent reduction in QIDS-SR from baseline than those in the iCBT arm, $t(589) = 2.73$, $p = 0.007$, even after controlling for imbalances in baseline severity,

F(1, 588) = 4.36, $p = 0.04$.[2] **Figure 3.5C**, **3.5D** show the weekly percentage distribution of participants achieving early response, response, and remission throughout the study for each of the two treatment arms. For the iCBT arm, by week 4, 39% of participants have achieved early response (i.e., a 30% reduction), 17% of participants have achieved response (i.e., a 50% reduction) and 13% of participants have achieved remission (i.e., a score ≤ 5). Participants in the antidepressant arm exhibited a significantly higher rate of early response at 51%, $X^2 = 5.09$ (1), $p = 0.02$, as well as rate of response at 29%, $X^2 = 7.19$ (1), $p = 0.007$, but no significant difference in their remission rate of 11%, $X^2 = 0.24$ (1), $p = 0.62$. In terms of absolute score change, in the iCBT arm, QIDS-SR depression scores were significantly reduced by an average of 3.09 points (SD = 4.31) (21%), t(500) = 16.06, $p < 0.001$, with a moderate effect size, $d = 0.72$. In the antidepressant arm, this reduction was larger with an average of 5.18 points (SD = 5.11) (31%) on the QIDS-SR, t(91) = 9.73, $p < 0.001$, with a large effect size, $d = 1.01$. A two-way ANOVA confirmed this difference was significant, F(1, 591) = 17.26, $p < 0.001$ (**Figure 3.5A**).



[2] N = 2 in the iCBT arm identified as outliers were removed for this analysis (i.e., they have QIDS % changes larger than ± 3 standard deviations from mean).

Figure 3.5. Clinical change in QIDS-SR. (**A**) Pre-post 4-week QIDS-SR score reduction. Both treatment arms experienced significant decreases in depression score measured by QIDS-SR from the baseline to the final assessment. (**B**) Effect sizes and statistical significance of clinical symptom reduction in both treatment arms. All clinical symptoms reduced significantly from the baseline to final assessment in both treatment arms except for schizotypy, eating disorder symptoms, and impulsivity in the antidepressant arm. (**C**) Percentages of early response, response, and remission achieved by participants in the iCBT arm at each study timepoint. (**D**) Percentages of early response, response, and remission achieved by participants in the antidepressant arm at each study timepoint.

In terms of general functional impairment (WSAS), there were modest but significant improvements in both treatment groups. Participants in the iCBT arm saw their self-reported impairment reduce by 1.57 points (SD = 7.55) (8%), $t(499) = 4.65$, $p < 0.001$, $d = 0.21$ and those in the antidepressant arm reported reductions of 3.09 points (SD = 8.46) (14%), $t(79) = 3.27$, $p = 0.002$, $d = 0.37$. These percentage changes did not differ significantly across the treatment arms, $t(572) = 1.36$, $p = 0.17$.[3] Consistent with a transdiagnostic perspective on mental health, clinical gains extended beyond depression symptoms and daily functioning. Analyses revealed significant reductions in most clinical symptoms gathered in both treatment arms (all $p < 0.05$), except for schizotypy ($p = 0.10$), impulsivity ($p = 0.95$), and eating disorder symptoms ($p = 0.05$) in the antidepressant arm (**Figure 3.5B**).

3.3.5 Study Schedule Compliance

On average, participants in both arms completed the baseline assessment ~ 1 day after initiating treatment (Antidepressant: Mean = 1.24, Median = 1, SD = 1.64, range = -3 to + 5 days; iCBT: Mean = 0.77, Median = 1, SD = 1.45, range = -2 to + 4 days). For the iCBT cohort, we had the benefit of some objective data from the iCBT provider to complement the self-report. The median difference between self-reported treatment start date and the day participants first registered on the iCBT platform was + 1 day (Mean = 1.76, SD = 2.90, range = -1 to + 20) (i.e., they registered for iCBT on the platform 1 day before they reported to us they would start treatment). The due date for all subsequent assessments were based on the self-report treatment start date, regardless of whether their last assessment was completed slightly early or late. Weekly check-in assessments and the final assessment were provided to participants 1 day before they were due with the instruction to complete them on the following day. Despite this

[3] N = 4 in the iCBT arm and N = 1 in the antidepressant arm identified as outliers were removed for this analysis (i.e., they have WSAS % changes larger than ± 3 standard deviations from mean).

instruction, we found that many participants completed them immediately upon receipt (i.e., 1 day before due date). From the treatment initiation date, weekly check-in 1 was completed on average on day 6 (Mean = 6.78, SD = 1.33), but there were a handful of longer intervals (range = 3–15), weekly check-in 2 was completed on average on day 13 (Mean = 13.88, SD = 1.40, range = 8–22), and weekly check-in 3 was completed on average on day 20 (Mean = 20.81, SD = 1.35, range = 15–28). The final assessment, which was more time-consuming than the weekly check-ins, was completed on average on day 28 (Mean = 28.76, SD = 1.74, range = 23–37). The median interval between treatment initiation and final assessment was 28 days (Mean = 28.70, SD = 1.59, range = 24–36) in the iCBT arm and 29 days (Mean = 29.15, SD = 2.42, range = 23–37) in the antidepressant arm (**Figure 3.6A**, **3.6B**, and **eFigure 8.2.2.2** in **Supplementary Materials 8.2.2**).

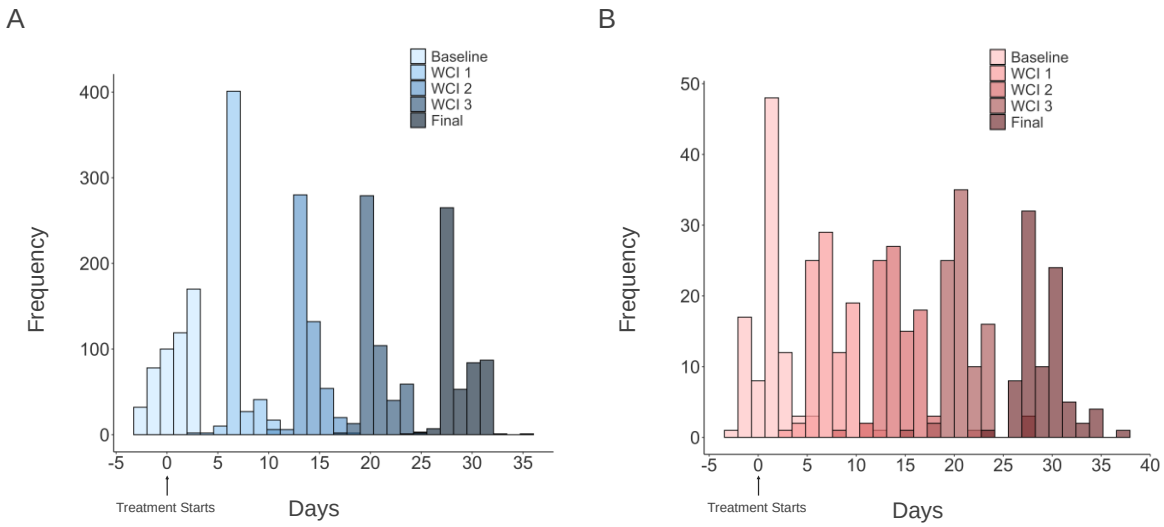


Figure 3.6. Distributions of overlapping completions dates of each study section for (**A**) the iCBT arm and (**B**) the antidepressant arm. Day '0' depicts treatment start date.

Participants were requested to complete the baseline and final assessments in a single sitting, taking short breaks between sections. However, for a variety of practical and technical reasons, some participants were only able to partially complete their baseline or final assessment before returning later to complete the remaining sections. We defined participants as completing the section in 1 sitting if they did not take a break exceeding 4 hours between any of the study sections. Using this cut-off, 9% (N = 47) of participants in the iCBT arm and 23% of participants (N = 21) for the antidepressant arm did not complete the baseline assessment in a single sitting. The trend is similar for the final session, 9% (N = 43) in the iCBT arm and 16% (N = 15) in the antidepressant arm did not

complete it in a single sitting. Of those who completed their assessments in a single sitting, the median time it took participants to complete the baseline assessment was 1.63 hours (SD = 0.77) and the median time to complete the final assessment was 1.30 hours (SD = 0.85). Participants were more likely to complete the brief weekly check-ins during daytime (6am-6 pm: 76%) when compared to the baseline and final assessments (both 59%, $X^2 = 62.71$, $p < 0.001$).

3.3.6 Data Quality

At baseline, 66% reported being distracted in some way (iCBT: N = 310, 65%; Antidepressant: N = 61, 75%). Overall, the most common types of distractions endorsed were family and friends (37%), background noise (32%), and phone (28%). In relation to intoxicating substances, at baseline, just 3% of participants informed us that they had taken 1 of our defined substances within 5 hours of starting the study (iCBT: N = 14, 3%; Antidepressant: N = 4, 5%). Of the very few participants who reported any form of substance use, 13 had consumed alcohol (2%), 5 reported marijuana use (< 1%) and 2 people reported using opiates (< 1%) (see **eTable 8.2.2.7** and **eTable 8.2.2.8** in **Supplementary Materials 8.2.2** for similar trends in distraction and substance use items at the final assessment).[4] In terms of our inattention 'catch questions', 11% (N = 63) of participants failed at least 1 of the 6 attention checks embedded in the study (iCBT: N = 51, 10%; Antidepressant: N = 12, 13%). The majority of inattentive participants were only inattentive at one time (N = 47, 8% of total sample), with just 3% of the total sample (N = 16) failing more than 1 attention check. People were more likely to be inattentive at certain timepoints in the study, $X^2 = 25.32$ (5), $p < 0.001$. The longer, more burdensome assessment sessions had more attention lapses (baseline: check 1 N = 17, check 2 N = 35; final: check 1 N = 21) than the 3 brief weekly check-ins (1 check at each: N = 12, N = 11, N = 10, respectively). To further assess data quality, we examined participants' consistency in reporting their height at the baseline and final assessments. Height reports were reliably measured across the two time points (ICC1 = 0.97), and there was no significant difference in the absolute size of the discrepancy of height reports based on whether participants were classed as inattentive or not, t(67) = -1.66, $p = 0.10$ (**Figure 3.7A, 3.7B**). Finally, we examined the internal consistency of self-report symptom assessments. At baseline, Cronbach's alpha was good

[4]At baseline, N = 33 were missing data for distraction and substance use data quality item checks.

for all scales (i.e., r > 0.7), ranging from 0.71–0.95, and this rose to 0.81–0.95 at the final assessment (**Figure 3.7C, 3.7D**). Full results are presented in **eTable 8.2.2.9** in **Supplementary Materials 8.2.2**.

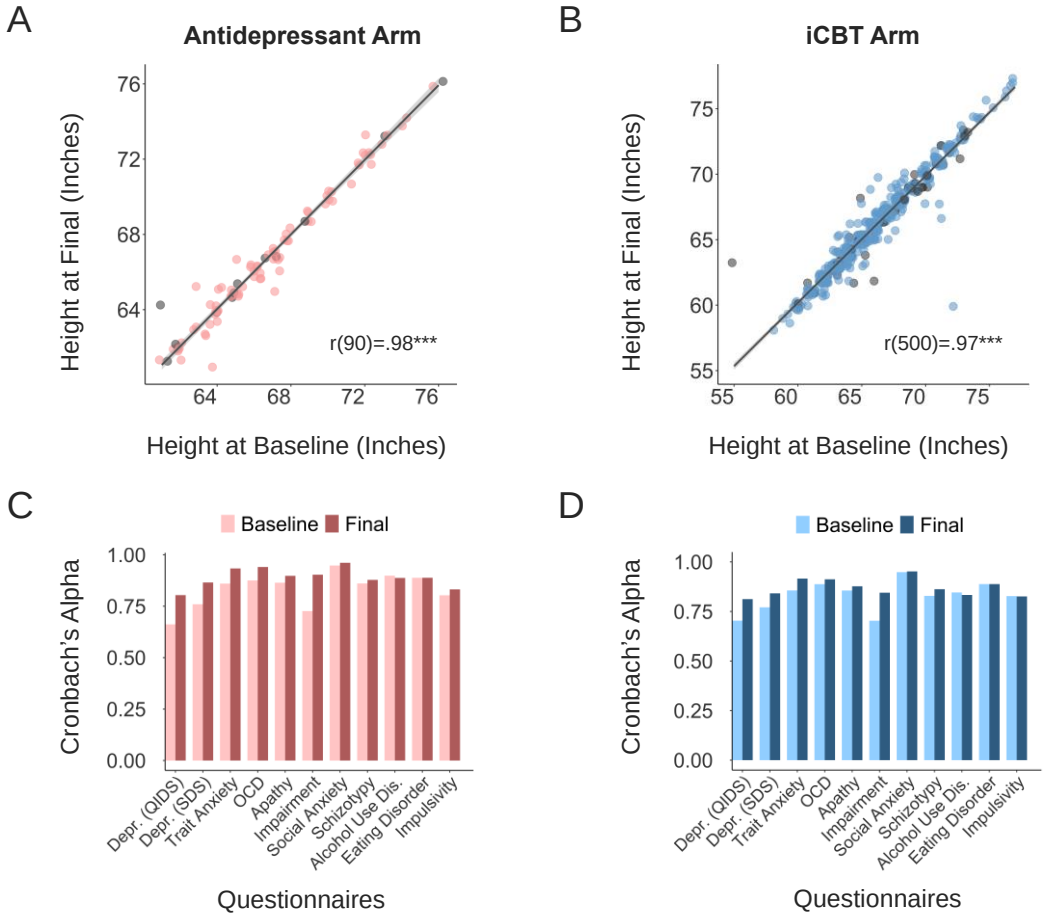


Figure 3.7. Data Quality indicators. Correlation of height (in inches) was gathered at the baseline and final assessments in (**A**) the Antidepressant arm, $r = 0.98$ and (**B**) the iCBT arm, $r = 0.97$. Participants who failed at least 1 attention check are coloured grey. Internal consistency of the self-report questionnaires (Cronbach's alpha) for the (**C**) Antidepressant arm and the (**D**) iCBT arm

3.3.7 Qualitative Feedback

Of 155 invited, 135 participants completed the online feedback survey from 19th June 2020 to 13th October 2020, giving a response rate of 87%. Data were analysed for 4 open-ended free-text questions concerning what participants liked and disliked about the study, what they suggested could be added to the study, and whether they found the payment schedule to be satisfactory (see section 'Qualitative Data Analysis' in **Supplementary Materials 8.2.2**). When asked *"What did you like about the study?"*, the most prominent theme emerging from the responses relates to *Self-Reflection* (40%). Participants liked how the study prompted them to reflect on aspects of their mental health they would not

have done otherwise and helped them keep track of treatment progress through the weekly check-ins. Another major theme was that participants found the study *Easy to Complete* (33%), citing convenience in terms of both online accessibility and flexibility with respect to assessment completion times and dates, the inclusion of breaks and email reminders, and the clarity of instructions. A proportion of respondents said they liked the *Gamified Tasks* (24%) which some found interactive and challenging, and a further 20% of participants reported feeling aligned in general with the *Study's Mission*. Only 4% of participants reported *Payment* as what they liked about the study. Although 24% of participants reportedly liked the games, when asked *"What did you dislike about the study?"*, a large number (89%) also cited the *Gamified Tasks* (89%), which were felt to be tedious (e.g., "repetitive", "boring", "lengthy"), and in some cases confusing, frustrating, and too difficult. The second most prominent theme in response to this question about dislikes referred to the overall *Study Design and Mechanics* (13%). Some did not like the length of baseline and follow-up assessments and overall time-commitment involved, while others had problems with study coordination, administrative or other logistical problems.

In terms of suggestions for additions to the study in future, most of the respondents suggested additional aspects of *Self-Report* (83%). Recommendations ranged from the inclusion of free-text and experience sampling to including a broader range of questions on treatment information, psychological states and behaviours, demographics, lifestyle, physical health, environmental factors, own perceptions of change/symptoms/problems, and positive mood. Another area for improvement pertained to the *Study Mechanics and Design*, including extending the overall study duration and including a longer-term follow-up assessment. A good proportion of participants (30%) reported there was *Nothing* they wished to add to the study. In relation to the payment schedule, participants were overwhelmingly satisfied, with only 7% citing a negative experience (e.g., missed/delayed payments, not worth the time-commitment). Most participants (71%) rated the payment schedule positive or very positive and some (33%) were neutral about it (e.g., "fine", "no complaints", "appropriate").

3.4 Discussion

The adoption of online data collection in psychiatry research has seen a dramatic increase in recent years (Gillan & Daw, 2016), but much of this research remains cross-sectional and correlational in nature. The Precision in Psychiatry (PIP) study extends this

conventional approach to create a foundation for longitudinal treatment prediction research in psychiatry. We recruited and screened a large sample of individuals receiving iCBT (baseline N = 600, final N = 502) and a smaller sample receiving antidepressant medication (baseline N = 110, final N = 92). For eligible patients, we acquired an extensive range of self-report and cognitive measures (> 600 variables) at baseline and after 4 weeks of treatment, in addition to brief weekly check-ins. In what follows, we discuss the benefits and limitations of this approach and put forward recommendations for future studies (**Table 3.4**).

3.4.1 Recruitment at Scale, at Speed

The major success of the study is that it enabled us to recruit a large cohort of patients undertaking treatment for depression in a relatively short period of time. This was most evident in the iCBT arm, where we reached a maximum recruitment rate of 59 patients completing the baseline assessment per month. This corresponded to just over 500 full study completers within one year and a half. While the antidepressant arm was slower and more expensive to recruit for, we nonetheless gathered data from close to 100 individuals in 13 months, which compared to conventional strategies for recruiting participants with clinical diagnoses, was rapid. This approach also benefitted from high retention rate, where 93% of participants were retained for 3 weeks, and that dropped to 84% at week 4, following the final (lengthier) assessment (1.5 hours). If future research does not require detailed cognitive and clinical follow-up data (i.e., studies focused purely on prediction), this suggests one can expect retention > 90%, if follow-up assessments are brief (e.g., restricting to 1 or 2 self-report outcome measures). Qualitative feedback from users suggests that the flexibility of the study design may have helped us to recruit and retain participants. When participants are able to take part from any location, and at any time, this reduces logistic challenges associated with traditional, in-person data collection methods (e.g., travel to in-person locations, 9–5 participation hours) and makes research available to people often underrepresented in research (e.g., those from rural areas, socially anxious, more severely disabled). Treatment adherence for those who remained in the study was very high at >= 97%.

3.4.2 Compatibility with Digital Therapy

Digital psychological interventions such as iCBT are becoming increasingly popular as they allow greater access to care at a reduced cost, while demonstrating similar

effectiveness for those requiring low intensity intervention (Andersson et al., 2019a; Kumar et al., 2017; Webb et al., 2017). There exists, however, little basic research examining the mechanisms of therapeutic change in iCBT, how it affects cognition, brain function, or indeed, who it is best suited to and why. We see this as an important opportunity for future work for several reasons. iCBT lends itself well to systematic research as the therapeutic content that patients have access to is standardised and reproducible, which solves issues of both inter-clinician and intra-clinician variability in the delivery of in-person CBT and leads to more generalisable insights. Individual variability in engagement with the online platform can be tracked precisely via granular and objective treatment data (e.g., what modules, when, and for how long), which can be mined to understand moderators of treatment success (Chien et al., 2020; Enrique et al., 2019). This may be particularly useful for researchers and clinical providers aiming to identify active ingredients of successful CBT, for personalisation, precision and more. This combination of digitised therapy and digitised research may thus provide a much more direct route to real-world clinical integration than other less integrated approaches.

3.4.3 Non-Random Assignment in Naturalistic Design

The observational nature of this study reflects the 'real world' of treatment allocation (i.e., non-randomised), which places a fundamental limit on causal inference. Though it does not solve the problem of non-random assignment, we included more than one observational arm, which allows us to assess the generalisability and specificity of any treatment prediction model we develop to new cohorts and treatments. In terms of demographics, participants in both iCBT and antidepressant arms were primarily white, female, employed, third level educated, which limits the generalisability of these findings. These sample characteristics are comparable to other large-scale studies recruiting for antidepressant treatment (Trivedi et al., 2006) and iCBT (Richards et al., 2020), indicating that this lack of generalisability is not a problem unique to digital treatment research, but something that all research in this area needs to work to address. Participants in the antidepressant arm were marginally younger and more likely to be unemployed, but most notably they had more severe clinical presentations and symptoms than their iCBT counterparts. This was expected as it follows the guidelines of the Improving Access to Psychological Therapies (IAPT) program in NHS, where iCBT is typically prescribed first for mild to moderate depression before antidepressant medication is considered (NICE, 2022). Given this, the finding that by week 4, participants in the antidepressant

arm experienced greater symptom reduction, should be interpreted with caution. Prior research suggests comparable effectiveness of the two treatments (Forand et al., 2019) and the short timeframe of our study and the self-paced nature of the iCBT intervention may have made for a weaker overall 'dose' for this arm. Some participants were receiving concurrent, overlapping treatments (8% of patients in the iCBT arm taking medication, and 36% of the antidepressant arm receiving some form of psychotherapy). This is a significant limitation of the inclusive study design we adopted, and insights regarding specificity or generality of any effects should be supported with sensitivity analyses (i.e., excluding participants undertaking concurrent treatments).

3.4.4 Lack of Experimental Control

While we could not control when and if participants would complete each study section as per their schedule, to make participation as convenient and flexible as possible, we issued each study section one day before due-date and allowed participants a 4-day window to complete it. As a result, participants on average completed assessments 1 day earlier than they were due, and participants overall differed in the intervals between starting treatment and completing baseline and subsequent sessions. While these differences were minimised, issues of timing are some of the most challenging for researchers working with internet-based methods to manage. As previously mentioned, online studies primarily rely on self-report, rather than clinician-assigned diagnoses or severity assessments. This raises legitimate concerns regarding the reliability and validity of online data gathered in a less-controlled environment when compared to traditional, in-lab/in-clinic settings. Online studies can be susceptible to inattentive and careless responding, as is the case for other forms of online research (e.g., crowdsourcing) (Zorowitz et al., 2021). At a minimum, prior research suggests that individuals tend to follow task instructions better when tested in-person versus at home (Ramsey et al., 2016). Our analyses of the quality of the baseline data we gathered revealed that some participants failed to complete their assessments in a single sitting (11%), most reported being distracted during the study session (63%), and a small few had even consumed intoxicating substances (3%). On the extreme end, some online platforms (e.g. Amazon's Mechanical Turk) are suffering from major quality control issues, with a recent paper finding that > 50% of respondents reported their own gender identity inconsistently at two time-points (Donegan & Gillan, 2022). We believe these more serious risks can be mitigated by adopting a more targeted recruitment protocol such as that described here,

i.e., advertising the study to only those individuals eligible, using validation steps (prescription upload, registration requirement for iCBT), and ensuring that the mission of the study is aligned to the goals of the participants (i.e., improving mental health treatment) (Donegan & Gillan, 2022). In terms of our more objective quality checks, just 11% of subjects were caught by any of our attention checks when filling in online questionnaires. Overall, we found the data to be of excellent quality; height (in inches) self-reported at week 0 and week 4 had near perfect inter-rater reliability. This exceeds the 2-week test–retest reliability of height (in adolescents) gathered using paper booklets ($r = 0.93$) (Brener et al., 2003). Internal consistency of the self-report questionnaires administered was also high. QIDS had the lowest Cronbach's alpha at baseline of 0.71, comparable to that observed in another patient sample at treatment outset (alpha = 0.72) (Rush et al., 2003). In terms of measurement properties, inter-rater reliability (clinician agreement) for some of the most common mental health conditions, including depression, is relatively low for the criteria of the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) (Regier et al., 2013), while self-report assessments enjoy much higher reliability, both in-person (Geschwind et al., 2021) and online (Shapiro et al., 2013). Although self-report has these advantages, it may be less valid for use in mental health populations where insight is compromised.

3.5 Conclusion

Depression is a highly heterogeneous disorder for which no single treatment intervention is universally effective. We need to move from a trial-and-error approach to treatment to one that is precise and where possible personalised. To this end, researchers are currently exploring the potential of developing clinical decision tools by training machine learning algorithms to predict clinical outcomes. In order to obtain robust predictions, we need substantially larger sample sizes than is typical in the field. Our data suggest that Internet-based methods can achieve this, allowing us to gather rich, complex datasets from large cohorts, with measurable indicators of treatment adherence and engagement. We hope that the detailed data we have provided in this paper provides a working template for future Internet-based treatment studies in psychiatry.

Table 3.4 Practical Guidance for Internet-Based Treatment Prediction Research

Practical Guidance for Internet-Based Treatment Prediction Research

1. **Keep Assessments Brief**

Retention was high for brief, self-report assessments and in particular weekly check-ins were well-received by patients wishing to track their progress through treatment. Cognitive tests were by far the most disliked component of our study. Considerable work is needed to make these more tolerable for participants.

2. **Ensure Incentives are Aligned**

The key to quality data in an online environment is to keep incentives aligned. Participants in our study resonated with the mission of the study and/or enjoyed the opportunity for self-reflection. Future research should be sensitive to these motivations and (i) communicate the mission of the study clearly, early, and often, (ii) supply participants with information about study outcomes at the time of publication, (iii) solicit feedback from participants and (iv) consider a graphical display where service users can visualise their progress throughout treatment.

3. **Make Participation Easy**

The ease of participation is imperative to achieving successful online recruitment, for example, allowing participants to complete assessments remotely and at a time convenient to them. In addition to a PC/laptop, smartphone and smartwatch may be incorporated in future for increased convenience in online data collection. They can further facilitate the collection of different sorts of data, such as mobility data, sleep, and experience sampling data.

4. **Issue Regular Reminders, be Flexible and Pragmatic**

To encourage retention, a timely reminder for each assessment should be delivered a day prior to due date, and a small window for completion may be provided to increase flexibility for participants to complete each assessment. Sensitivity analyses can be used to ensure late or early assessments do not confound results.

5. **Data Quality is not a Given**

Data quality indicators (e.g., catch questions, distraction probes, and stable variables for high test-retest reliability analysis) should continue to be included for assessing the quality of self-report online data. The online research environment changes and is potentially vulnerable to BOTS or dishonest respondents. To reduce the threat this poses to valid research, recruitment should be targeted to those initiating treatment and include a validation (prescription photo, iCBT registration).

Chapter 4 – Machine Learning Prediction of Depression Symptom Change Following Internet-Delivered Cognitive Behavioural Therapy

4.1 Introduction

Depression, as one of the leading causes of disability worldwide, is a debilitating but common condition affecting millions of individuals every year (World Health Organization, 2023). Psychotherapy such as cognitive behavioural therapy (CBT) is considered a frontline treatment alongside pharmacology (NICE, 2020), and it is increasingly being delivered online to address issues like limited access to and availability of in-person care (Kazdin & Blase, 2011; McHugh & Barlow, 2010). Despite continuous development of internet-based CBT (iCBT), rates of patients benefiting from it remain poor. Similar to antidepressant medication (Trivedi et al., 2006) and face-to-face psychotherapy (Cuijpers et al., 2021a), roughly half of the patients respond to iCBT, with just a third achieving clinically meaningful remission (Andersson et al., 2019a). A considerable reason for this heterogeneity in treatment response may be due to the trial-and-error approach often prescribed in treatment allocation to patients. Identifying individuals likely to benefit from treatment before they start is therefore crucial for advancing this field (Andersson, 2016; Cuijpers et al., 2012).

Researchers have been working to discover markers that could predict depression improvement following iCBT. These include biomarkers such as therapygenetics (e.g., genetic risk scores; Andersson et al., 2019) and neuroimaging markers (e.g., brain region activity and volume; Webb et al., 2018), which may be effective but resource-intensive to collect in practice. Using cognitive data as a proxy for direct brain measurement is more practical, but their signal for predicting treatment response remains unclear (e.g., cognitive flexibility; Lindner et al., 2016; Silfvernagel et al., 2018). Studies commonly investigate sociodemographic and clinical variables such as sex, age, education, marital status, employment, psychiatric history, and psychosocial functioning, as well as treatment-related factors like credibility and expectation of success. Collectively, however, they all show inconsistencies in predicting treatment response (Button et al., 2012; Edmonds et al., 2018; El Alaoui et al., 2016; Hadjistavropoulos et al., 2016). The severity of baseline depression symptoms is a relatively strong predictor; several individual patient data meta-analyses found that more severe baseline depression indicated larger effects of iCBT (Karyotaki et al., 2017, 2018, 2021). Nevertheless, each

of these variables, even when statistically significant, holds only a small amount of explanatory power, making it difficult to translate them into clinical use (Chekroud et al., 2021; Marek et al., 2020).

This presents a unique opportunity for machine learning to be used to develop multivariable prediction models from large-scale, longitudinal data (Chekroud et al., 2021; Rost et al., 2023). Early successful machine learning models came from studies predicting antidepressant medication response. Chekroud and colleagues' model (2016) used 25 clinical and demographic variables to predict depression remission status, resulting in ~60% accuracy in hold-out data. They found baseline depression severity as well as depression item/subscales to carried important predictive signals. Iniesta and colleagues (2016; 2018) extended this line of work by showing that model predictions may be drug-specific, where an escitalopram-trained model predicted escitalopram response better than nortriptyline, and vice versa. They also demonstrated that combining multimodal data (i.e., genetics, socio-demographic, clinical features) can significantly improve predictive accuracy (i.e., ~70 accuracy in external validation).

Aside from a few exemplars, research to-date has suffered from significant methodological issues (Chekroud et al., 2021; Ermers et al., 2020; Lee et al., 2018; Sajjadian et al., 2021). Studies often retrospectively reanalysed clinical trial data with limited variables, did not assess out-of-sample estimates of model fit, and lacked large sample sizes necessary for accurate, robust predictions. The outlook is even less favourable when considering psychotherapy specifically; a recent scoping review revealed only 44 studies (mostly proof-of-concepts) have been carried out in the area, of which 14 had N>200 and only 3 externally validated their models (Aafjes-van Doorn et al., 2021). To get the large, multivariable data that we need for accurate machine learning predictions, one promising solution is to leverage online research methods. Digitising treatment prediction research can dramatically upscale and accelerate rich, longitudinal data collection, while increasing the depth and breadth of data being collected from diverse samples (Lee et al., 2023; Rutledge et al., 2019). iCBT and online research complement each other by enabling the remote, standardised collection of data at scale and at speed, which can be integrated efficiently. In this sense, both digitised intervention and research lend themselves well to the application of predictive models with machine learning (Chekroud et al., 2021; Lee et al., 2023).

The use of machine learning in predicting iCBT treatment response is rapidly emerging, albeit uncommon still. Some initial proof-of-concepts can be found predicting iCBT response in obsessive-compulsive disorder (N=61; Lenhard et al., 2018), social anxiety (N=26; Månsson et al., 2015), and body dysmorphia (N=88; Flygare et al., 2020). These studies reported high accuracies (75-92%) with no external validation, likely attributable to model overfitting during training procedures (Isacsson et al., 2023). In depression, Pearson and colleagues (2019) built an ensemble model with self-report data (N=283) that outperformed a benchmark linear regression model in predicting depression level post-iCBT (R^2=0.25%; +0.08% gain), and found iCBT engagement data (e.g., module usage, therapist access) to be important predictors. Their results suggest that incorporating change data during treatment may improve prediction performance. Their model, however, was not externally validated using hold-out data. Critically, a model relying on treatment data cannot be applied before treatment starts to guide and individualise treatment allocation. Hornstein and colleagues (2021) evaluated several machine learning models that used only pre-treatment variables to predict post-iCBT responder status (N=970). Their winning model achieved a modest 60% balanced accuracy when tested in hold-out data, likely because they only included a very limited range of self-report data. In this sense, the merit of incorporating multimodal data to enhance prediction performance is demonstrated by Wallert and colleagues (2022) who used demographics, clinical, process (e.g., time of day completing assessment), and genetic data in machine learning to predict remission following iCBT (N=894). The final model retained predictors belonging to all 4 types of data and yielded 65.6% accuracy in external validation. One major drawback of this study, however, is the absence of a treatment comparison group to assess the model's specificity, key to determining whether the predictions are unique to iCBT or generally applicable to other treatments. To our knowledge, while studies have previously tested for drug-specificity of their models (Iniesta et al., 2018), no research has directly assessed the specificity of iCBT prediction models to other depression treatments, such as antidepressant medication.

To address this gap, the Precision in Psychiatry (PIP) study used machine learning to predict early depression response for patients initiating iCBT using a wide range of self-report and cognitive data gathered at baseline. The study adopted a fully internet-based method to recruit and longitudinally assess patients as they progressed through iCBT. We further tracked a smaller cohort of patients initiating antidepressant medication to test the

treatment specificity of our model predictions. For both arms, early treatment response was assessed at week 4, which has been shown to be a reliable indicator of overall response (Belanger et al., 2023; Enrique et al., 2019; Nierenberg et al., 2000). The PIP study aims to predict early iCBT response with comparable performance to that of post-iCBT, all while offering the advances of being much more closely aligned with real-world implementation.

4.2 Methods

4.2.1 Study Design and Participants

The PIP study is a 4-week longitudinal observational study that aimed to use machine learning to identify baseline self-report and cognitive predictors of early treatment response to iCBT for depression. Using a fully internet-based protocol, we recruited and tracked patients initiating iCBT intervention (or antidepressant medication) for common mental health issues for the first 4 weeks of their treatment. A detailed overview of the study design and procedures, recruitment, sample characteristics, outcome assessments, interventions, clinical changes, treatment information, study feasibility and attrition, data quality, as well as ethical approval has been published elsewhere (Lee et al., 2023). Briefly, our primary study arm comprised patients initiating clinician-guided iCBT on the SilverCloud platform, digitally recruited from a mental health charity in Dublin, Ireland and from a National Health Service (NHS) Talking Therapies clinic in London, UK. SilverCloud provides low-intensity, evidence-based iCBT interventions for a broad range of mental health conditions, of which efficacy in improving clinical outcomes have been shown previously (Richards et al., 2020). A smaller, secondary study arm comprised patients initiating a variety of antidepressant medication, globally recruited using a combination of online and in-print advisements. Participants in both arms were included if they were between 18-70 years old, fluent in English, had access to a computer, were starting treatment within 2 days of study sign-up, and scored ≥10 on the Work and Social Adjustment Scale (Mundt et al., 2002), a transdiagnostic measure of functional impairment as a result of mental health. Eligible participants invited to take part in the study were sent an individualised study schedule and web-link for completing the baseline assessment, 3 consecutive weekly check-ins, and the final assessment (see **Figure 4.1A**).

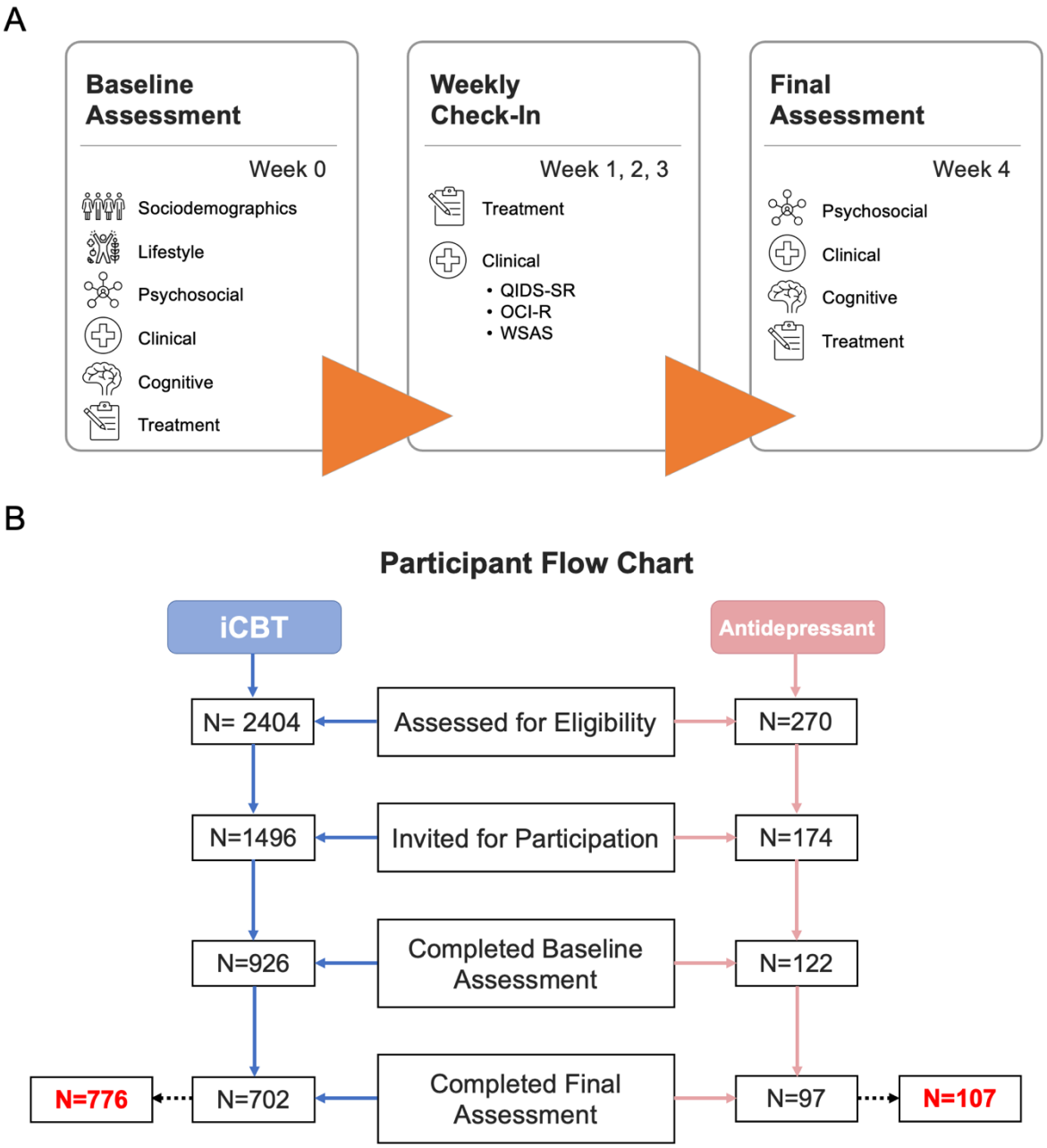


Figure 4.1 A) PIP study design overview. At week 0, eligible participants were invited to complete the baseline assessment comprising self-report questionnaires and gamified cognitive tasks online. Subsequently, for 3 consecutive weeks, they were invited to complete a weekly assessment on clinical symptoms and treatment adherence, before completing a final assessment at week 4, an abbreviated version of the baseline assessment. **B) Participant flow chart.** Participants were progressed onto the next stage of the study upon full completion of each assessment. Upon imputation of missing QIDS-SR score at the final assessment, the sample sizes included for analyses increased from N=799 to N=883.

Participant flow chart of the study is shown in **Figure 4.1B**. We adopted a complete cases approach, where only participants who completed all assessments were included.

Nevertheless, to maximise the sample sizes for training and validating our treatment prediction model, we imputed the score of our primary depression outcome, measured on the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR) (Rush et al., 2003), for those with missing data at study end-point. We predicted the missing final QIDS-SR score at week 4 in the training sample using QIDS-SR score at week 3 in a linear regression (b=0.88, SE=0.03, p<.001). We then applied the same formula for imputing missing final QIDS-SR score in the test samples. This increased our final sample size from N=799 (iCBT N=702, antidepressant N=97) to N=883 (iCBT N=776, antidepressant N=107).

The majority of participants from both arms resided in Ireland and the UK, and tended to be young, female, single, employed, and perceived themselves to be of middle social class status. In relation to clinical characteristics, while most participants reported having no family members with mental health history, they themselves reported currently having ≥1 psychiatric disorders (primarily depression and anxiety) with minimum 2 lifetime mental health episodes, and the most common age of onset being in their adolescence. Most participants have had prior experience with mental health treatments, and self-reported an average rating of treatment success expectation with their current intervention. Some participants reported taking concurrent treatments to their interventions; more so psychotherapy than antidepressant medication (see **Table 4.1** for patient characteristics of study completers in the iCBT and antidepressant arm).

Table 4.1 Baseline characteristics of study completers.

Sample Characteristics	**Study Arms**		**t / X^2 (df)**	**p**
	iCBT[a]	Antidepressant[b]		
Sex, *N (%)*			2.03 (2)	.363
Female	600 (77.5)	78 (72.9)		
Male	168 (21.7)	27 (25.2)		
Other	6 (0.8)	2 (1.9)		
Country, *N (%)*			124.09 (2)	<.001
UK	648 (83.7)	42 (39.2)		
Ireland	109 (14.1)	46 (43.0)		
Other	17 (2.2)	19 (17.8)		
Age, *Mean (SD)*	31.8 (11.0)	30.1 (10.4)	1.52 (877)	.128
Marital Status, *N (%)*			1.60 (5)	.902
Single	303 (39.1)	47 (43.9)		
In a Relationship	226 (29.2)	31 (29.0)		
Married	196 (25.3)	23 (21.5)		
Divorced	28 (3.6)	3 (2.8)		
Separated	19 (2.5)	3 (2.8)		
Widowed	2 (0.3)	0 (0)		
Education, *N (%)*			4.81 (2)	.090
<Third Level	181 (23.4)	15 (14.0)		

Some/Complete	412 (53.2)	63 (58.9)		
Third Level				
>Third Level	181 (23.4)	29 (27.1)		
Employment, *N (%)*			15.38 (2)	<.001
Employed	555 (71.7)	57 (53.3)		
Unemployed	213 (27.5)	48 (44.8)		
Retired	6 (0.8)	2 (1.9)		
Subjective Social Status (0-9), *Mean (SD)*	4.17 (1.68)	4.40 (1.98)	-1.30 (879)	.193
No. of Current Diagnoses, *N (%)*			34.11 (2)	<.001
None	263 (34.0)	8 (7.5)		
One	276 (35.6)	45 (42.0)		
>One	235 (30.4)	54 (50.5)		
Types of Diagnoses[c], *N (%)*			9.84 (5)	.080
None	263 (34.0)	8 (7.5)		
Depression	367 (47.4)	81 (75.7)		
GAD	300 (38.8)	58 (54.2)		
Panic Disorder	32 (4.1)	5 (4.7)		
PTSD	28 (3.6)	14 (13.1)		
OCD	36 (4.7)	4 (3.7)		
Others	55 (7.1)	13 (12.1)		
Family with Mental Disorders, *N (%)*			3.44 (3)	.329
None	320 (41.3)	37 (34.6)		
One	238 (30.8)	33 (30.8)		
Two	122 (15.8)	18 (16.8)		
≥Three	94 (12.1)	19 (17.8)		
No. of Lifetime Episodes, *N (%)*			7.54 (2)	.023
<2	81 (10.6)	8 (7.6)		
2-5	371 (48.7)	40 (37.7)		
>5	310 (40.7)	58 (54.7)		
Age of Onset, *N (%)*			10.04 (2)	.007
Childhood	132 (17.4)	27 (25.7)		
Adolescence	315 (41.6)	51 (48.6)		
Adulthood	311 (41.0)	27 (25.7)		
Current Episode Duration (Days)[d], *Median (Range)*	199 (0-19,490)	200 (2-15,904)	0.19 (797)	.847
History of Past Treatment, *N (%)*			10.11 (3)	.018
Never Before	353 (45.6)	34 (31.8)		
Psychotherapy & Medication	170 (22.0)	34 (31.8)		
Medication only	123 (15.9)	15 (14.0)		
Psychotherapy only	128 (16.5)	24 (22.4)		
Concurrent Treatment, *N (%)*				
Antidepressant Medication	62 (8.0)	9 (8.4)	0.02 (1)	.881
Psychotherapy	182 (23.5)	42 (39.3)	12.40 (1)	<.001
Treatment Expectation (0-9), *Mean (SD)*	5.06 (2.05)	4.77 (1.93)	1.40 (879)	.161

[a]Excluded missing data for descriptive analyses (N=2 for sex, country, marital status, education, employment, subjective social status, number of diagnoses, types of diagnoses, family with mental health disorders, treatment history, and treatment expectations, N=3 for age, N=14 for number of lifetime episodes, N=18 for age of onset, N=76 for current episode duration). [b]Excluded missing data for descriptive analyses (N=1 for age, N=1 for number of lifetime episodes, N=2 for age of onset, N=8 for current episode duration). [c]The total number of diagnoses type exceeds the sample size of completers (i.e., participants have the option to pick more than one diagnosis). [d]Median and range reported instead of mean and standard deviation due to extremely skewed data.

4.2.2 Outcome Measure

The study operationalised early treatment response as the change score measured on the QIDS-SR from the baseline assessment (week 0) to the final assessment (week 4). The QIDS-SR is a widely used 16-item self-report instrument designed to assess the severity of DSM-IV criterion symptoms for major depressive disorder. The measure has been shown to enjoy highly acceptable reliability and concurrent validity (Rush et al., 2003). Of the study completers, participants in the antidepressant arm presented with more severe baseline depression severity (M=16.38, SD=4.19) than those in the iCBT arm (M=13.72, SD=4.30), t(881)=-6.04, p<.001. Baseline depression severity was also similar between the original completer sample (N=799, M=14.02, SD=4.35) and those for which final QIDS-SR were imputed (N=84, M=14.25, SD=4.61), t(881)=0.47, p=.64. Participants in both the antidepressant arm (M=4.98, SD=5.11, t[96]=9.61, p<.001) and the iCBT arm (M=3.03, SD=4.17, t[701]=19.23, p<.001) improved in depression symptoms after 4 weeks of treatment.[5] This reduction was significantly larger in the antidepressant arm when compared to the iCBT arm, t(797)=-4.19, p<.001, but participants between-arms were not different on their QIDS-SR score at study end-point, t(797)=-1.56, p=.012.[1]

4.2.3 Predictors of Interest

At baseline, participants completed a comprehensive battery of self-report assessments and gamified cognitive tasks in randomised order. Baseline predictors of interest spanned six categories of data, including socio-demographics, psychosocial factors, lifestyle and general health factors, clinical data, treatment data, and cognitive data. To ensure model optimisation, all interval and continuous variables were standardised and nominal data binarised, with the exception of 'age of onset' which was dummy transformed. Variables with rare endorsement, defined as <10% in any response option for binary items and ≥90% in any response option for interval/continuous variables, were eliminated. Variables that were highly collinear (r≥0.8) and had near zero variance were also removed.

It is widely known optimal feature selection remains a challenge in the application of machine learning in treatment prediction. The predictive utility of features with different levels of information may depend on the type of algorithm used. The study therefore adopted an exploratory strategy to feature selection, testing several combinations of feature sets with varying levels of granularity based on:

[5]This analysis was carried out with the original sample of completers (iCBT N=702, antidepressant N=97) excluding those for which the final QIDS-SR scores were imputed.

(i) An item-level approach (357 predictors), where each item was entered as an individual predictor.

(ii) A subscale score approach (100 predictors), where psychometric subscale scores (total scores if not available) were entered as predictors.

(iii) A total score approach (78 predictors), where psychometric total scores were entered as predictors.

(iv) A filter-based approach employing the random forest variable importance calculation method when entering all items as predictors in the model, to circumvent human bias in feature selection.

Of note, both the total and subscale feature sets incorporated 9 symptom-specific QIDS-SR items whilst the item and filter-based feature sets incorporated all 16 items of the scale. A full list of predictor variables collected in the study is detailed in the online variable directory (see **Supplementary Materials 8.2.1**).

4.2.4 Model Derivation

We tested a combination of linear and non-linear machine learning models, including elastic net regression, random forest, eXtreme gradient boosting machine (XGBoost) in linear and tree-form to explore the optimal model fit with regards to diversity in model complexity, interpretability, assumptions, handling of dimensionality, and detection of higher-order effects and non-linear data patterns.

Elastic net regression is a parsimony-oriented model that employs regularisation hyperparameters (lasso and ridge) for penalising regression weights to address feature multicollinearity, by combining the predictive utility of a high number of variables and in turn prevent overfitting (Friedman et al., 2010). By relying on linear functions, elastic net regression allows for improved interpretability of the predictor-outcome relationship. Both the random forest and XGBoost tree-form are ensemble learning methods, combining multiple weak learners (i.e., random subsamples of decision trees of predictors) to model complex, non-linear relationships between the predictors and outcome, which may compromise interpretability. While random forest algorithm constructs independent decision trees of random selected predictors before combining and averaging their predictions which is an approach resilient to overfitting (Breiman, 2001), XGBoost tree trains its decision trees in a sequential manner (each new tree aims at fitting the residual error of the previous one), which may lead to better predictive performance.

XGBoost further allows the incorporation of linear models as base learners in addition to decision trees (XGBoost linear) with regularisation, which may further improve interpretability and generalisation of models. By comparing a range of diverse algorithms with varying approaches in the selecting and weighting of variables, it would allow us to identify the optimal predictive model.

4.2.5 Model Training and Cross-Validation

Figure 4.2 demonstrates the processing and analysis pipeline for the study. Each model was trained on 70% of the iCBT completer dataset (N=543). To internally evaluate and compare performance in predicting depression symptomatology improvement, we carried out a 10-fold nested repeated cross validation (nrCV) for each model, where the outer and inner cross-validation folds were stratified based on our outcome measure of QIDS-SR change. nrCV is a resampling technique that separates the processes of hyperparameter optimisation and model selection (inner folds) and model performance evaluation (outer folds) to minimise overfitting and optimistic bias in performance estimation (Lewis et al., 2023). Repeated iterations were implemented to reduce the impact of random variability and to stabilise performance estimates.

In brief, with each of the 10 outer folds, the data are partitioned into outer training sets (90%) and outer validation sets (10%). Within each outer fold, a 10-fold repeated cross-validation with 10 iterations takes place in the outer training set, i.e., an inner fold. Within each inner fold, a model is trained and tuned using various hyperparameter configurations using the inner training set (90%). The best model configuration is selected based on the lowest root mean square error (RMSE) on the inner validation sets (10%) in the inner folds. This model is then brought to fit on the whole outer training set and evaluated on the outer validation set of the outer fold. With 10 outer folds, this process is repeated 10 times, generating 10 models. The performances across these 10 models are averaged together to give a measure of a final RMSE value to use for comparison across models (Lewis et al., 2023). Lastly and independently, one last round of 10-fold cross-validation is performed on the whole training dataset to determine the optimal hyperparameters to fit the final model, used for prediction with external data.

nrCV was applied to each machine learning model with each feature set separately. For the filter-based approach, a random forest-based filter was applied to all items entered as predictors in the model, which ranks each of their importance by computing the average

decrease in prediction accuracy when a particular variable is randomly permuted with others remaining unchanged. The higher the decrease in prediction accuracy as a result, the more important the predictor (Breiman, 2001). This filter was embedded within the outer training folds of the nrCV to prevent leakage of information into the outer validation folds. Furthermore, for each machine learning algorithm type, we estimated a benchmark model comprising only baseline depression severity, age, and sex as predictors. This allows us to quantify the improvement of prediction from additional predictors by comparing their performance to a minimal set of variables. Altogether, 20 machine learning models were trained and compared based on performance estimates resulting from the nrCV analyses. For our continuous outcome measure, we used the coefficient of determination R^2, which quantifies the proportion of variance in QIDS-SR change explained by the model, averaged across the predictions on the outer validation folds, to determine the model's expected predictive utility in unseen data . We considered a benchmark R^2 of ≥6.3% to be clinically meaningful, in line with a pre-established consensus on clinical significance (Uher et al., 2012).

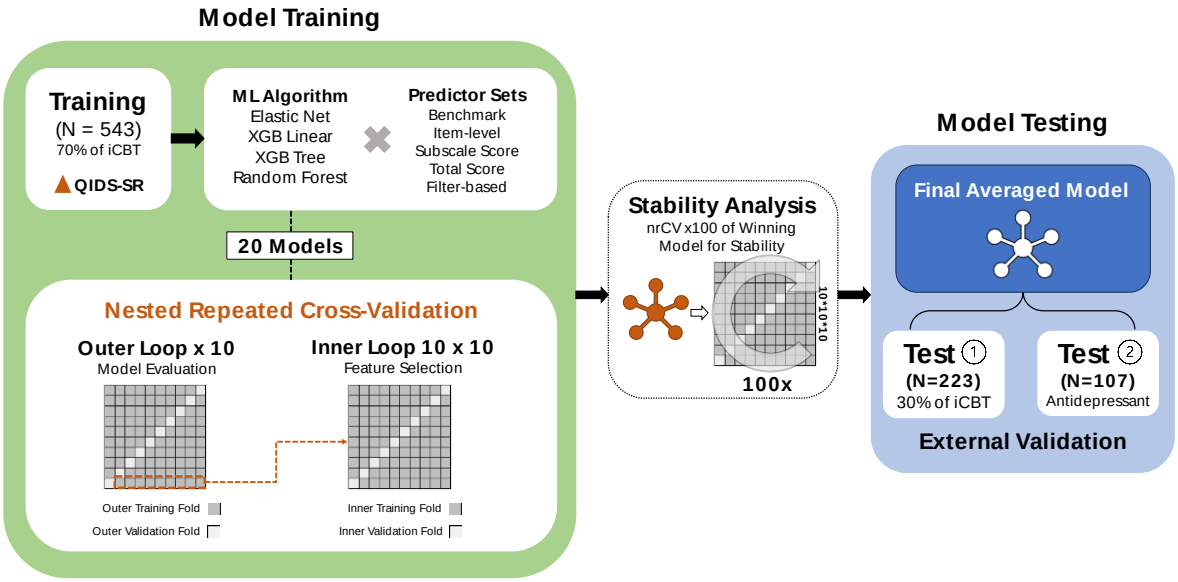


Figure 4.2 Predictive modelling pipeline. We used 70% of the iCBT sample (N=543) to train machine learning models that predict depression improvement indexed by the QIDS-SR during the first 4 weeks of treatment. We tested 4 machine learning algorithms, each applied to 5 different sets of features with varying level of granularity. We applied nested repeated cross-validation (nrCV) to each model, involving 10 outer loops for model evaluation and 10 inner loops (with 10 iterations each) for hyperparameter optimisation and feature selection. We determined the best model explaining the most variance in the outcome, and confirmed its performance via 100 runs of the nrCV. Finally, we averaged the feature coefficients over the 100 runs and tested the averaged model's generalisability on

2 separate hold-out data sets: 30% of the iCBT sample (N=223) and the antidepressant medication sample (N=107).

4.2.6 Model Testing and External Validation

Permutation-based significance testing. The final model was statistically compared to the benchmark model of the same algorithm type using permutation-based testing to quantify any additional variance explained by the predictors beyond basic variables of baseline severity, age, and sex. We focused on reporting the results of the final model here due to the computationally intensive nature of this analysis. To test this, we assumed predictions from the benchmark and final model were equal, and generated a null distribution of R^2 differences between the two models after creating 1000 random pairs of predictions, where the individual predictions are randomly taken from either the benchmark or the final model. We then used this null distribution to derive the p-value for true R^2 difference estimated from both models. Cross-validation (and not nested cross-validation) was carried out for this analysis as hyperparameter optimisation was not required.

Stability analyses. As machine learning algorithms may be sensitive to random data partitioning between training and validation sets during cross-validation, to ensure the robustness of our results, we repeated the nrCV analyses 100 times for the final model and brought forward the averaged coefficients of retained features for external validation.

External validation. To test the final averaged model's generalisability to unseen data, we applied it to two hold-out datasets to predict their QIDS-SR change scores and measured the model's performance in R^2 and RMSE. The first test dataset comprised the remaining 30% of the iCBT completer sample (N=233). The 70:30 random split of the iCBT sample for model training and testing was stratified in a way so that recruitment site (Ireland/UK), age, sex, baseline and final QIDS-SR scores were balanced between the two. The second test dataset comprised the antidepressant sample (N=107) to test the treatment-specificity of our model prediction. Crucially, the test datasets were not involved at any stage of data preparation, model development, and internal model validation.

4.2.7 Predictor Importance

To understand the underlying factors driving the predictions of the final model, we determined the variable importance of the retained features of the final model in a model-agnostic approach using Shapley additive explanation (SHAP) values (Lundberg & Lee,

2017). SHAP values calculates the average marginal contribution of each predictor to the model performance across all possible combinations of predictors.

4.2.8 Missing Data

For missing data, we imputed the median for continuous/ordinal variables and the mode for binary variables. The rates of missingness for all predictors were between 0-5%, with the exception of 'current episode duration' (9%). To avoid leakage of information into hold-out datasets at this step, we imputed the median/mode values derived from the training dataset and applied them to missing data in the test sets.

All data processing and analyses were carried out using R (version 4.2.2). We made extensive use of R packages including *caret* (Kuhn, 2008), *glmnet* (Friedman et al., 2010), *xgboost* (Chen & Guestrin, 2016*)* for predictive modelling, and *fastshap* for calculating SHAP (Greenwell & Boehmke, 2020). Statistical significance was determined at $p<.05$.

4.3 Results

4.3.1 Cross-Validation Performance and Final Model Derivation

Table 4.2 shows the performance measures of all 20 models assessed during nrCV analyses. All model algorithms, regardless of the feature set employed, predicted substantially beyond the clinical threshold (i.e., R^2=6.3%), with the exception of the XGBoost linear model employing the item score feature set (R^2=4%) and filter-based feature set (R^2=5.2%). Comparisons of performance measures revealed the best performing model as the elastic net regression employing the total score feature set, which was associated with the lowest prediction error (RMSE=.924, MAE=.736) and the highest proportion of variance explained (R^2=14.6%) (see **Figure 4.3A**). Our non-parametric permutation significance test further showed this model outperformed the elastic net regression benchmark model (R^2=9%, RMSE=.960, MAE=.754) significantly by 5.6% ($p<.001$) (see **Figure 4.3B**). Stability analyses repeating nrCV 100 times revealed, on average, the elastic net regression with total score features predicted 14% variance in change in QIDS-SR (mean R^2=14%, SD=.008, 95% CI [.138, .142]) with the lowest RMSE (mean RMSE=.927, SD=.004, 95% CI [.926, .927]).

Table 4.2 Model performance results from nested cross-validation analyses predicting change in QIDS-SR score in the training sample (N=543).

Type of Algorithm	Feature Set	Performance Metrics		
		R^2	RMSE	MAE
Elastic Net Regression	Benchmark	.090	.953	.752
	Total Scores	**.146**	**.924**	**.736**
	Subscale Scores	.113	.941	.749
	Item Scores	.125	.935	.747
	Filter-Based	.121	.938	.748
Random Forest	Benchmark	.079	.960	.754
	Total Scores	**.142**	**.929**	**.732**
	Subscale Scores	.138	.931	.736
	Item Scores	.092	.954	.754
	Filter-Based	.086	.957	.756
XGB Linear	Benchmark	.077	.961	.758
	Total Scores	**.119**	**.938**	**.743**
	Subscale Scores	.089	.957	.756
	Item Scores	.040	.984	.781
	Filter-Based	.052	.975	.771
XGB Tree	Benchmark	.084	.957	.754
	Total Scores	.112	.944	.736
	Subscale Scores	**.116**	**.940**	**.736**
	Item Scores	.083	.957	.754
	Filter-Based	.089	.954	.756

Note: R^2=R-squared; RMSE=Root Mean Square Error; MAE=Mean Absolute Error

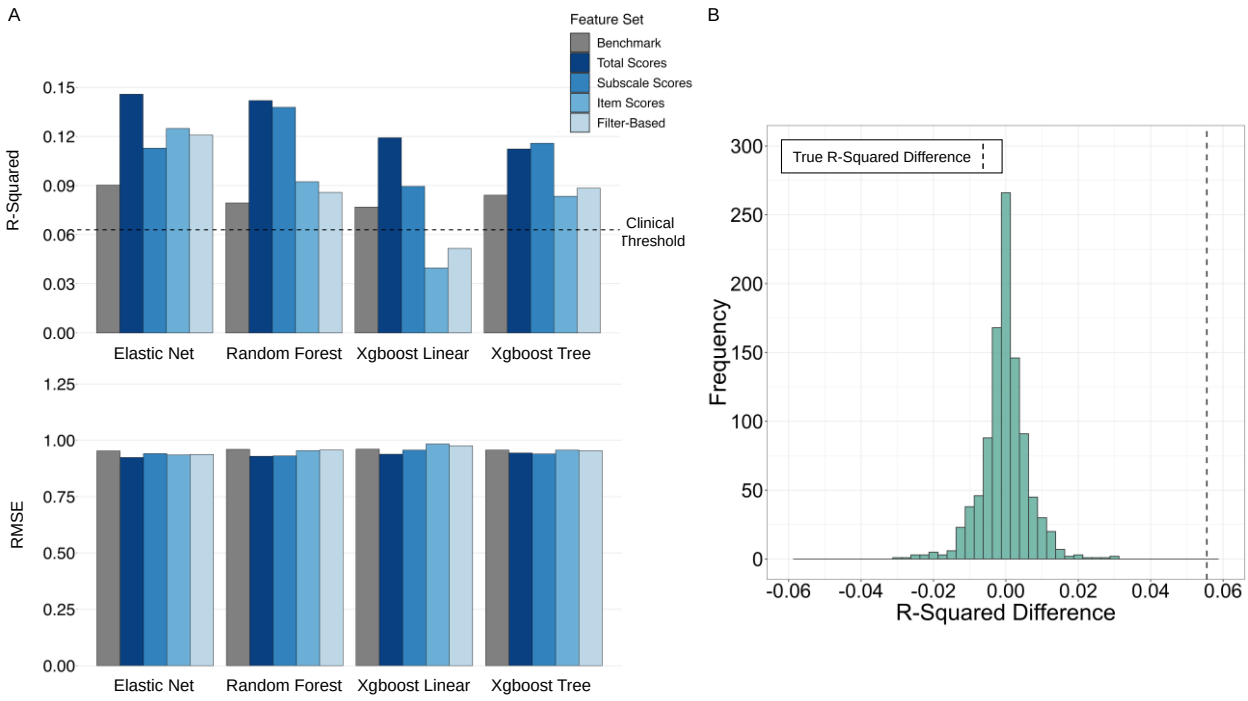


Figure 4.3 Model performance from cross-validation. **A)** R^2 and RMSE values obtained from nested cross-validation analyses for elastic net regression, random forest, XGBoost linear and tree models employing the benchmark, total scores, subscale scores, item scores, and filter-based feature sets,

respectively. **B)** Null distribution of 1000 permutated R^2 differences between the elastic net regression with the benchmark and total score feature sets, with the true R^2 difference score residing outside of the distribution.

4.3.2 External Validation

Bringing forward the final model from the stability analyses to external validation, we found that the model explained more variance in QIDS-SR change from the baseline to final assessment than in the cross-validation when applied to the iCBT hold-out dataset (R^2=18.5%, RMSE=.878); this prediction was not treatment-specific however, with the model explaining also more outcome variance in the antidepressant group than in cross-validation (R2=17.7%, RMSE=1.102) (see **Figure 4.4A and 4.4B**).

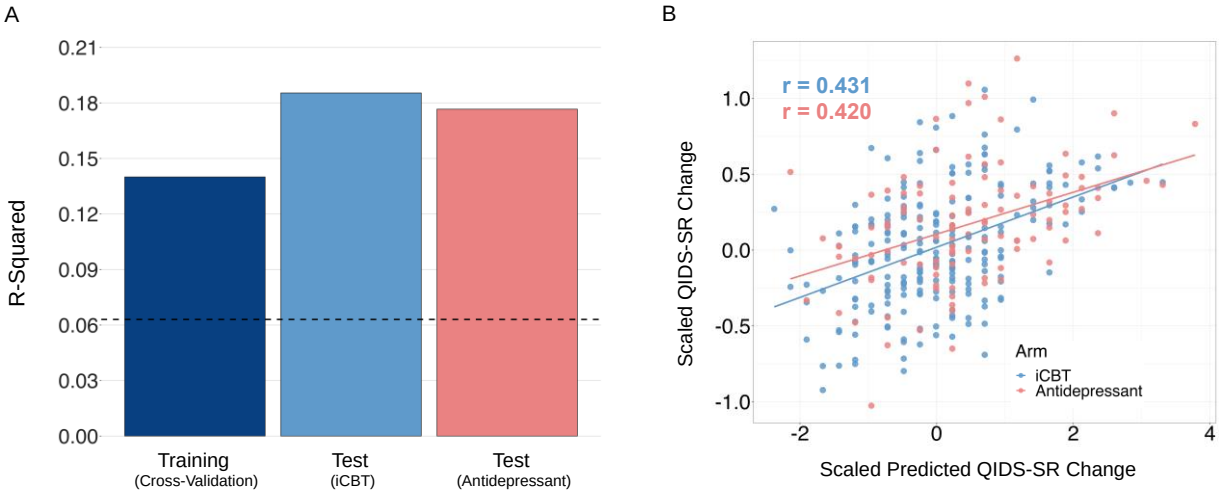


Figure 4.4 Final averaged model performance during external validation. **A)** Model performance of 1) the final averaged model across 100 runs of nrCV (R^2=14%) and during external validation with the iCBT hold-out dataset (R^2=18.5%) and the antidepressant hold-out dataset (R^2=17.7%). **B)** Pearson correlation between the actual values of QIDS-SR change scores (scaled) and the predicted QIDS-SR change scores (scaled) in both the iCBT hold-out dataset and the antidepressant hold-out dataset, respectively.

4.3.3 Predictor Importance

Of the 78 baseline predictors included in the final elastic net regression model, out of the 100 repeats of the nrCV, the same 27 variables were retained every time (alpha=.737, lambda=.053), with 4 extra variables retained 61 times (alpha=.684, lambda=.053) (see **eTable 8.3.1.1** in **Supplementary Materials 8.3.1**). **Figure 4.5** demonstrates the averaged SHAP values of the retained 31 variables in the final model, the averaged

coefficients of which are shown in **eTable 8.3.1.2**, **Supplementary Materials 8.3.1**. The top 2 performing predictors were baseline depression severity measured on both the QIDS-SR and the Zung Self-Rating Depression (SDS) scale, albeit in opposing direction (greater QIDS-SR and decreased SDS total scores at baseline lead to greater change in QIDS-SR). Expectation of treatment success as a treatment-related variable featured highly following baseline depression severity, where higher ratings are associated with greater improvement in depression symptomatology. The rest of the predictors primarily comprised demographics (e.g., sex), physical health and lifestyle (e.g., exercise, diet), environmental factors (e.g., stressful life events, social support), depression-specific symptoms (e.g., low mood, lack of interest), as well as transdiagnostic clinical characteristics (e.g., apathy, impulsivity, functional impairment), The model further retained several cognitive variables (e.g., metacognition efficiency, model-based and model-free index, volatility learning), but all were featured low in comparison to the rest of the predictors.

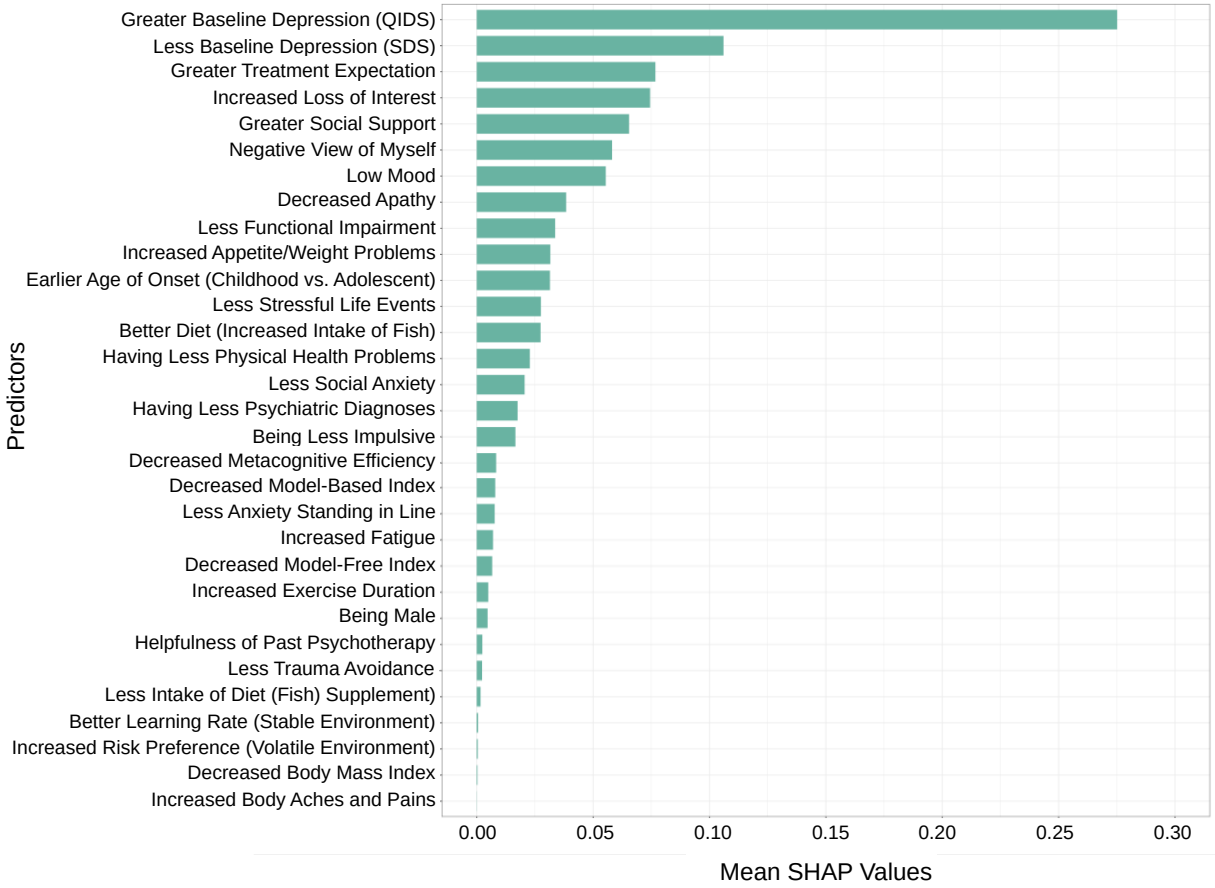


Figure 4.5 Ranked mean SHAP values of 31 retained predictors in the final model, averaged over 100 iterations of the model. Greater the mean SHAP values, the more important the predictor.

4.4 Discussion

Improving the rates of treatment response in depression has not proved to be an easy feat, yet pursuits using machine learning within the realm of precision medicine have shown promise. The aim is to reliably estimate the likelihood of treatment response for each individual from the get-go, which in turn may guide decisions surrounding treatment allocation (Rost et al., 2023). To this end, the PIP study integrated digitised research and intervention to satisfy the 'big data' requirement of machine learning (Lee et al., 2023). While other prediction studies have employed machine learning, our study was the first to predict early treatment response in depression for patients undertaking iCBT in a observational setting, and test the specificity of model predictions against another type of treatment (i.e., antidepressant medication). Our findings may have implications for timely adjustment of treatment based on early response and potentially improve the overall response rate of iCBT (Sajjadian et al., 2021).

Using a comprehensive battery of baseline self-report and cognitive data all gathered remotely online before treatment initiation, we tested and compared a diverse array of linear and non-linear machine learning algorithms with varying levels of feature granularity, as it was important to explore the relationship between the nature of the data and algorithm type. Our findings from the internal nested cross-validation analyses revealed that all models performed relatively similar, well exceeding the pre-established benchmark for clinical significance (i.e., $R^2 \geq 6.3\%$; Uher et al., 2012), with the exception of the XGBoost linear model employing the item score and filter-based feature sets. Apart from the XGBoost tree, each model algorithm performed the best respectively when the total score feature set was incorporated. While one may assume a richer dataset with increased number and granularity of predictors would result in more accurate predictions, this increases the risk of topological overlap which may in turn translate to an elevated level of measurement noise, thus yielding suboptimal performance instead. Of all 20 models tested during nested cross-validation, the best performing model was the elastic net regression employing the total score feature set, explaining 14.6% variance in early depression symptomatology change following 4 weeks of iCBT treatment. This dropped slightly to 14% when evaluated 100 times via stability analyses, but overall the model still had the lowest prediction error amongst the rest. The superiority of elastic net regression relative to other machine learning algorithms has been previously highlighted (Webb et al., 2020). As a variant of linear regression models with regularisation, it is

robust against overfitting while more easily interpretable at the same time, making it the ideal candidate algorithm for generalising to and predicting outcomes in new unseen data (Delgadillo et al., 2017).

Case in point, when we brought the final model averaged from the stability analyses forward to external validation in the iCBT hold-out dataset, it explained a greater proportion of variance (R^2=18.5%) in depression symptomatology change than in nested cross-validation. Our final model, using just 31 self-report and cognitive predictors at baseline to predict 4 weeks of iCBT response, demonstrated comparable/better performance to similar studies that predicted response after a full course of treatment or at longer follow-up timepoints. These included populations seeking treatments in antidepressant medication (Iniesta et al., 2016), psychotherapy (Buckman et al., 2021) and digital interventions (Hornstein et al., 2021; Jacobson & Nemesure, 2021). In cases where our performance was slightly inferior in comparison, studies either incorporated change data during treatment (Pearson et al., 2019) or other multimodal data that may be otherwise more inaccessible (i.e., genetics data; Wallert et al., 2022). Our findings successfully demonstrated the potential applicability of machine learning, using pre-treatment information easily obtained online, for clinicians to estimate from the get-go, the extent a patient may benefit from iCBT in their first few weeks of treatment. The PIP study represents a significant stride towards incorporating algorithmic tools to inform clinical practice in early stages of treatment. However, improvement in the prediction performance is further warranted before any implementation of the tool can take place.

Of note, not only did the final model generalise well to the iCBT hold-out dataset, it did also to the antidepressant medication hold-out dataset (R^2=17.7%), suggesting that the model's predictions were not treatment-specific. Further, we cannot ensure that our findings do not reflect (i) susceptibility to placebo effects or (ii) the naturalistic course of depression over 4 weeks. To our understanding, no prior research has investigated the ability of machine learning models to differentiate treatment response between different intervention types (e.g., medication vs. psychotherapy). Here, we speculate several reasons for our finding. First, the ability for models to differentiate between treatment class and/or type may depend on the type of predictors incorporated. In antidepressants, it has been found that models trained with genetic and clinical data (Iniesta et al., 2018) and EEG markers (Tozzi et al., 2020; Williams et al., 2015) may be class-specific, but not resting-state EEG (Rajpurkar et al., 2020) or neuroimaging markers (Goldstein-Piekarski

et al., 2018). In our study, we did not consider predictors that may be most relevant for iCBT (e.g., computer fluency, engagement) but a wide range of self-report such as socio-demographics and transdiagnostic clinical factors previously associated with depression treatment response (e.g., psychiatric comorbidity and pre-treatment disability; Kessler et al., 2017; Smits et al., 2012), which may explain why the model with these general, robust predictors also generalised well to antidepressant sample. In addition, the observational nature of the PIP study provides a lack of control in sampling due to non-random treatment assignment based on routine clinical decisions (Lee et al., 2023; Rost et al., 2023). This may lead to increased heterogeneity in the data and in turn model predictions (Webb et al., 2020). Unlike in clinical trials, where a systematic comparison of different treatment arms often result in treatment-specific predictions that are not readily generalisable to other treatments (Rost et al., 2023). Related to this is the fact that some participants in our study were undertaking concurrent treatments simultaneously (e.g., iCBT with antidepressant medication and/or traditional psychotherapy), which limits our ability to restrict our inference to iCBT treatment response (Lee et al., 2023). Those in the antidepressant sample also reported significantly higher baseline depression severity when compared to the iCBT sample, the increased variance of which may infer more signal for the model to make predictions on. Future studies should adopt a randomised controlled trial design with consideration to treatment-specific predictors to determine differential treatment response.

One major challenge facing precision in psychiatry is the lack of transparency and explainability in the underlying mechanisms of predictions made by machine learning (Rost et al., 2023). Findings of this study helped elucidate this by identifying baseline patient characteristics that contribute to treatment response in early stages of iCBT. Perhaps not surprisingly, baseline QIDS-SR depression severity was the strongest predictor of depression symptomatology improvement, as established robustly in a host of machine learning prediction studies (Chekroud et al., 2016; Li et al., 2023a; Pearson et al., 2019; Wallert et al., 2022; Webb et al., 2020), and also in studies investigating specifically iCBT treatment response (Button et al., 2012; Edmonds et al., 2018; Karyotaki et al., 2017, 2018, 2021; Hadjistavropoulos et al., 2016). As expected, higher baseline QIDS-SR scores predicted greater reductions in QIDS-SR at week 4. This response definition, which relies on change, is a composite measure heavily influenced by the baseline score. Therefore, both measures are susceptible to mathematical coupling

(i.e., a variable correlates with a change score that includes that same variable). While it may be the case that baseline score can associate with its change score (or a follow up score) in ways that reflect important recovery mechanisms, the confounding effects arising from their mathematical coupling and the potential for ceiling effects need to be considered (Browne et al., 2010; Oldham, 1962; Terluin, 2012). What we didn't expect, however, is the paradoxical finding that baseline depression measured on the SDS predicted depression change in the opposite direction. A previously study by Scodari and colleagues (2023) reported a similar finding, attributing it to measurement variances among the two different scales of depression. This is likely also in our case, where the QIDS-SR scale assess a range of depression symptoms pertaining to a DSM-IV diagnosis of Major Depressive Disorder (Rush et al., 2003), while the SDS scale taps into the pervasive and disturbing effect of depression, focussing more on the physiological and psychomotor domains (Zung, 1965), the latter for which iCBT may not be as effective as a treatment that seeks to correct unhelpful cognitive thinking and behaviours (Hofmann, 2012).

Importantly, we found that the final model explained early treatment response significantly more than a benchmark model with just baseline depression severity, age, and sex, highlighting the added predictive utility of other predictors included in the study. For instance, high expectations of treatment success ranked the highest following baseline depression severity in predicting early iCBT response. As a common factor for treatment response in psychotherapy (Flygare et al., 2020; Pearson et al., 2019; Webb et al., 2020), it has been suggested to substantially contribute to therapeutic gains regardless of the specific psychotherapeutic approach used (Mogoașe et al., 2017; Wampold et al., 2002). This suggests that early success in iCBT for depression strongly relies on the patient's subjective belief in whether the intervention will work for them, highlighting the implication for clinicians to encourage buy-in of the intervention prior to treatment assignment. In line with the literature, specific symptoms such as loss of interest, negative view of self, low mood, appetite/weight problems, anxiety waiting in line, trauma avoidance, and fatigue largely drove the prediction (Hornstein et al., 2021; Koutsouleris et al., 2016; Wallert et al., 2022; Webb et al., 2020), demonstrating the merit for considering individual symptom profiles in predictive efforts (Fried & Nesse, 2015). In addition, a range of transdiagnostic clinical variables spanning different physical and psychiatric conditions and chronicity was also featured (i.e., apathy, functional

impairment, physical and psychiatric co-morbidities, age of onset of mental health episodes, social anxiety, impulsivity, body aches and pains), which is largely consistent with literature highlighting comorbid psychopathology and physical impairment as important predictors of symptom change in depression (Kessler et al., 2017). While socio-demographic variables are typically used in traditional approaches to investigate depression treatment response (Rost et al., 2023; Vieira et al., 2022), only sex (i.e., being male) predicted early iCBT response in our study. Rather, the model featured some lifestyle and physical health variables not so commonly considered in similar studies, such as exercise, diet, and body mass index. Interestingly, we see cognitive features related to volatility learning, metacognition, model-based and mode-free planning also contributing to the model's prediction. In contrast to prior work, educational attainment was not retained in the final model, and one possibility is that cognitive variables accounted for the variance educational attainment may explain. Albeit ranking lower than most self-report variables in the model, this provides merit for the multimodal approach to predicting treatment response as previously endorsed (Chekroud et al., 2021b; Iniesta et al., 2018; Rost et al., 2023; Wallert et al., 2022). As we continue to progress prediction efforts in the field, it may be necessary to incorporate self-report and cognitive factors addressed during treatment to construct a more relevant psychological patient profile (Vieira et al., 2022).

Our results should be considered in light of certain limitations. The observational, uncontrolled nature of the PIP study implies that we cannot assume the model prediction is causally specific to early stages of iCBT, especially since some of our participants were undertaking concurrent treatments as iCBT. While this may be typical in a naturalistic, observational setting (Webb et al., 2020), it means clinicians may not use this model to recommend iCBT vs. other treatments for patients to try out. Future randomised controlled trials should seek to develop machine learning algorithms for differentiating response to various treatments, so that they can be of use to determine the optimal interventions for each individual (Gillan & Whelan, 2017). Furthermore, few issues persist with regards to the study's outcome measure of early depression change. Firstly, the measure was solely dependent on subjective patient ratings, not corroborated by formal/external post-treatment assessments. The outcome is also not completely independent of its predictors (e.g., baseline depression), which may induce measurement and incorporation bias (Rutjes et al., 2006; Schmidt & Factor, 2013). One could argue in

favour for a binarised and clinically meaningful outcome measure irrespective of predictor bias, such as remission/non-remission status, but such would be unrealistic to consider and achieve during early stages of treatment. While the PIP study endeavoured to collect an extensive range of pre-treatment predictors (i.e., 357 individual items), it was still missing some important variables (e.g., race, income) or variables of different modalities (e.g., genetics or neuroimaging markers) that may capture some unexplained variance in the outcome. Nevertheless, the PIP study considered data types such as self-report and gamified cognitive factors that could be seamlessly administered and assessed in an internet-based research method, which has the potential to scale up future treatment prediction research (Lee et al., 2023). In this regard, future studies should consider incorporating smartphone-based and/or passive sensing data (e.g., app usage, heart rate variability, geographic location) that can be easily integrated into digitised research and intervention. These novel objective measures offer a less resource-intensive and burdensome alternative over other highly dimensional but costly data (e.g., genetics and neuroimaging), potentially unveiling previously undetected patterns that might improve predictive accuracy (Chekroud et al., 2021; Rost et al., 2023). Lastly, if the translation and implementation of algorithmic-driven tools such as our model of early iCBT treatment changes are ever going to take place in real-life clinics, there is a need to prove their added predictive utility over and above the expertise of clinicians, with high acceptability and feasibility. Future studies should seek to explicitly test the accuracy of machine learning prediction models against those generated by clinicians themselves.

4.5 Conclusion

In conclusion, the application of machine learning in digitised research and intervention represents a promising frontier in improving precision in psychiatry. Our findings provided insights into a multitude of baseline patient characteristics spanning self-report and cognitive data that are predictive of iCBT treatment effectiveness at an early stage, with modest and comparable accuracy to similar studies investigating a full course of mental health treatment or at longer follow-ups. The PIP study adds to the growing body of research utilising algorithmic approaches to enhance personalised treatment in depression, highlighting the potential machine learning has to predict and optimise early treatment outcomes for individuals with depressive symptomology.

Chapter 5 – Estimating the Prognostic Value of Cross-Sectional Network Connectivity for Treatment Response in Depression

5.1 Introduction

Depression is one of the most prevalent and debilitating mental illnesses worldwide (World Health Organization, 2023). Despite the availability of effective treatments, a considerable proportion of individuals with depression still fail to achieve an adequate and sustained improvement (Cuijpers et al., 2021a; Rush et al., 2006; Trivedi et al., 2006). Unfortunately, there are few reliable and robust characteristics that distinguish those who respond and do not respond to treatment with most efforts to-date focused on standardised clinical measures and demographics (Maj et al., 2020; McMahon, 2014; Rost et al., 2023). Many have suggested this is a consequence of the way we conceptualise depression as a latent phenomenon that causes observed symptoms like sadness and anhedonia, which we typically sum to produce an overall depression score. Network theory of psychopathology forwards a different perspective and posits that symptoms are interacting components of a dynamical system (Borsboom, 2017; Borsboom & Cramer, 2013), which can result in positive feedback loops that propel people into episodes of illness. The greater the connectivity of these symptom networks, the lower the psychological resilience one has, with more connected networks reacting more strongly to perturbations and taking longer to recover. A key prediction of network theory emerging from this is that individuals with tightly connected networks should have greater vulnerability to depression, poorer prognosis, and more treatment resistance (Cramer et al., 2016; Pe et al., 2015; van Borkulo et al., 2015).

Several studies tested this using cross-sectional network analysis. van Borkulo and colleagues compared baseline connectivity differences between persisters (n=253) and remitters (n=262) of depression after two years (van Borkulo et al., 2015). In line with network theory, persisters had tighter network connectivity compared to remitters at baseline. This was replicated in a child/adolescent sample (n=566/174) (McElroy et al.,, 2019), but there have also been null findings, for example, in adolescents (n=232/233) (Schweren et al., 2018), and when depression and anxiety symptoms were examined together (n=956/1466) (O'Driscoll et al., 2021). On a more granular level, some studies have shown that the severity of symptoms that are more 'central' (i.e. important) is associated with non-response (Elliott et al., 2020; Hagan et al., 2021), and that

improvements in central symptoms predict changes in other symptoms (Papini et al., 2020; Robinaugh et al., 2016; Rodebaugh et al., 2018). Findings regarding the centrality hypothesis, however, are not univocal (O'Driscoll et al., 2021; Spiller et al., 2020), and it remains unclear whether centrality measures perform better than other network/non-network metrics when compared directly. Finally, contrary to network theory, a host of studies have reported that connectivity increases (rather than decreases) after treatment (Beard et al., 2016; Berlim et al., 2021; Blanco et al., 2020; Bos et al., 2018; Curtiss et al., 2021).

One of the common critiques of the network literature is the over-reliance on cross-sectional data and methods; estimating correlations between symptoms across-subjects rather than within-subject (Contreras et al., 2019; Robinaugh et al., 2020), and often employing small samples (Schumacher et al., 2022). This introduces two issues. First, it is uncertain if cross-sectional relationships between symptoms correspond to intraindividual relationships (Epskamp & Fried, 2018; Fisher et al., 2018). Second, cross-sectional studies typically construct just two networks of differential treatment response for comparison. This precludes controlling for potential confounds such as symptom severity and variance. Variance is of particular interest as it relates to the strength of the association that can be observed between symptoms. Cross-sectional networks are typically estimated from the partial correlations between symptom-pairs (Fried et al., 2016), and the correlation between any two nodes is their covariance proportional to their total variance. This leaves network estimation susceptible to change in variance, which can be introduced artificially when creating sub-groups of participants (Bos & De Jonge, 2014; Fried et al., 2016; Terluin et al., 2016). Prior research has shown that connectivity differences remain when groups are matched on baseline severity (McElroy et al., 2019; van Borkulo et al., 2015), but to our knowledge, none have assessed the impact of variance.

Our study sought to fill this gap by examining baseline network differences in N=40,518 patients who received internet-delivered cognitive behavioural therapy (iCBT) for depression. Leveraging our large sample, we adopted a novel subsampling approach so to conduct parametric analyses for 160 independent responder and non-responder networks with n=250 unique patients per subsample. Importantly, these subsamples naturally varied in levels of baseline network connectivity, symptom severity, and variance. This allowed us to assess if differences in cross-sectional network connectivity are better

explained by differences in depression severity and/or variance, which have not been separable using standard methods comparing single dyads of responder-non-responder networks. Additionally, using the independent networks from the subsampling method, we assessed whether other network metrics such as symptom centrality related to treatment success and contextualised their effect sizes against simpler metrics such as mean and variance of individual symptoms. Finally, findings were tested for generalisation to partially overlapping samples receiving iCBT of a longer duration (8-12 weeks) and to networks constructed from anxiety symptoms in patients receiving anxiety-relevant iCBT.

5.2 Methods

5.2.1 Study Setting and Intervention

We examined an archival dataset of patients who received iCBT from SilverCloud Health between January 2015 to December 2020, as part of the Improving Access to Psychological Therapies program within the National Health Service in England. The intervention followed NICE guidelines and have shown efficacy in improving clinical outcomes with sustained effects (Palacios et al., 2022; Richards et al., 2020). Patients provided their consent for their anonymized data to be used in routine service evaluations.

5.2.2 Outcome Measure

Depression was measured by the Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001). PHQ-9 was administered to patients at the beginning of each iCBT session, but patients were able to skip and return to these assessments later.

5.2.3 Study Sample

Figure 5.1A illustrates the process from which we derived our final study sample. First, patients were excluded if they did not have at least one PHQ-9 completed in a timeframe of 4-8 weeks since treatment initiation. We included patients completing relatively short durations of treatment (i.e., 4 weeks) due to the self-paced nature of iCBT (Lawler et al., 2021). The last PHQ-9 completed within the 4-8 week window was deemed as the follow-up assessment. As the study focused on examining the association between depression network characteristics and clinical changes following treatment for depression, patients were further excluded if they scored <10 on the PHQ-9 (i.e., did not reach ‘caseness’ for depression) at baseline, and if they were enrolled in any other type of

iCBT program not purposed for treating depression. Most patients were clinician-guided, meaning treatment progress was monitored and facilitated by a clinician. As prior studies have shown differential efficacy of iCBT when guided vs. unguided (Karyotaki et al., 2021), we excluded data from patients who were unguided. Furthermore, patients who satisfied the responder and non-responder status defined in our study were included. Patients were classified as Responder if 1) they recovered (i.e., transitioned from 'caseness' to 'non-caseness' post-treatment), and 2) their score reduction was greater than the Reliable Change Index of ≥6 on the PHQ-9 (Jacobson & Truax, 1991). Patients were classified as Non-Responders if they met neither of these criteria, and patients who met only one of these criteria were treated as intermediate cases that were removed from analyses. Finally, as network estimation is influenced by sample size (Burger et al., 2022), we yielded equal-sized groups for Responders and Non-Responders by matching the cohorts using 1:1 propensity score matching (n=20,259 per group), where each patient with a specific number of days in treatment in the Responder group was matched to another patient with the same number of treatment days in the Non-Responder group, independent of their clinical scores.

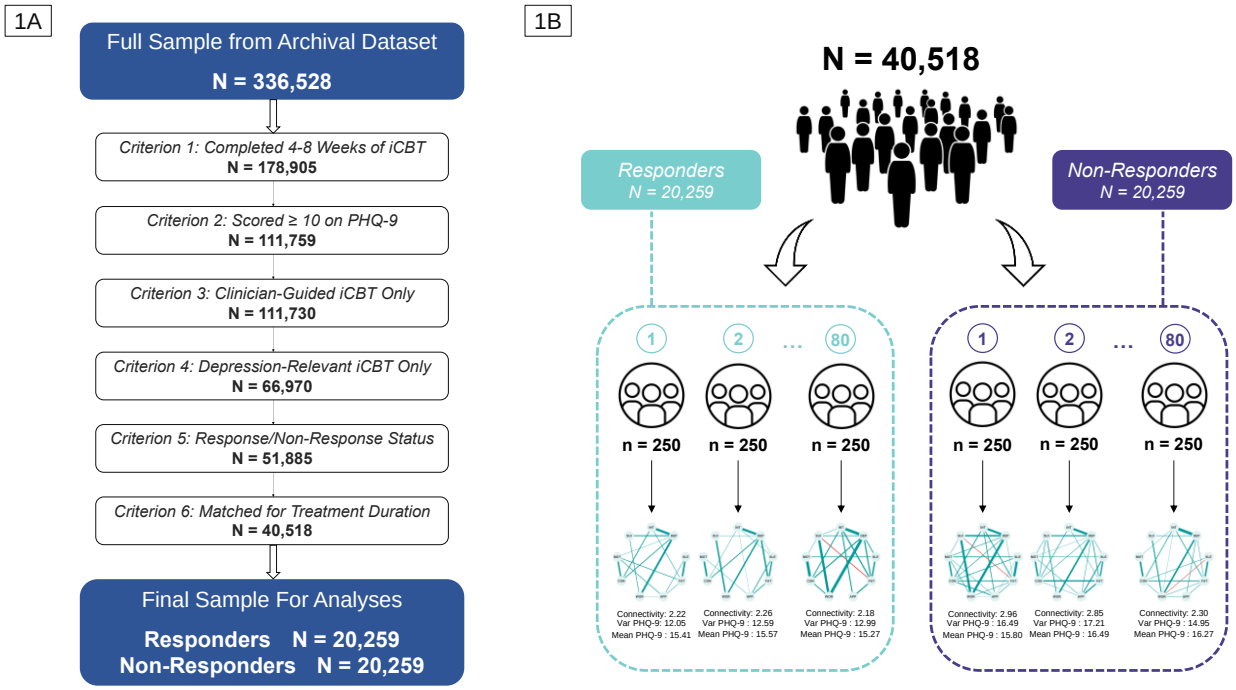


Figure 5.1 Sampling procedures for analyses. (A) Final study sample flow chart with inclusion and exclusion criteria. **(B)** Subsampling procedure for parametric analyses testing whether baseline depression severity and variance explained the association between network connectivity and treatment response. The Responder and Non-Responder samples were divided into 80 sets of n=250, respectively, where each set differed naturally in PHQ-9 baseline mean and variance. NOTE: INT =

'loss of interest/pleasure', DEP = 'depressed mood', SLE = 'sleep', FAT = 'fatigue', APP = 'appetite', WOR = 'worthlessness', CON = 'concentration', MOT = 'psychomotor problems', SUI = 'suicidality'.

5.2.4 Data Analysis

Baseline and Pre-Post Score Analyses. Differences in PHQ-9 sum and item scores at baseline and follow-up, along with treatment engagement, were compared across Responders and Non-Responders using t-tests and ANOVA.

Network Analysis. Cross-sectional networks using Gaussian Graphical Models were estimated for Responders and Non-Responders at baseline and follow-up using all items of the PHQ-9 (Epskamp & Fried, 2018; Epskamp et al., 2018a). Relationships between symptoms (nodes) were estimated using partial correlations (edges) (i.e., the relationship between two symptoms after controlling for the others within the same network). The glasso regularization penalisation technique based on the Extended Bayesian Information Criterion was performed during model selection (Chen & Chen, 2008). A tuning hyperparameter ($\gamma = 0.5$) was employed to find the optimal balance between parsimony and goodness of fit of the network. Network connectivity was defined as the weighted sum of the signed associations between nodes. For symptom centrality, we focused on examining node strength as one of the most evaluated and intuitive metrics in psychological networks. It quantifies the strength of a node's direct connections to other nodes in the network (Bringmann, 2021). Statistical significance testing on network connectivity, edge-specific, and centrality differences were conducted using the Network Comparison Test (NCT) (van Borkulo et al., 2022). The NCT is a two-tailed resampling-based permutation test that compares network differences between two independent, cross-sectional networks (responders and non-responders). Edge-difference networks (i.e., subtracting two network covariance-matrices) were used to illustrate significant edge differences between networks.

Power Estimation. To determine the required sample size to detect connectivity differences between Responders vs. Non-Responders at baseline, we repeated the NCT 1000 times for random subsets of n=250, n=500, n=750, and n=1000 per group and reported the statistical power, i.e., the proportion of samples in which a significant difference was detected.

Subsampling Analysis. To test if the relationship between connectivity and treatment response is explained by baseline severity and/or variance differences, we divided our sample into 160 independent subsamples of n=250, of which 80 were Responders and 80 were Non-Responders (**Figure 5.1B**). Each subsample naturally varied in baseline PHQ-9 sum score mean and PHQ-9 sum score variance, which allowed us to treat these networks as unique observations in linear regressions predicting network connectivity from response status, with baseline PHQ-9 sum score mean and PHQ-9 sum score variance as covariates. Using these independent subsamples, we further tested the added prognostic value of network metrics for treatment success; we contextualised the magnitude (i.e., effect size) of the association between baseline network connectivity and treatment response by comparing it to other baseline measures in univariate regressions, with response status as the IV, and the severity and variance of PHQ-9 sum and item score as well as strength centrality of individual symptoms as DVs. We repeated this procedure to test for differences in network connectivity, prior to and after treatment.

Generalisation Test. To test if our main results generalised, we applied the same analytical procedures to two other samples from our dataset. This included 1) a smaller group of patients (N=22,952) who underwent a longer course of iCBT (8-12 weeks) for depression to examine treatment duration effect (**Supplementary Materials 8.4.1**) and 2) a larger group of patients (N=70,620) who received iCBT for anxiety to probe whether observed findings were disorder-specific, where response status and networks were based on the Generalised Anxiety Disorder-7 (GAD-7) (Spitzer et al., 2006) (**Supplementary Materials 8.4.2**). The main dataset partially overlapped with both of these datasets (33% for the 8-12 weeks iCBT sample, 49% for the GAD sample).

All data processing and analyses were conducted using *R* (version 4.1.1). We used specific packages such as *MatchIt* for group matching (Ho, et al., 2011), *qgraph* for network visualisation (Epskamp et al., 2012), *bootnet* for network estimation (Epskamp & Fried, 2018), and *NetworkComparisonTest* for network comparisons (van Borkulo et al., 2022).

5.3 Results

5.3.1 Sample Characteristics

Non-Responders had significantly higher baseline PHQ-9 sum score mean and PHQ-9 sum score variance (M=16.26, SD=4.03) compared to Responders (M=15.33, SD=3.56)

(mean difference: t[40516]=24.64, *P*<.001; variance difference: F=1.28, *P*<.001). Non-Responders also scored higher on all PHQ-9 items and had greater variance in 'loss of interest/pleasure', 'depressed mood', 'psychomotor problems', and 'suicidality' (**Table 5.1, eFigure 8.4.1**). By definition, Responders exhibited a larger reduction post-treatment (M=10.06, SD=3.47) than Non-Responders (M=0.13, SD=3.36) in PHQ-9 sum score, t(40516)=292.83, *P*<.001, even after controlling for imbalance in baseline PHQ-9 sum score mean, F(1, 40515)=121473.12, *P*<.001 (**Figure 5.2A, eTable 8.4.1, eFigure 8.4.2**). On average, Responders were in treatment one day longer (M=44.17, SD=7.93) than Non-Responders (M=43.07, SD=8.22), t(51883)=-15.36, *P*<.001. There were more Non-Responders (68%) receiving depression-only iCBT vs. comorbid depression-anxiety iCBT than Responders (65%), X^2=64.09 (2), *P*<.001.

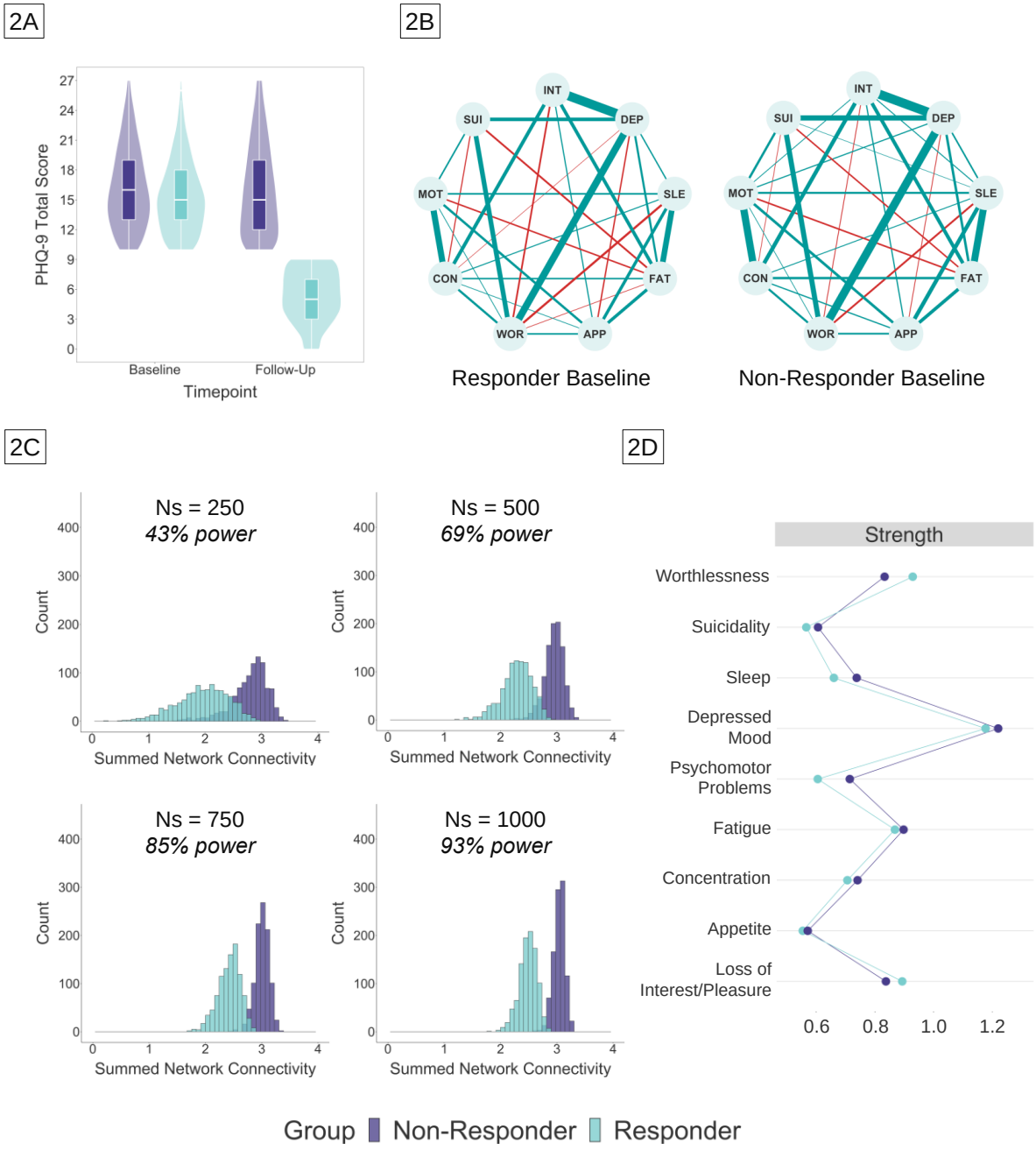


Figure 5.2. Full-Sample Network Differences at Baseline. (A) PHQ-9 sum score before and after iCBT by responder group. **(B)** Baseline network visualisations by responder group, where green and red edges denote positive and negative partial correlations, respectively. NOTE: INT = 'loss of interest/pleasure', DEP = 'depressed mood', SLE = 'sleep', FAT = 'fatigue', APP = 'appetite', WOR = 'worthlessness', CON = 'concentration', MOT = 'psychomotor problems', SUI = 'suicidality'. **(C)** Power analyses for detecting network connectivity differences at baseline. Responder and Non-Responder groups were randomly subsampled 1000 times each at N=250, N=500, N=750, and N=1000. **(D)** Strength centrality comparisons of PHQ-9 symptom nodes between Responders and Non-Responders at baseline.

Table 5.1. Comparisons of PHQ-9 item and sum score means and variances of Responders and Non-Responders at baseline.

PHQ-9 Variable	Response Status		t / F (p)
	Responders (N=20,259)	Non-Responders (N=20,259)	
Loss of Interest/Pleasure			
Mean	1.90	1.96	8.27 (*P*<.001)
Variance	0.62	0.67	1.07 (*P*<.001)
Depressed Mood			
Mean	2.01	2.06	7.28 (*P*<.001)
Variance	0.60	0.63	1.06 (*P*<.001)
Sleep			
Mean	2.18	2.29	13.28 (*P*<.001)
Variance	0.75	0.70	0.93 (*P*<.001)
Fatigue			
Mean	2.32	2.42	13.69 (*P*<.001)
Variance	0.56	0.52	0.93 (*P*<.001)
Appetite			
Mean	1.76	1.90	14.03 (*P*<.001)
Variance	0.98	0.92	0.95 (*P*<.001)
Worthlessness			
Mean	2.02	2.07	6.07 (*P*<.001)
Variance	0.75	0.74	0.99 (*P*=.35)
Concentration			
Mean	1.84	1.72	12.92 (*P*<.001)
Variance	0.80	0.80	1.00 (*P*=.89)
Psychomotor Problems			
Mean	0.92	1.05	14.40 (*P*<.001)
Variance	0.85	0.93	1.10 (*P*<.001)
Suicidality			
Mean	0.51	0.66	20.01 (*P*<.001)
Variance	0.54	0.72	1.35 (*P*<.001)
PHQ-9 Total			
Mean	15.33	16.26	24.64 (*P*<.001)
Variance	12.70	16.21	1.28 (*P*<.001)

Note: All p-values indicated above for PHQ-9 item comparisons have been adjusted for multiple significance testing using the Hochberg method.

5.3.2 Full Sample Network Differences at Baseline

The Non-Responder network had greater connectivity than the Responder network at baseline (3.15 vs 2.70, S=0.44, *P*<.001) (**Figure 5.2B**). This effect was small; a power analysis revealed that n=750 per group was required to achieve 85% power to detect this (**Figure 5.2C**). When we further matched both groups on baseline PHQ-9 sum scores, thereby matching on both PHQ-9 sum score mean and PHQ-9 sum score variance (n=18,281 per group; mean difference: t[36560]=0, *P*=1; variance difference: F=1.00, *P*=1), connectivity differences disappeared between Responders and Non-Responders

(2.73 vs. 2.72, S=0.008, *P*=.80), suggesting that sum score mean and/or variance drive the effect. We found 10/36 edges were significantly different between-groups (all *P*<.05) (**eTable 8.4.3**, **eFigure 8.4.4A**). The Non-Responder network had two more edges present, while the Responder network had five weaker positive edges and two stronger negative edges. With regards to strength centrality (**Figure 5.2D**), 'depressed mood' was the most central symptom in both networks (1.18 vs 1.22, *P*=.17). Responders exhibited greater strength in 'worthlessness' (0.93 vs 0.83, *P*=.004) and 'loss of interest/pleasure' (0.89 vs 0.84, *P*=0.047), while 'sleep' (0.66 vs 0.74, *P*=.02) and 'psychomotor problems' (0.61 vs 0.71, *P*=.002) were significantly more central in the Non-Responder network (**eTable 8.4.2**).

5.3.3 Parametric Analysis of PHQ-9 Sum Score Mean, Variance and Network Connectivity

Responders and Non-Responders differed in baseline PHQ-9 sum score and symptom mean, PHQ-9 sum score and symptom variance, and network connectivity. To disentangle these features, we drew 160 independent samples of n=250 Responders and n=250 Non-Responders (i.e., 80 subsets per group). We found that baseline PHQ-9 sum score mean and PHQ-9 sum score variance were positively correlated in the networks of both Responders (r=0.47, *P*<.001) and Non-Responders (r=0.35, *P*=.002), where the greater the PHQ-9 sum score means within each subsample, the higher the PHQ-9 sum score variances (**Figure 5.3A**). We estimated networks for each subsample and found that networks were more connected in Non-Responders (β=-1.35, SE=0.11, *P*<.001) (**Figure 5.3D**). However, network connectivity across these subsamples was positively associated with baseline PHQ-9 sum score mean (Non-Responders, r=0.23, *P*=.04; Responders, r=0.20, *P*=.08; **Figure 5.3B**) and PHQ-9 sum score variance (Responders r=0.41; *P*<.001; Non-Responders r=0.58, *P*<.001; **Figure 5.3C**). Taking these network characteristics forward to a multiple linear regression analysis, group differences in network connectivity survived after controlling for baseline PHQ-9 sum score mean (β=-0.71, SE=0.26, *P*=.007, **Figure 5.3E**), but not PHQ-9 sum score variance (β=-0.28, SE=0.19, *P*=.14, **Figure 5.3F**).

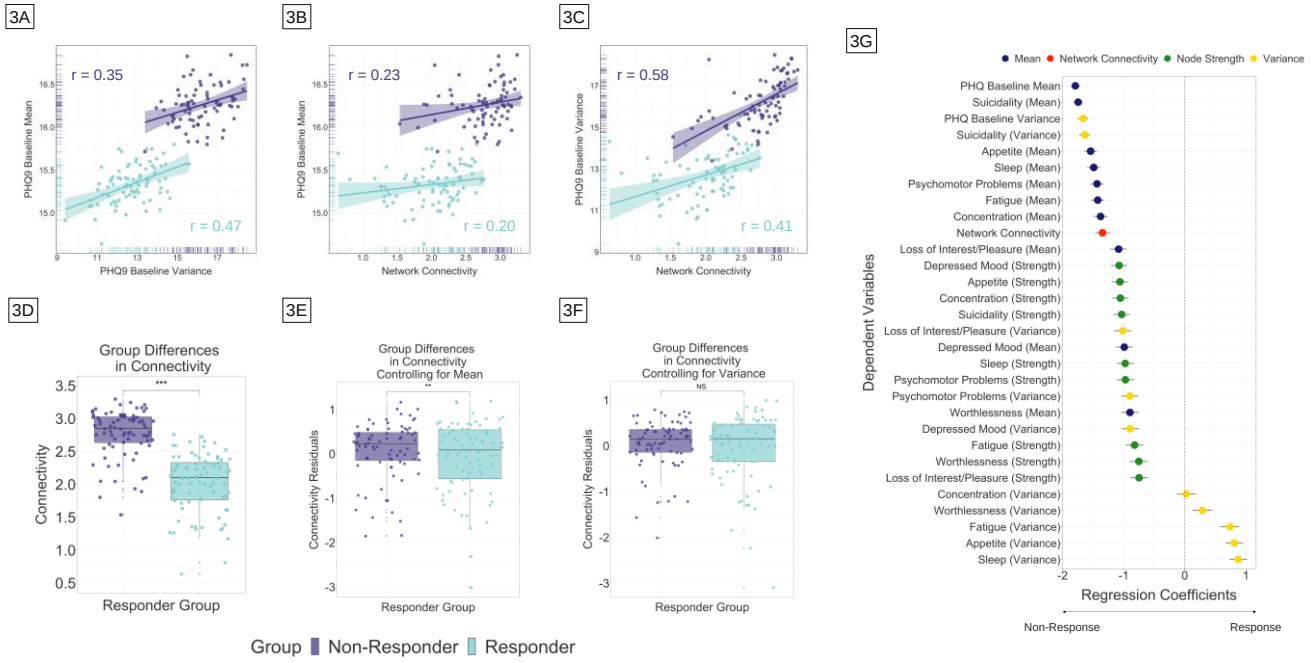


Figure 5.3. Parametric analyses on independent subsamples of Responders and Non-Responders. **(A)** Correlation between baseline PHQ-9 mean and variance of 160 independent samples of N=250 participants used to construct networks, split by responder group. **(B)** The same analysis was carried out for baseline PHQ-9 mean and baseline network connectivity of these networks, and **(C)** baseline PHQ-9 variance and network connectivity. **(D)** Network connectivity differences between 80 Responder and 80 Non-Responders networks overall, and **(E)** after controlling for baseline PHQ-9 mean (p=0.007) and **(F)** variance (p=0.14), respectively. **(G)** Regression analyses with response status (Responder, Non-Responder) as IV and individual symptom features (mean, variance, centrality) as DVs. All regressions were statistically significant (all p<0.05), except for symptom variance in concentration (p=0.89) and worthlessness (p=0.07).

5.3.4 Parametric Analysis of Symptom-Level Data

The subsampling analysis further revealed between-group differences in symptom strength, where the centrality of all symptoms were higher in the Non-Responder subsets (all *P*<.001, **Figure 5.3G**). To contextualize these differences, we compared their effect sizes relative to the mean and variance of individual symptoms, and the aggregate measures from the prior section. We found that baseline PHQ-9 sum score mean was the most strongly associated with response status (β=-1.79, SE=0.07, *P*<.001), where Non-Responders had greater baseline severity. This was followed by 'suicidality' mean (β=-1.74, SE=0.08, *P*<.001), and baseline PHQ-9 sum score variance (β=-1.67, SE=0.09, *P*<.001) (**Figure 5.3G**). Notably, the mean score of every symptom (except 'depressed mood') was more associated with treatment response than its centrality. The strength of

'depressed mood', the most central symptom at baseline for both groups, had the highest signal for treatment response of all other symptom strengths, but was still weaker than 7/9 measures of item means.

5.3.5 Network Connectivity Changes Following Treatment

Examining changes following treatment, the overall network connectivity of the full sample increased from baseline to follow-up (2.97 vs. 4.08, S=1.10, *P*<.001). These effects were evident separately in both the Responder networks (2.70 vs 3.25, S=0.55, *P*<.001) (**eFigure 8.4.3A, 8.4.3C, 8.4.4C; eTable 8.4.4**), and Non-Responder networks (3.15 vs 3.52, S=0.38, *P*<.001) (**eFigure 8.4.3B, 8.4.3D, 8.4.4D, eTable 8.4.5**). At follow-up, Non-Responders continued to have a more connected network (3.52 vs 3.25 S=0.27, *P*<.001) compared to Responders (**eFigure 8.4.3C, 8.4.3D, 8.4.4B, eTable 8.4.6**). In the subsampling analysis[6], we examined network connectivity in both groups, pre- and post-treatment. A repeated measures ANOVA revealed a significant main effect of Group, where Non-Responders had overall more connected networks, F(1, 156)=197.23, *P*<.001. There was also an effect of Time, where networks increased in connectivity from baseline to follow-up, F(1, 156)=545.45, *P*<.001. Finally, there was a Group by Time interaction, F(1, 156)=37.44, *P*<.001, driven by the fact that Responders had greater increases in connectivity (M=1.05, SD=0.52), t(78)=-18.14, *P.adj*<.001, compared to Non-Responders (M=0.62, SD=0.37), t(78)=-14.79, *P.adj*<.001. PHQ-9 sum score variance decreased over time for Responders but increased for Non-Responders, both likely a function of the small range of values required to qualify for 'response' (**eFigure 8.4.5**). Correlational analyses revealed that changes in network connectivity were not associated with changes in PHQ-9 sum score mean for Responders (r=0.09, *P*=.44) nor for Non-Responders (r=0.06, *P*=.60). For both cohorts, changes in network connectivity were positively associated with changes in PHQ-9 sum score variance (Responder r=0.42, *P*<.001; Non-Responder r=0.49, *P*<.001).

5.3.6 Replication and Generalisation

To test the robustness of our main findings, we repeated the core analyses for two partially overlapping datasets including 1) patients receiving iCBT for 8-12 weeks

[6]The subsampling analysis pertaining to network connectivity changes following treatment included 79 Responder subsamples and 79 Non-Responder subsamples due to the removal of n=1 subsample network that had a non-positive definite covariance matrix, which is unstable.

(N=22,952) and 2) where networks were based on anxiety symptoms (N=70,620). We replicated our results in both sensitivity analyses: at baseline, the full sample Non-Responder network was more connected than the full sample Responder network in both patients undergoing longer treatment (3.08 vs. 2.74, S=0.34, *P*<.001) and in those receiving iCBT for anxiety (2.68 vs 2.42, S=0.26, *P*<.001). Parametric analyses revealed that, in both cases, connectivity differences between Responders and Non-Responders were no longer significant when sum score variance was accounted for in the model (patients undergoing longer treatment: β=0.23, SE=0.20, *P*=.25; patients undergoing treatment for anxiety: β=-0.19, SE=0.12, *P*=.13). Baseline sum score mean and sum score variance were also once again more predictive of treatment response than baseline network connectivity in both patients undergoing longer iCBT (mean: β=-1.53, SE=0.14, *P*<.001; variance: β=-1.50, SE=0.14, *P*<.001, connectivity: β=-1.07, SE=0.18, *P*<.001) and those receiving iCBT for anxiety (mean: β=-1.86, SE=0.05, P<.001; variance: β=-1.78, SE=0.06, P<.001; connectivity: β=-1.67, SE=0.08, *P*<.001). Lastly, correlational analyses examining network changes following treatment confirmed an association between sum score variance changes and network connectivity changes in both patients undergoing longer iCBT (Non-Responders: r=0.34, *P*=.02; Responders: r=0.23, *P*=.127) and patients receiving iCBT for anxiety (Non-Responder: r=0.59, *P*<.001; Responder: r=0.49, *P*<.001) (see **Supplementary Materials 8.4.1 and 8.4.2** for a detailed report of sensitivity analyses findings).

5.4 Discussion

Prior work has suggested that patients with more tightly connected symptom networks are more treatment resistant (Cramer et al., 2016; Pe et al., 2015; van Borkulo et al., 2015). However, existing studies are based on comparisons of single responder vs. non-responder cross-sectional networks (Fisher et al., 2018), with relatively low samples (Forbes et al., 2017), that do not account for symptom variance (Terluin et al., 2016). We addressed these gaps in a sample of N=40,518 that was analysed as a whole, and also divided into subsamples, thereby permitting parametric analyses of the role of variance in connectivity estimates, separate to that of severity. In the single network comparison, we found that connectivity was greater for Non-Responders than Responders at baseline. This effect was small, however, requiring n=750 per group to reliably detect it, and we identified two potential confounds: Non-Responders had greater depression severity and variance at baseline. To disentangle these effects, we created 160 independent networks

of Responders and Non-Responders (n=250 each), and tested across networks if severity and/or variance explained connectivity differences. While the Non-Responder networks were on average more connected than the Responder networks, after controlling for sum score variance, the association between connectivity and treatment response was no longer significant. We replicated this result in two partially overlapping generalisation samples, one with patients undergoing a longer duration of iCBT (8-12 weeks) and another based on anxiety, not depression (4-8 weeks).

This paper highlights an important confound that is under-studied in the network literature (Bos & De Jonge, 2014; Terluin et al., 2016); network estimation is based on (partial) symptom correlations which depend on the variance of these symptoms, not just their covariance. Imbalances in variance may be an inherent clinical characteristic of treatment-resistant groups (Friedman et al., 2012), but can also be easily introduced when subgrouping patients based on treatment response (leading to range restriction of items) (Linn, 1968). That said, it is important to recognise that variance contributes to, but is not the same as, network connectivity. For example, correlations between network connectivity and variance were moderate (r=0.41-0.58), and recent work examining temporal, intraindividual networks found associations with symptom change over time that survived controlling for variance (Kelley et al., 2023). Moreover, networks actually became more connected following treatment despite reductions in both symptom severity and variance, a counter-intuitive finding most consistent in the network literature (Beard et al., 2016; Berlim et al., 2021; Blanco et al., 2020; Bos et al., 2018; Curtiss et al., 2021). One explanation is that increased symptom connectivity is not *necessarily* bad; a highly connected network should theoretically lead to a more malleable system, but not necessarily worsening mental health (Fried et al., 2016b; McElroy et al., 2019), as therapeutic gains may be due to systems becoming less 'stuck' and more open to change. Recent work examining personalised network dynamics in healthy individuals supports this; individuals with more connected depression networks tended to have greater fluctuations in depression over 6 weeks, but these went in both positive and negative directions of change (Kelley et al., 2023).

Consistent with prior work (Hagan et al., 2021; O'Driscoll et al., 2021; Robinaugh et al., 2016), 'depressed mood' had the highest strength centrality for both groups at baseline. However, we found that the severity of all symptoms (except 'worthlessness') was more strongly linked to treatment response than the strength of 'depressed mood', and that the

severity of each symptom was more predictive than its corresponding strength centrality. Both baseline severity and variance were also more predictive of treatment success than network connectivity. The lack of added prognostic value of network metrics was previously challenged by Spiller and colleagues (2020), who found that both mean symptom severity and count were more predictive of symptom changes than centrality indices. Together these findings question the real-world prognostic utility of cross-sectional network metrics, over and beyond basic self-report symptom data readily available at baseline.

There are several limitations of this study. Firstly, this was a retrospective, observational study with no control group. Information on patient demographics and concurrent treatment such as medication status were also not available. It therefore remains unclear whether the observed results can be generalised to networks estimated with patients undergoing alternative treatment (e.g., antidepressant medication). Our main study sample was also limited to patients who, on average, scored on the cusp of the cut-off for determining less/more severe depression at baseline (i.e., 16 on the PHQ-9), and therefore may not be representative of all patients with depression enrolled in primary care (NICE, 2022). In addition, while the PHQ-9 is widely used for detecting and monitoring depression symptoms within routine care settings, the instrument is primarily purposed for screening depression symptoms against the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV) by combining related symptomology into single items (Harrison et al., 2021). Future research should consider gold-standard symptom assessments such as semi-structured clinical interviews specifically designed and validated for in-depth assessment of individual symptoms of depression (e.g., Wing et al., 1990). As previously noted, networks estimated from cross-sectional data do not always generalise onto an individual-level (Hamaker, 2012; von Klipstein et al., 2021), and indeed, differences in baseline sum score mean and sum score variance can be introduced systematically by the binary definition of 'response' that is required for cross-sectional network analysis. The crucial next step for network theory is to move towards a dynamical account of psychopathology afforded by personalised, within-subject networks for each patient undergoing treatment overtime.

5.5 Conclusion

In a large sample of >40,000 patients, we determined that network connectivity differences between iCBT responders and non-responders are small, requiring hundreds

of patients to be appropriately powered. We highlighted that symptom variance is an important confound to interpreting cross-sectional network effects and may drive prior findings of increased baseline connectivity in treatment non-responders. The baseline mean and variance of depression sum and symptom scores fared better at predicting response than both overall network connectivity and individual symptom strength centrality.

Chapter 6 – General Discussion

6.1 Summary

We have evidence that iCBT is a flexible, stigma-free, and scalable solution that can widen and expedite access to evidence-based support in depression (Andersson et al., 2019; Cuijpers et al., 2010; 2018; Hobbs et al., 2017; Karyotaki et al., 2021; Spek et al., 2007). Despite its increase in clinical use, our understanding of who it best benefits, and why, remains limited. It is clear a single mental health treatment does not work equally for everyone, but progress towards *personalised medicine* for psychotherapy has been remarkably slow (Simon & Perlis, 2010). Mullarkey and Schleider (2021) allude that the primary reason for this lack of progress is the critical mismatch between our aspiration for clarity about which treatment works best for whom, and the often ill-equipped study designs and methods that we use to test this. Treatment decisions are still being made on the basis of trial-and-error, personal experiences, and potentially unreliable anamnestic information, rather than objective and replicable markers that can reliably form a patient profile for which iCBT is optimal for. Notably, a common methodological challenge in the realm of precision psychiatry to provide such markers is limited sample power.

To this end, this thesis emphasises the need for a shift towards a *big data* approach to delve deeper into the mechanisms and predictors of iCBT for depression. Using large patient datasets, 1) I investigated the real-world effectiveness of iCBT for important subgroups to inform who we should *prescribe* iCBT to; 2) I harnessed the power of algorithmic, data-driven tools to *predict* iCBT treatment response in depressed individuals assessed prior to treatment, and lastly, 3) I applied cross-sectional network analysis to depression symptoms to attempt to *understand* why iCBT works for some, but not others. Below, I provide a summary of each chapter, encompassing their specific aims, methodologies, and findings. I then discuss this comprehensive body of evidence in light of implications concerning the research and utilisation of iCBT in mental health

service delivery. I address some important strengths and limitations of this thesis, recommend avenues for future research for expanding the scope of my work, and hope to conclude with valuable insights that can be judiciously applied to clinical decision-making regarding iCBT.

6.1.1 Summary on Chapter 2

So what sort of patient data can we use to inform initial treatment decisions in iCBT? We first began our enquiry by focussing on a particularly vulnerable, high risk subgroup prone to developing depression: individuals suffering from chronic long-term conditions. In **Chapter 2**, I investigated the effectiveness of iCBT for real-world patients with LTC and comorbid depression and anxiety symptoms. Access to appropriate psychological care for this subgroup is especially challenging. These individuals already bear the physical and financial burden of their LTC, and healthcare services often struggle to meet both their physical and mental health needs (Naylor et al., 2012). In this sense, the increased flexibility and accessibility of iCBT for this cohort is particularly pronounced (Beatty & Lambert, 2013; Naylor et al., 2012; van Beugen et al., 2014). Plenty of controlled trials support the use of iCBT to treat comorbid depressive symptoms in LTC patients across conditions like chronic pain, diabetes, cardiovascular diseases, and cancer (Adhikary et al., 2023; Mehta et al., 2018; van Beugen et al., 2014). But these studies vary in quality (Beatty & Lambert, 2013; Charova et al., 2015), and often have small sample sizes, making it difficult to compare with other interventions (Moshe et al., 2021). Importantly, robust real-world evidence was still largely missing (Kumar et al., 2017). Considering the challenges in translating trial findings to clinical settings (e.g., overestimation of effects and reduced sample generalizability; Moshe et al., 2021), we thought it crucial to investigate the effectiveness of iCBT for patients with LTC with adequate statistical power when the intervention is implemented in a naturalistic service context.

To do this, we retrospectively analysed a large-scale dataset comprising 4 years of routine care data from 21,051 patients, including 4024 with LTC, who enrolled in low-intensity CBT interventions (i.e., iCBT, guided self-help, and group therapy) under the UK's Improving Access to Psychological Therapies (IAPT) program. IAPT provides treatment to >500,000 patients per year within the public health system; its IAPT-LTC pathway is one successful example that integrates both physical and mental health support. The service follows the stepped-care model in treatment delivery, which many sees as a

starting point for tailoring treatment options to patients with varying clinical presentations (Clark, 2018). In this naturalistic context, we found low-intensity CBT-based interventions as a whole, delivered at IAPT step 2 and 3, improved depression more in LTC patients and improved anxiety more in non-LTC patients, but the differences between which were small (i.e., <1 point difference on the scales) and unlikely to make a clinical difference; they are in general just as effective in treating depression and anxiety in patients with and without LTC. This underscores the merit of integrating these interventions in mainstream healthcare service delivery, with high ecological validity that aligns with broader findings concerning psychotherapies in general (Cuijpers et al., 2023).

In regular routine care, a key enquiry for healthcare professionals and service providers is whether we can expect to see any additional benefits from iCBT in comparison to the array of alternative interventions for patients with LTC. In this sense, we found iCBT to outperform guided self-help and group therapy, again regardless of whether someone has a LTC or not. iCBT combines psychoeducation with clinician support, helping patients better understand health-related information which can motivate them to improve related behaviours (Murray, 2008). With the convenience of it being online, it offers easy, stigma-free access for individuals with LTC to access psychological help who may otherwise struggle with in-person care (van Beugen et al., 2014). This work offers important insights for managing patients with LTC and comorbid depression and anxiety symptoms. Clinicians now have highly valuable, robust guidance when it comes to choosing the most appropriate treatment for this vulnerable cohort, and can confidently consider iCBT as a viable option amongst other interventions.

Above is a clear example of how knowing which subgroups for whom iCBT is more suitable than other interventions, can personalise care pathway for patients with depression. One major caveat here is that we cannot determine whether additional treatment personalisation (i.e., tailored interventions, disease-specific measurement) might change our observations, given evidence that suggests a tailored approach may be more preferable when treating psychological comorbidities in this cohort (Hind et al., 2010). The study also lacked details about each patient's specific chronic illness; what it was, how many they had, how long were they diagnosed for, all of which could interact with their treatments and potentially influence study findings. Moreover, while physical comorbidities is a reliable marker for depression treatment response (Jankowsky et al., 2022), there are many more characteristics about a person that carry important prognostic

signal for their likelihood to respond to iCBT. Trying to personalise treatment based on a one-at-a-time approach does not get us (and has not gotten us) anywhere, when each factor on their own improves treatment outcome prediction by a tiny amount (Chekroud et al., 2021; Simon & Perlis, 2010). Even in attempts that try to examine more than a few predictors at a time, samples are too small and unrepresentative, often leading to false positives and overestimation of effect sizes (Forstmeier et al., 2017).

In this sense, **Chapter 3 and 4** recognise the importance of this multivariable, big-data approach, which underlies the core aims of the *Precision in Psychiatry* (PIP) study introduced in these chapters. If we wish to capture the complexity and heterogeneity of depression and reliably estimate a person's likelihood of responding to iCBT, we need to move beyond underpowered studies and single-factor approaches, and acquire a substantial amount of data from a large and diverse pool of individuals. A critical challenge therefore is how we can get the large, high-quality data needed for this endeavour. The good news is, given the scalability of iCBT for rapid acquisition of patients at scale while standardising longitudinal treatment outcome monitoring, the field has been very excited about its potential to bring about a novel experimental paradigm that appeals to the application of data-driven algorithmic prediction tools (Chekroud et al., 2021; McNamara et al., 2022). To be able to develop robust, multimodal prediction models for treatment response in psychiatry, we propose that aspects of digitisation, like in iCBT, should be adopted into the way we collect our data, which are much more time- and cost-effective than traditional, lab-based methods.

6.1.2 Summary on Chapter 3

Chapter 3 introduced and tested an innovative, online-based methodology in the PIP study for collecting data from individuals undergoing iCBT and antidepressant medication during the first 4 weeks of their treatments. Online data collection is gaining traction in psychiatry research, due to its potential to cost-effectively and rapidly gather data from large, diverse samples for repeated longitudinal assessments (Gillan & Rutledge, 2021). We showed this clearly in the PIP study, where our fully internet-based protocol significantly expedited remote data collection from treatment-seeking samples with excellent study retention (85%) and treatment adherence (≥97%) in just ~1 year of active recruitment, surpassing that in traditional lab-based and research consortia settings. Pairing this method with iCBT allowed for more expansive enrolment of participants compared to antidepressant medication (a monthly average of 47 baseline iCBT

completers compared to 15 in the antidepressant arm), making it a powerful tool for scaling up recruitment of large samples needed for precise, reliable prediction models (Gillan & Whelan, 2017). Although the study lasted only 4 weeks, both treatment arms showed significant improvements not only in depression but also in 11 transdiagnostic psychiatric symptoms. Even with no in-person elements to the protocol, study schedule compliance was excellent: there was only ~1 day difference on average between the scheduled and the actual date of assessment completion.

Online data collection offers not only speed and convenience but also depth. In the PIP study, it facilitated collection of a broad range of mental health relevant variables (>600 variables per person) from treatment samples shown to be demographically and clinically representative over time (Berinsky et al., 2012; Shapiro et al., 2013). While self-report data are typically more popular in that they offer meaningful insights and practical value for real-world clinical integration (Chekroud et al., , we can also easily obtain estimates of neural correlates and brain function using computerised cognitive tasks, which are compatible with online administration (Coughlan et al., 2019; McNab et al., 2015; Rutledge et al., 2014). Naturally, there were concerns about increased noise and bias in online data, which we attempted to mitigated using regular objective attention checks. While a good proportion of participants reported being distracted during assessments (~66%), this did not impact negatively on our data quality; indeed, only a small percentage failed attention checks (8% failed 1, and 3% failed >1), and the 4-week reliability of self-report assessments were generally high (in height and all clinical assessments). Qualitative participant feedback was also mainly positive, with many finding the assessments flexible, easy to complete, and reflective, although some regarded the gamified tasks to be somewhat tedious, while others suggested longer study durations and more self-report assessments.

In brief, an internet-based methodology is promising for longitudinal treatment-oriented research in depression, permitting rapid, cost-effective and in-depth data collection with good acceptability and high data quality. Importantly, this study provides a template for future research combining digitised research methods and intervention designs in computational psychiatry research, a field dominated by cross-sectional research methods. Nonetheless, there were limitations to this approach. The PIP study was observational, which precludes us from making strong causal claims due to the non-random allocation of participants to their treatment arms. Participants in both arms had

differences in baseline characteristics and symptom improvement, and some were also undertaking other treatments for their mental health. Keeping these limitations in mind, we brought forward our online sample for an algorithmic-approach to predict iCBT response in chapter 4.

6.1.3 Summary on Chapter 4

Chapter 4 saw us applying machine learning techniques to develop prognostic models for early depression treatment response in patients initiating iCBT, using pre-treatment self-report and cognitive data. The focus on using pre-treatment data permits us to apply the model to guide treatment decisions before treatment starts. These models can assess many complex factors at once and gauge their relative importance towards prediction, providing more insightful information about iCBT response compared to post-hoc explanatory methods (Yarkoni & Westfall, 2017). We compared an array of ML methods with varying levels of feature granularity through rigorous cross-validation, tested the generalisability of the best model in hold-out data, and assessed how specific our predictions were to iCBT response. The winning model, an elastic net regression with 31 self-report (total scores) and cognitive features, explained 14.6% variance in 4-week iCBT treatment outcomes during cross-validation, significantly surpassing a clinical criterion (R^2>6.3%; Uher et al., 2012) and outperforming a simple benchmark model by 5.6%. The model did even better when we tested it on hold-out data (18.5% variance explained). Its performance was similar to other studies that had longer follow-up periods (Buckman et al., 2021; Chekroud et al., 2016; Hornstein et al., 2021; Iniesta et al., 2016), but with added implications for optimising iCBT treatment personalisation based on early responses. In an important advance over prior work, we tested the specificity of our model and found that it performed comparably in the antidepressant group (17.7% variance explained), suggesting that our predictions were not specific to iCBT but also applicable to antidepressant medication treatment response. Successful prediction of this kind marks an important step forward in clinical practice; it implies that these models could inform optimal treatment allocation decisions before treatment even begins, saving clinicians and patients time, energy, and resources.

This prediction performance of our best model, while good, still has room for improvement before it can be effectively used for clinical decision-making. Deo (2015) notes that acceptable predictive model performance in clinical psychology comes from 1) measuring the right things and 2) measuring enough of them. There is often a trade-off

between innovative, theoretically driven measurement of rich phenotypes and having a sufficiently large dataset. For instance, studies using neuroimaging data can easily lead to overfitting due to a high ratio of features to participants who are costly to recruit. Conversely, studies with large N tend to rely on electronic health records with limited available measurements. Even with a substantial sample size of N>70,000 patients, Bone and colleagues (2021) showed that advanced ML algorithms failed to outperform traditional logistic regression using routine electronic health records. At a minimum, a simulation study by Luedtke and colleagues (2019) suggests we need at least N=500 patients per arm to approximate a treatment assignment rule. This figure is also likely optimistic, considering it was only compared to random treatment assignment. More participants would be required if the data is noisy and sophisticated methods are used to detect any nonparametric and interactional effects between predictors (Jacobucci & Grimm, 2020; Luedtke et al., 2019; McNamara et al., 2022). In this sense, we demonstrate how moving towards digital methods and intervention can help bridge this gap, allowing the collection of theory-based, multimodal data from hundreds of patients in a relatively fast pace.

Our work in the PIP study effectively uncovered specific patient profiles that are more likely to benefit from early stages of iCBT treatment. Of note, baseline depression severity, as well as a host of specific depressive symptoms, played a big part in determining whether someone will respond to iCBT early or not, which is consistent with existing depression treatment literature (Chekroud et al., 2016; Hornstein et al., 2021; Iniesta et al., 2016; Webb et al., 2020). The rest of the profile featured mostly self-report data; closely following in rank was treatment expectation as well as other transdiagnostic psychiatric symptoms. Various general health, lifestyle, and environmental factors also featured on the list, but the model only included 1 commonly investigated sociodemographic variable (i.e., gender). While cognitive variables did not feature as high as most other self-report variables in the model, their inclusion still supports the value of using a multimodal approach for predicting treatment response. In this sense, ML is helpful in identifying what baseline characteristics a person has that can influence how well they respond in the early stages of iCBT. But these prediction-focused applications are not a panacea; depression and the therapy process itself are beyond intricate, and we still lack a complete understanding of how these factors influence each other, especially when it comes to the interplay between individual depression symptoms (e.g., how

motivation affects guilt or how sleep affects mood). We must adapt our methods to better understand how they interact with each other to affect outcomes.

6.1.4 Summary on Chapter 5

To do this, we took to an alternative conceptual lens, namely the network theory of psychopathology in **chapter 5**. The theory proposes that depression symptoms interact as components of a dynamical system, and it is this interaction that leads to the rise, maintenance, and persistence of depression (Borsboom, 2017). We tested a key prediction that has significant implications for depression prognosis: individuals who do not respond to treatment have more interconnected symptoms (i.e., elevated network connectivity) at baseline. Prior research investigating this premise, which relied heavily on cross-sectional data, produced mixed evidence (McElroy et al., 2019; O'Driscoll et al., 2021; Schweren et al., 2018; van Borkulo et al., 2015), likely due to insufficient sample sizes used to compare single responder vs. non-responder networks. As we delved deeper, it became evident that these factors not only affect the stability and robustness of network estimation, but that cross-sectional network comparisons are inherently inadequate to understand how intra-individual differences unfold. As a result, we sought to determine if prior research was affected by confounds such as baseline severity and variance of symptoms.

Once again leveraging the scalability of digital interventions, we featured the largest sample of iCBT patients to-date to investigate this, with >20,000 responders and non-responders who underwent 4-8 weeks of iCBT treatment. In line with network theory, the non-responder network at baseline exhibited greater connectivity than responders; the caveat being this effect was small, required hundreds of patients per group to detect it, and also disappeared once we adjusted for differences in depression variance (not severity) between-groups. This is particularly relevant because prior research has linked greater network connectivity with higher psychological vulnerability (i.e., higher mean levels of symptom severity) (van Borkulo et al., 2015). Mathematically speaking, network estimation is correlation-based and so is sensitive to differences in symptom variance, not just covariance. When symptoms become more severe, they can reach higher scores, leading to increased variance and in turn increased connectivity (Fried et al., 2016b; Terluin et al., 2016). Non-responders presented more severe and varied symptoms at baseline compared to responders, which may explain their elevated level of baseline network connectivity.

Though this presents a confound, it is important to note that while connectivity can be affected by variance, it is not synonymous with it. Indeed, we observed increased connectivity post-treatment even when variance decreased across the same timeframe, just like many studies that demonstrated elevated connectivity after successful depression treatment (Beard et al., 2016; Berlim et al., 2020; Blanco et al., 2020; Bos et al., 2018; Curtiss et al., 2021). Some have attributed this to methodological issues like low statistical power and ceiling/floor effect in symptoms (Fried et al., 2016b), or as a result of psychotherapy effects such as response shift bias (albeit also happens in antidepressant treatment), both of which are unlikely in our case. Rather, we think our finding lends support to an emerging alternative interpretation of what network connectivity means in a psychopathological context: rather than elevated psychological vulnerability, it infers increased adaptability/malleability instead. This would make sense, if one considers a tightly connected network where symptoms have the ability to highly influence each other, they can change in either direction together. For example, symptoms may deteriorate quicker with one another when triggered by external stressors, but the reverse can also happen after successful treatment, leading to 'positive spiralling' whereby improvements in one symptom can rapidly lead to improvements in others, again through a tightly connected network (Fried et al., 2016b; Kelley et al., 2023.; McElroy et al., 2019). That being said, the way symptoms change over time during treatment may not be fully understood by examining cross-sectional networks at single timepoints.

Lastly, we found that for both groups, 'depressed mood' was the most central symptom. Before we can take this forward for clinical use, we need to understand the actual predictive value of network metrics in context. Studies investigating the prognostic utility of centrality in cross-sectional networks yielded contentious findings (Hagan et al., 2021; O'Driscoll et al., 2021; Robinaugh et al., 2016)). Some even suggest that the centrality-outcome relationship was not stronger than simple feature of symptom count (Silk et al., 2019; Spiller et al., 2020). In this regard, we found that both baseline severity and variance were more predictive of treatment response than network connectivity, and that the severity of each symptom was more predictive than its corresponding centrality. This raises questions about the practical use and real-world utility of depression cross-sectional networks in clinical settings; if they do not enhance our ability to predict treatment success compared to readily available baseline data, what is their real value?

6.2 Implications

6.2.1 Testing for Treatment Specificity

Numerous studies exist when it comes to examining response prediction and mechanisms, not limited to just iCBT but across various depression treatments (Andersson et al., 2019c; Domhardt et al., 2021; Godlewska & Harmer, 2021; Herzog et al., 2018; Kazdin, 2007; Marwood et al., 2018; Roiser et al., 2012; Rost et al., 2023). In that regard, we believe efforts to determine treatment specificity are crucial to ensure the observed effects are specific to iCBT. Despite the observational nature of our work in this thesis, we endeavoured to test the generalisability and treatment specificity of our findings to iCBT. Firstly, in our subgroup analysis involving patients with LTC, we directly compared the effectiveness of iCBT with other low-intensity interventions, which provided valuable information for clinicians when choosing treatments for this cohort. In the PIP study, we trained and tested our model predicting early depression change with iCBT samples, but also externally validated it with an independent antidepressant dataset. While some studies have explored the drug-specificity of machine learning models in antidepressant medication (Chekroud et al., 2016; Iniesta et al., 2018), the PIP study is among the first to investigate the treatment specificity of a machine learning model comparing psychotherapy and antidepressant treatment.

To this end, our model generalised well on patients receiving antidepressant treatment. While this could be due to the observational nature of the PIP study, it is possible that psychotherapies like iCBT and antidepressant medication may share common but also dissimilar treatment mechanisms that operate on different levels of therapeutic action (DeRubeis et al., 2008). Then, it is perhaps not surprising to find that both iCBT and antidepressant medication induced similar transdiagnostic treatment effects in improving a range of clinically relevant symptoms other than depression. For example, behavioural activation in psychotherapy vs. neurotransmission alteration in antidepressant medication are both involved in modifying distorted cognition (Hollon, 2019), while therapeutic alliance in psychotherapy is evidently missing in pharmacology (Wampold et al., 2002). A recent study by Dunlop and colleagues (2023) showed that reduced affective network connectivity with motor systems was a shared process important for recovery with both CBT and antidepressant medication. But they also found differential brain-based effects for both treatments; medication was independently associated with broadly inhibitory effects and CBT with networks involved in cognitive control. It would be imperative to

explore this further, particularly with treatments that have distinct mechanisms such as neuromodulation techniques (e.g., transcranial magnetic stimulation, electroconvulsive therapy), to determine if specificity exist in those cases (Miloseva & Richter, 2022). Furthermore, it is also important to challenge the fundamental assumption that iCBT works the same way as traditional CBT simply because they share similar strategies and protocol. The delivery of iCBT is technologically-mediated and has varying levels of therapist involvement. These differences in turn can influence how iCBT and traditional CBT mechanisms work and affect symptom outcomes (Mogoașe et al., 2017).

Just being able to predict treatment response is no longer sufficient; we need to be able to determine which treatment is best for each individual. Prognostic models that can help us do that already exist, and are in development towards implementation, such as the personalised advantage index (PAI) based on multivariate prediction models (Cohen & DeRubeis, 2018; DeRubeis et al., 2014). The PAI determines the treatment expected to be more effective for an individual patient, and quantifies the extent of this predicted advantage. Data from RCTs have shown that individuals who were assigned to their predicted optimal treatment yielded significantly better outcomes compared to those assigned to their predicted non-optimal treatment, in different depression treatment conditions (antidepressant medication vs. CBT; (DeRubeis et al., 2014)) and different psychotherapy treatment conditions (cognitive vs. interpersonal therapy; (Huibers et al., 2015). These individualised prediction models are more clinically informative and can assist both clinicians and patients in selecting the most suitable treatment. As advanced statistical modelling approaches, new predictive factors, and big data become more accessible, these models hold the potential to enhance outcomes in depression treatment (Cohen & DeRubeis, 2018).

6.2.2 Embracing an Online, Data-Driven Approach

Traditional, theory-driven approaches to predicting and understanding depression treatment responses struggle to address the complexity and diversity of the condition (Chekroud et al., 2021b; Simon & Perlis, 2010). We think it is time to embrace a digital shift in the way we conduct research (Gillan & Daw, 2016; Lazar et al., 2017). We showed in the PIP study, that internet-based methods, when paired with iCBT, work really well in ensuring smooth and swift enrolment to enhance recruitment. Worth noting here is that the more detailed assessments at baseline and final timepoints were only accessible via a computer/laptop, which may not be readily available for everyone. We

think even better retention rates can be achieved if the entire study can be integrated in easily accessible devices like smartphones or smartwatches. These technologies can make research even more convenient and burdenless for individuals to participate (Gillan & Rutledge, 2021). We should also expand on other means of facilitating online testing, such as forming research collaborations with mental health providers to integrate online testing into patient referrals (Gillan & Daw, 2016; Gillan & Whelan, 2017). An industry-academia collaboration in this thesis illustrates the benefits of such partnerships; by collaborating with an iCBT provider, we seamlessly integrated our online study into their patient referral system. We also had access to an untapped, vast archival database of iCBT users (N>300,000), which made extensive investigations of iCBT treatment response using advanced statistical modelling possible.

Based on our comprehensive evaluation of the feasibility, acceptability, participant perspectives of online testing, we believe the *secret* to successful online testing is two-fold; make participation as easy as possible, while maintaining data integrity and quality. For remote dense and longitudinal sampling, minimising participant burden is vital for maintaining retention. Assessments should be as brief, non-repetitive, and easy to complete. Gamifying data collection, particularly with cognitive tests, can help keep participants engaged. Some successful examples include studies investigating cognitive markers like model-based planning in psychopathology (Donegan et al., 2023) and disentangling active cognitive components of CBT (Norbury et al., 2023). Relatedly, online testing enhances methodological standardisation to data collection which minimises unintended measurement variations. This is a double-edge sword in a way, as any and all administerial and technical errors can uniformly affect participants. To mitigate this, there should be clear, user-friendly assessment instructions in place. Repeated testing of the data collection pipeline before implementation and ongoing spot-checks during recruitment are also essential precautions.

Data quality in online testing is diminishing. Over time, studies have shown that participants completing studies online are becoming prone to distractions and may employ heuristic response strategies to save time and effort during assessments. While large online samples can help offset random measurement noise (Gillan & Rutledge, 2021), the impact of poor data quality is in some cases systematic. Studies show that inattentive/careless participants tend to endorse more mental health symptoms than the general online population on scales that have a negative skew, potentially increasing the

risk of false-positive errors (Chandler et al., 2020; Ophir et al., 2020). Zorowitz and colleagues (2021) found that because of this, inattentive/careless participants can induce spurious correlations (i.e., inflate/reverse observed effects) between task behaviours and symptom measures. These correlations reduced with each quality check, and even disappeared when these participants were effectively screened out.

This is nonetheless worrying, coupled with the fact that concerns about data quality from online labour markets are growing (Chmielewski & Kucker, 2020). Recent research by Burnette and colleagues (2022) and Donegan & Gillan (2022) shows that data obtained via Amazon Mechanical Turk (AMT) were of insufficient quality for drawing reliable conclusions. Both studies had to exclude hundreds of participants in their studies as a result. Furthermore, the increasing sophistication of bots in the age of artificial intelligence has rendered conventional protective measures like catch questions, attention checks, and response consistency requirements less effective (Burnette et al., 2022; Zorowitz et al., 2021). Donegan & Gillan (2022) posit the way forward for online testing should be with caution and with scrutiny; there should be robust control measures, comprehension tests, and sensitivity analyses excluding inattentive/careless participants throughout the research process. Payments should be contingent on eligible participants completing specific study sections. Understanding participants' intrinsic motivation for undertaking online research is also key. In the PIP study, participants expressed a desire to contribute to mental health research, with money incentive a secondary consideration. When participants' motivation aligns with study goals, the research process becomes more seamless and efficient. Online testing has made a transformative impact on psychiatry research, its potential in treatment prediction research response is only just beginning. The field is evolving, and researchers must continuously adapt to emerging technological advances and the changing landscape of online research (Donegan & Gillan, 2022).

6.2.3 Clinical Use of Prognostic Models

Once we are 'powered' enough with high quality, large, and rich data, the hope is that we can use it to train data-driven, prognostic models (Lazar et al., 2017). These prediction approaches, when rigorously assessed and validated, are set to complement post-hoc explanatory approaches. The PIP model was developed to predict early iCBT response using a wide range of baseline characteristics. It identifies prognostic factors at the screening stage, and provides clinicians an objective, algorithmic-based tool to guide

treatment allocation and referral decisions before treatment even begins. In a practical scenario, picture a patient seeking depression treatment at a healthcare provider, possibly on a large waitlist. The patient's information could be obtained in advance and fed into the PIP algorithm, which predicts how much improvement they will experience in the early stages of iCBT. This prediction helps the clinician decide whether to offer them iCBT based on a predefined clinical threshold of improvement. If the algorithm predicts significant improvement, the patient can start iCBT immediately. But if it predicts a higher risk of poor outcomes, they can be referred straight into alternative treatments, possibly more intensive, in-person care. A predictive model focused on early stages of treatment helps the clinician anticipate how patients will respond to treatment early on, and provides baseline recommendations meaningful to the clinician that are easy to implement (Penedo et al., 2022; Li et al., 2023; Sajjadian et al., 2021). This approach allows for timely adjustments for improved effectiveness, as shown by Delgadillo and colleagues (2018) when they monitored early on the patients for whom non-response is predicted.

Before we can think about deploying machine learning algorithms into clinical practice, we need to consider several prerequisites. Firstly, the model should rely on data readily accessible by clinicians without incurring additional resources. The PIP model used just 31 baseline variables of self-report and cognition to predict a modest amount of change in depression after 4 weeks of iCBT. These data types can be easily implemented and evaluated by clinicians or via self-ratings. In contrast, biological measures like neuroimaging or genetics are expensive, cumbersome, and difficult to assess, making them less suitable for routine clinical testing. Their inclusion should only be justified if they substantially improve performance (Rost et al., 2023). Relatedly, in clinical practice, it is important to strike a balance between model interpretability and complexity. While high accuracy is valuable, understanding why a treatment works is more often more crucial. While not always the case, complex models with high accuracy might use irrelevant features that hinder clinicians from interpreting and drawing meaningful inferences. In that regard, its clinical utility is limited, further impeding our understanding of depression's etiology, development, and maintenance (Vieira et al., 2022; Zhang et al., 2021).

The concept of explainability does not just apply to the features, but also to the treatment outcome. Across depression treatment prediction studies, there are substantial variations

in the definitions of the outcome that models predict. Typically, psychiatry research uses a dichotomous approach, which involves setting a response threshold based on the amount of improvement (Hornstein et al., 2021; Lenhard et al., 2018), a clinical cut-off for remission (Chekroud et al., 2016; Iniesta et al., 2018; Wallert et al., 2022), or a blend of the two (van Breda et al., 2017). While the dichotomous approach is arguably more clinically intuitive and relevant for treatment decision-making, it is criticised for losing valuable information (Altman et al., 1994; Royston et al., 2006). Less attention, on the other hand, has been given to predicting continuous outcome variables (Jacobson & Nemesure, 2021; Pearson et al., 2019; Webb et al., 2020). Continuous metrics such as change in depression (as in the PIP study) or final depression score post-treatment are strongly influenced by baseline severity, which is a robust predictor of iCBT treatment outcome (Button et al., 2012; Edmonds et al., 2018; Hadjistavropoulos et al., 2016) at the same time a mathematical confound to the prediction (the calculation of change includes the baseline score, and the final score post-treatment is contingent on how high baseline severity is). This confound raises concerns specifically about the practical utility of baseline scores in a predictive context. While some suggest that baseline scores may still hold intrinsic value in predicting change/followup scores that reflects therapeutic recovery processes, it is crucial we consider how ceiling effects may interact with score variability, potentially leading to confounding associations between them (Browne et al., 2010; Oldham, 1962; Terluin, 2012). We currently do not have consensus regarding a standardised definition of treatment response, and this poses a challenge for comparing the performances of different machine learning algorithms. This also goes for variation in what is deemed an 'acceptable' threshold for model accuracy; some showed that model accuracy of >65% was deemed clinically actionable and useful (Forsell et al., 2022), but no guidance has been set for predicting continuous measures in this regard. A way around this alludes to a combination of the two; train the model to predict continuous scores first (thus maximizing information available), and then categorise it post-hoc based on a threshold (thus maximizing clinical utility) (Vieira et al., 2022).

The PIP model is a significant step towards precision in predicting iCBT treatment response for depressed individuals. Like many similar studies, its performance remains modest, and we need more research to fully ascertain the usefulness of prognostic models like this as a practical tool for clinicians (Rost et al., 2023). Recent studies are just starting to apply these models in real-world settings to allocate treatments, but results are

mixed. A study by Delgadillo and colleagues (2022) randomised therapists to provide care to 951 patients in an algorithm-based stratified care arm vs. stepped care as usual arm (treatment as usual). They found patients in the stratified care arm had higher depression remission rates (52.3%) compared those in the standard stepped care arm (45.1%). Similarly, Browning and colleagues (2021) randomised depressed patients into algorithmic-informed care vs. usual care for citalopram. Difference in recovery rates between the two arms (55.9% in algorithmic-informed care vs. 51.8% in usual care) here was similar to the above study, but did not reach statistical significance. This suggests that while algorithm-assisted care shows promise, we need more effective algorithms before we proceed to their full implementation in real-world clinical practice.

To truly understand the value of these models, we also need to directly compare their predictions with those made by human expertise. Research suggest clinicians often generate their predictions based on intuition (Perlis, 2016), and therefore prone to bias (Lutz et al., 2022). If we can demonstrate that data-driven models can improve prediction accuracy beyond human experts, then they hold clinical utility (Vieira et al., 2022). Model performance matters even more especially when we consider the detrimental impact poor ML recommendations can have on treatment decisions. An experiment conducted by Jacobs and colleagues (2021) investigated the impact of ML recommendations on 220 clinicians' antidepressant treatment choices in a series of simulated patient vignettes. They did not find that treatment selection accuracy improved when clinicians received ML recommendations vs. not, but critically, incorrect ML recommendations furthered lowered their treatment selection accuracy. ML performance also influenced perceived utility of the tool; they found that correct ML recommendations correlated significantly with higher perceived utility of the tool when compared to incorrect recommendations. This highlights the importance of considering how clinicians will use machine learning models in real-world treatment decision-making, beyond just assessing the accuracy of ML tools in isolation. Nevertheless, ML in psychiatry is meant to assist clinicians, not replace them. Researchers and clinicians both have a care obligation to ensure these algorithms in treatment decision-making does not cause harm, in particularly when it comes to the sensitive nature of mental health data. These models can easily inherit biases in data, such as underrepresentation of sample characteristics (e.g., gender or race), which can potentially lead to profound discrimination if not handled sensitively. If these challenges are effectively addressed, the integration of big

data, machine learning, and theory-driven development of treatments holds significant promise for advancing our prediction and understanding of iCBT for depression (Rutledge et al., 2019).

6.3 Limitations and Future Research

6.3.1 Cross-Sectional vs. Intra-Individual

This thesis primarily relied on baseline patient data for prognostic profiling or single pre-post treatment comparisons. Cross-sectional data, commonly used in network research, is limiting in many ways. Indeed, it can offer insight into symptom associations at a group-level, such as in our study where we analysed responder and non-responder networks pre-post iCBT to study depression symptom changes. Cross-sectional networks in this sense, are useful for generating exploratory hypotheses about symptom relations, but they are purely correlational and cannot imply direct, causal effects between symptoms, or generalize to individual-level changes due to treatment (Epskamp et al., 2018b; Hamaker, 2012; Klipstein et al., 2021). This is because between-subject variations and within-subject variations are only equivalent under very stringent conditions and are highly unrealistic when we apply these conditions in psychological research (Molenaar & Campbell, 2009). Cross-sectional networks further preclude independent exploration of confounding variables like baseline severity and variance during network estimation. We attempted to account for them through parametric testing using independent subsamples of responders and non-responders, which helped discern how these variables contribute to connectivity differences between-groups. Within the remit of what cross-sectional data can do, however, we could not provide in-depth insights into the dynamic symptom relationships and treatment effects, and the role of variability in network connectivity during treatment.

For these reasons and more, the field is beginning to move past cross-sectional tests of network theory towards personalised/idiographic methods. These networks comprised longitudinal, within-person time-series data, such as ecological momentary assessment (EMA). This approach allows researchers to model the intricate, dynamical interactions between symptoms over time and assess Granger causality. This can illuminate how treatment affects the entire system of depression within an individual, and test if that impact is different in responders and non-responders (Burger et al., 2022; Jones & Robinaugh, 2021; Wright & Zimmermann, 2019). In this regard, some small-scale studies

like Levinson and colleagues' (2021) identified highly variable and person-dependent influential symptoms for treatment targets in eating disorders (N=34), while Hoffart and colleagues (2019) related specific therapeutic mechanisms in exposure therapy for PTSD to hypervigilance and physiological reactivation (N=65). A notable study in depression, conducted by Komulainen and colleagues (2020), pooled prospective data from 8 antidepressant trials (N=3559). The authors found that improvements specifically in depressed mood, insomnia, and suicidality had bigger impact on subsequent improvement in the rest of the depression symptoms (i.e., greater out-strength centrality) in the treatment arm compared to placebo arm.

Having said this, as the personalised network approach gains momentum, it becomes crucial to collect enough data points for everyone so to ensure reliability in estimated parameters (Epskamp, 2020). One way we can do this is to make the network meaningful for everyone so that they are motivated to engage in intense longitudinal data collection. Moving forward, future studies should adopt a transdiagnostic network approach that transcends beyond the boundaries of categorical psychiatric diagnoses (Roefs et al., 2023). This involves considering not only variables related to the primary diagnosis of depression, but also comorbid disorders (e.g., anxiety, OCD), and individual-specific contextual variables encompassing various life aspects such as treatments (Bekhuis et al., 2018; Blanken et al., 2019; Boschloo et al., 2019), life events (Cramer et al., 2012), or social activities (Wichers, 2014). Personalised networks offer a deep understanding of how these elements interact to maintain depressive presentation, thus enabling a higher degree of personalisation in tailoring treatments for each person (Roefs et al., 2023; Wichers et al., 2021). To achieve this, collaboration between patients and clinicians is vital when customising personalised networks. This collaborative process helps generate working hypotheses, choose the right intervention for targeting key symptoms, and develop a dynamic understanding of the patient's psychopathological profile (Burger et al., 2022; Fisher & Boswell, 2016).

6.3.2 Digital Behavioural Data & Phenotyping

We believe treatment prediction research in iCBT will truly be 'data-driven' once we successfully combine complementary data sources for prediction, reflecting the fact that treatment response likely comprise influences from multiple interacting biopsychosocial factors of the person (Iniesta et al., 2018; Wallert et al., 2022), as well as their interaction with their treatment (Pearson et al., 2019). Yet, practical clinical use of these models

necessitates cost-effective mass recruitment of patients, which rules out a lot of the complex biomarker measures like neuroimaging, hormones, and genetics (Gillan & Whelan, 2017). One unique advantage of digital interventions like iCBT is that they offer us automated, detailed, and quantifiable user engagement and behavioural data on their treatment platform, such as modules viewed, tool usage, program completion rates, and clinician interactions (Donkin et al., 2011). Studies consistently highlight a strong dose-response relationship; the more patients engage with iCBT, the more improvement they experience, which makes engagement a critical factor in determining outcomes (Donkin et al., 2011). Chien and colleagues (2020) found unique associations between 5 different engagement profiles and specific usage of essential iCBT components and outcomes. Those who improved in depression the most tended to complete key CBT components such as cognitive restructuring and behavioural activation within the initial 2 weeks, irrespective of time spent on the platform. Importantly, we can also use these data within predictive models to provide interpretable and actionable outputs. Pearson and colleagues (2019) identified specific usage of iCBT modules and therapist access as important predictors in a baseline machine learning model predicting depression level 8 weeks post-iCBT. Even though patient engagement is unknown prior to treatment, the authors posited the model may still be useful for predicting outcomes for new patients with varying levels of engagement.

Nonetheless, engagement in a digital health context remains a huge unknown. Attrition rates are notoriously poor; patients typically only complete 50-70% of the program (Christensen et al., 2009), and drop-out rates can range up to 50% in guided interventions and up to 74% when unguided (Christensen et al., 2009). However, some show that completing the entire program is not necessary for clinical benefits (Christensen & Mackinnon, 2006), and others argue that disengagement from iCBT is not entirely a bad thing. A qualitative study conducted interviews with iCBT dropouts revealed both negative (e.g., treatment doesn't work for them) and positive experiences (got the most out of treatment already) (Lawler et al., 2021). In the context of our study, we found a weak link between iCBT engagement metrics and depression improvement, which could be due to its short duration (4 weeks). A longer study duration aligned with the recommended dosage of iCBT (i.e., 6-8 weeks; Richards et al., 2014a; 2014b) might better capture the dose-response relationship. Furthermore, high treatment adherence rates in our study may have limited meaningful investigations into drop outs, albeit we

conducted sensitivity analyses excluding these individuals. Future research should delve deeper into engagement and adherence in digital interventions with a personalised approach, for instance, the optimal level of engagement for each individual, which would be of scientific and clinical importance for maximising therapeutic benefits (Moshe et al., 2021).

The potential of iCBT engagement data, however, is contingent on whether patients are active (or inactive) with their treatment, which may introduce systematic biases in sample selection, data completeness, and attrition. Relatedly, the field of treatment prediction is getting increasingly excited in passive data collection using digital devices such as smartphones and wearables, to assess patients' health-related behavioural patterns, a concept known as 'digital phenotyping' (Chekroud et al., 2021; Gillan & Rutledge, 2021). These devices can now reliably and automatically capture a wide range of sensor data, including physiological (e.g., heart rate variability, stress levels, and sleep quality), behavioural (app usage, social patterns, exercise, language use), and environmental information (geospatial location) (Moshe et al., 2021), which offer insights otherwise inaccessible via self-reports alone for treatment prognosis. For such high dimensional, information-dense data, data-driven prognostic models like machine learning are exceptionally well-suited for uncovering obscure patterns for tailoring prediction and personalisation efforts, such as rapid and long-term changes in symptom severity, which can be used to trigger the delivery of just-in-time adaptive interventions (i.e., the right type of support at the right time) (Garcia-Ceja et al., 2018; Nahum-Shani et al., 2017; Teepe et al., 2021). Collecting passive data this way is also very low burden; it requires minimal time and effort from participants, avoids active participation of data input, and greatly scales up the amount of intraindividual data collected (Rost et al., 2023). One example is illustrated by Kelley & Gillan (2022), whereby digital phenotyping of language via mining of social media data permitted them to gather sufficient time-series, intra-individual data spanning periods of depression to create personalised networks for each person (N=964), which has been a challenge in network research. All in all, digital phenotyping holds promise in personalising care delivery with clinically valuable predictions, broadens opportunities for tailored interventions, while furthering our understanding of mental illness (Chekroud et al., 2021; Gillan & Rutledge, 2021; Marsch et al., 2022). Future studies should assess whether including certain kinds of idiographic,

passive data improves our ability to predict iCBT treatment response beyond predictors collected in a single baseline session.

In the above context, both passive data and engagement data are not initially available at baseline for analyses. In the PIP study, our focus was on making predictions using baseline data only so that the model can be immediately applied before a patient starts iCBT. However, research in iCBT and psychotherapy suggests that incorporating symptom changes during the initial weeks of treatment, as in Adaptive Treatment Strategies, increases prediction accuracy (Bone et al., 2021; Forsell et al., 2020; Schibbye et al., 2014). In their recommendation on ML methodological choices in iCBT treatment prediction, Isacsson and colleagues (2023) tested 80 models with various combinations of baseline variables, weekly symptoms, and treatment activity (N=6695). They found that the best predictions for identifying patients at risk of not responding to iCBT involved treatment data up to week 4. By week 2, the model's accuracy exceeded a clinically useful benchmark (>67% accuracy), providing clinicians the ability to make personalised adjustments while treatment is ongoing. Having said that, treatment prediction work becomes less interesting and readily applicable when we need to consider a significant proportion of the treatment itself beforehand. In the PIP study, we highlight that the 'baseline' nature of the model is a strength in itself. It provides treatment recommendations before iCBT begins, reduces the risk of mid-treatment dropout, and avoids wasting valuable resources and unnecessary patient suffering. We believe these two approaches need not be in conflict; they can complement each other and contribute to the flourishing of data-driven precision in psychiatry (Wallert et al., 2022).

6.3.2 Experimental Control

A key limitation in this thesis is the use of observational study design throughout our work. This design follows a more naturalistic approach, observing patients who received iCBT based on routine clinical decisions. It reflects how iCBT is used in real-world settings and offers high ecological validity, but it may also introduce greater heterogeneity into the data (Rost et al., 2023; Webb et al., 2020). As researchers, we lacked direct control over certain study aspects, limiting our ability to make inferences about iCBT. Firstly, patients were not randomly assigned to treatment group in both our LTC subgroup analysis and the PIP study. Responders and non-responders in our network analysis were also individuals who were naturally referred to iCBT through their healthcare providers. The lack of randomisation in treatment assignment could have

introduced systematic differences between treatment groups, such as variations in sample characteristics and selection bias from referral. Furthermore, the lack of experimental control in observational studies makes it challenging to draw causal inferences about treatment mechanisms; in this regard, it is very possible that patients may be receiving additional mental health treatments alongside iCBT. For one, information about concurrent treatments was not available in our LTC subgroup and network analyses. We know this was true in the PIP study, where a good proportion of participants in both the iCBT and antidepressant arms reported taking concurrent treatments. Consequently, we could not entirely attribute our findings solely to the impact of iCBT.

Through experimental study designs, future research can move towards an innovative experimental paradigm that is mechanistically driven and component oriented so to better elucidate mechanisms of iCBT. This entails exploring the active ingredients and moderators of iCBT and how this may differ for different patients (Domhardt et al., 2021). Given the highly standardised nature of iCBT, it becomes possible to manipulate specific iCBT elements (e.g., treatment duration, self-reflection, therapeutic alliance, mood-monitoring) while controlling for others, to tackle domain-specific depression-relevant presentations such as rumination and distorted thinking. This helps us understand the causal mechanisms underlying change, by unfolding the precise steps and processes through which iCBT induces therapeutic action (Kazdin, 2007; Moshe et al., 2021). It also helps us to develop and identify treatment strategies that directly target these mechanisms while eliminating redundant ones (Furukawa et al., 2021).

A similar approach can also be applied to network studies. For instance, targeting highly central symptoms as initial treatment targets is a contested hypothesis, as the dynamics of psychopathological networks may function radically different in other network types (e.g., social networks) (Bringmann et al., 2019). In this sense, experimental studies that isolate and target central nodes of the network for each individual patient compared to standard treatment are required to determine their efficacy. Ongoing work in this area is actively exploring the potential of targeting central networks in personalised networks to inform case and treatment formulations in psychotherapy, such as PREMISE (Burger et al., 2022) and TheraNet (Hall et al., 2022). In these initiatives, clinicians in the intervention group receive an individualised network with centrality indices related to specific symptoms. As patients progress in treatment, these networks dynamically change and provide feedback to clinicians for integration into their care plan. The control

interventions in comparison would follow evidence-based techniques without the guidance of the personalised network. Interventions guided by the individual patient's dynamic symptom network are expected to enhance treatment success and decrease relapse, while shedding light on the temporal dynamics between symptoms that contribute to improvement during treatment (Roefs et al., 2023). Briefly, experimental study designs promote a more precise and personalised matching of intervention (components) to specific symptom presentations and vulnerabilities relevant to the individual (Gillan & Rutledge, 2021). To our knowledge, large-scale, controlled studies focusing on implementation have not been extended to investigate iCBT treatment response. This represents a promising research direction that we encourage others to explore.

6.4 Conclusion

In summary, this thesis sought to predict and understand treatment response to iCBT in depression. Through our work that collectively examined patients undergoing iCBT in real-world settings, we highlighted the advantages of a *big data* approach to advance precision in psychiatry. We first ascertained iCBT is effective for patients with chronic physical conditions in everyday clinical settings. Insights from this work comprised some of the first evidence from large-scale, real-world patient data to inform the prescription of iCBT relative to other low-intensity interventions for this cohort. Expanding on the big data approach, we tested the feasibility of a fully internet-based methodology for upscaling treatment prediction research. This method successfully helped us to gather rich, complex data from substantially larger sample sizes, with good data quality and measurable indicators of patient engagement and treatment adherence. Using this rich dataset in the PIP study, we developed and evaluated a multimodal machine learning algorithm aimed to predict early iCBT response, with modest performance that generalised well to antidepressant patients. Lastly, to conclude on an even bigger note, we investigated whether baseline cross-sectional networks of depression symptoms differentiated between iCBT responders and non-responders using the largest sample size in this research area to-date. We found that symptom connectivity differences at baseline between responders and non-responders were likely driven by differences in baseline depression variance. We also found that cross-sectional network metrics did not carry more prognostic value about iCBT treatment response than baseline measures that we already have. While big data holds promise, we need even bigger ideas to unlock the dynamic nature of depression. It is time to embrace innovative statistical methods and put

these ideas to the test, working together with clinicians and patients to achieve improved precision and personalisation in iCBT for every individual in the real world.

7 References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729/SUPPL_FILE/TPSR_A_1808729_SM1621.ZIP

Adhikary, D., Barman, S., & Ranjan, R. (2023). Internet-based cognitive behavioural therapy for individuals with depression and chronic health conditions: A systematic review. *Cureus*, *15*(4), e37822. https://doi.org/10.7759/cureus.37822

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *19*(6), 586–592. https://doi.org/10.1037//0278-6133.19.6.586

Al-Harbi, K. S. (2012). Treatment-resistant depression: Therapeutic trends, challenges, and future directions. *Patient Preference and Adherence*, *6*, 369–388. https://doi.org/10.2147/PPA.S29716

Altman, D. G., Lausen, B., Sauerbrei, W., & Schumacher, M. (1994). Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *JNCI: Journal of the National Cancer Institute*, *86*(11), 829–835. https://doi.org/10.1093/jnci/86.11.829

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Publishing.

Anderson, N., & Ozakinci, G. (2018). Effectiveness of psychological interventions to improve quality of life in people with long-term conditions: Rapid systematic review of randomised controlled trials. *BMC Psychology*, *6*(1). https://doi.org/10.1186/S40359-018-0225-4

Andersson, E., Crowley, J. J., Lindefors, N., Ljótsson, B., Hedman-Lagerlöf, E., Boberg, J., El Alaoui, S., Karlsson, R., Lu, Y., Mattheisen, M., Kähler, A. K., Svanborg, C., Mataix-Cols, D., Mattsson, S., Forsell, E., Kaldo, V., Schalling, M., Lavebratt, C., Sullivan, P. F., & Rück, C. (2019). Genetics of response to cognitive behavior therapy in adults with major depression: A preliminary report. *Molecular Psychiatry*, *24*(4), 484–490. https://doi.org/10.1038/s41380-018-0289-9

Andersson, G. (2016). Internet-delivered psychological treatments. *Annual Review of Clinical Psychology*, *12*(1), 157–179. https://doi.org/10.1146/annurev-clinpsy-021815-093006

Andersson, G., Carlbring, P., & Rozental, A. (2019a). Response and remission rates in internet-based cognitive behavior therapy: An individual patient data meta-analysis. *Frontiers in Psychiatry*, *10*, 749. https://doi.org/10.3389/FPSYT.2019.00749/BIBTEX

Andersson, G., Carlbring, P., Titov, N., & Lindefors, N. (2019b). Internet interventions for adults with anxiety and mood disorders: A narrative umbrella review of recent meta-analyses. *Canadian Journal of Psychiatry*, *64*(7), 465–470. https://doi.org/10.1177/0706743719839381

Andersson, G., & Hedman, E. (2013). Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. *Verhaltenstherapie*, *23*(3), 140–148. https://doi.org/10.1159/000354779

Andersson, G., & Titov, N. (2014). Advantages and limitations of Internet-based interventions for common mental disorders. *World Psychiatry*, *13*(1), 4–11. https://doi.org/10.1002/wps.20083

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019c). Internet-delivered psychological treatments: From innovation to implementation. *World Psychiatry*, *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

Armey, M. F., Schatten, H. T., Haradhvala, N., & Miller, I. W. (2015). Ecological momentary assessment (EMA) of depression-related phenomena. *Current Opinion in Psychology*, *4*, 21–25. https://doi.org/10.1016/j.copsyc.2015.01.002

Baguley, C., Farrand, P., Hope, R., Leibowitz, J., Lovell, K., Lucock, M., O'Neill, C., Pxton, R., Pilling, S., Richardson, D., Turpin, G., White, J., & Williams, C. (2010). *Good practice guidance on the use of self-help materials within IAPT services* (Issue March, p. 39).

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

Bauldry, S. (2015). Variation in the protective effect of higher education against depression*. *Society and Mental Health*, *5*(2), 145–161. https://doi.org/10.1177/2156869314564399

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research*, *21*(9), e14567.

Baxter, S., Johnson, M., Chambers, D., Sutton, A., Goyder, E., & Booth, A. (2018). The effects of integrated care: A systematic review of UK and international evidence. *BMC Health Services Research*, *18*(1), 350. https://doi.org/10.1186/S12913-018-3161-3

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

Beatty, L., & Binnion, C. (2016). A systematic review of predictors of, and reasons for, adherence to online psychological interventions. *International Journal of Behavioral Medicine*, *23*(6), 776–794. https://doi.org/10.1007/s12529-016-9556-9

Beatty, L., & Lambert, S. (2013). A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. *Clinical Psychology Review*, *33*(4), 609–622. https://doi.org/10.1016/J.CPR.2013.03.004

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience 2007 10:9*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

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

Belanger, H. G., Lee, C., Poliacoff, Z., Gupta, C. T., & Winsberg, M. (2023). Early response to antidepressant medications in adults with major depressive disorder. *Journal of Clinical Psychopharmacology*, *43*(1), 46–54. https://doi.org/10.1097/JCP.0000000000001638

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Berlim, M. T., Richard-Devantoy, S., Dos Santos, N. R., & Turecki, G. (2020). The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: A randomized clinical trial. *Psychological Medicine*, 1–15. https://doi.org/10.1017/S0033291720001002

Blanco, I., Contreras, A., Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2020). Positive interventions in depression change the structure of well-being and psychological symptoms: A network analysis. *The Journal of Positive Psychology*, *15*(5), 623-628.

Blanken, T. F., Van Der Zweerde, T., Van Straten, A., Van Someren, E. J. W., Borsboom, D., & Lancee, J. (2019). Introducing network intervention analysis to investigate sequential, symptom-specific treatment effects: A demonstration in co-occurring insomnia and depression. *Psychother Psychosom*, *88*(1), 52–54. https://doi.org/10.1159/000495045

Blazer, D. G. (2010). Protection from late life depression. *International Psychogeriatrics*, *22*(2), 171–173. https://doi.org/10.1017/S1041610209990895

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

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503

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

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of 'evidence that psychopathology symptom networks have limited replicability' by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, *126*(7), 989–999. https://doi.org/10.1037/abn0000306

Bos, E. H., & De Jonge, P. (2014). “Critical slowing down in depression” is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, 111(10), E878–E878. https://doi.org/10.1073/pnas.1323672111

Bos, F. M., Fried, E. I., Hollon, S. D., Bringmann, L. F., Dimidjian, S., DeRubeis, R. J., & Bockting, C. L. H. (2018). Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *Social Psychiatry and Psychiatric Epidemiology*, *53*(6), 617–627. https://doi.org/10.1007/s00127-018-1506-1

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., Jarrett, R. B., Kennedy, S. H., Miranda, J., Mohr, D. C., Simons, A. D., Parker, G., Petrak, F., Herpertz, S., Quilty, L. C., … Cuijpers, P. (2019). The symptom-specific efficacy of antidepressant medication vs. Cognitive behavioral therapy in the treatment of depression: Results from an individual patient data meta-analysis. *World Psychiatry*, *18*(2), 183–191. https://doi.org/10.1002/wps.20630

Boschloo, L., van Borkulo, C. D., Rhemtulla, M., Keyes, K. M., Borsboom, D., & Schoevers, R. A. (2015). The Network Structure of Symptoms of the Diagnostic and Statistical Manual of Mental Disorders. *PLOS ONE*, *10*(9), e0137621. https://doi.org/10.1371/journal.pone.0137621

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brener, N. D., McManus, T., Galuska, D. A., Lowry, R., & Wechsler, H. (2003). Reliability and validity of self-reported height and weight among high school students. *Journal of Adolescent Health*, *32*(4), 281–287. https://doi.org/10.1016/S1054-139X(02)00708-5

Bringmann, L. F. (2019). Person-specific networks in psychopathology: Past, present and future. Science of the Total Environment, 135907. https://doi.org/10.1016/j.copsyc.2021.03.004

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., Epskamp, S., Eronen, M. I., Hamaker, E., Kuppens, P., Lutz, W., McNally, R. J., Molenaar, P., Tio, P., Voelkle, M. C., & Wichers, M. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy*, *149*, 104011. https://doi.org/10.1016/j.brat.2021.104011

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

Bringmann, L. F., Lemmens, L. H., Huibers, M. J., Borsboom, D., & Tuerlinckx, F. J. P. M. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, *45*(4), 747-757.

Browne, J. P., van der Meulen, J. H., Lewsey, J. D., Lamping, D. L., & Black, N. (2010). Mathematical coupling may account for the association between baseline severity and minimally important difference values. *Journal of Clinical Epidemiology*, 63(8), 865–874. https://doi.org/10.1016/j.jclinepi.2009.10.004

Browning, M., Behrens, T. E., Jocham, G., O’Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Browning, M., Bilderbeck, A. C., Dias, R., Dourish, C. T., Kingslake, J., Deckert, J., Goodwin, G. M., Gorwood, P., Guo, B., Harmer, C. J., Morriss, R., Reif, A., Ruhe, H. G., van Schaik, A., Simon, J., Sola, V. P., Veltman, D. J., Elices, M., Lever, A. G., … Dawson, G. R. (2021). The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PReDicT): An open-label, randomised controlled trial. *Neuropsychopharmacology*, 1–8. https://doi.org/10.1038/s41386-021-00981-z

Buckman, J. E. J., Cohen, Z. D., O’Driscoll, C., Fried, E. I., Saunders, R., Ambler, G., DeRubeis, R. J., Gilbody, S., Hollon, S. D., Kendrick, T., Watkins, E., Eley, T. C., Peel, A. J., Rayner, C., Kessler, D., Wiles, N., Lewis, G., & Pilling, S. (2021). Predicting prognosis for adults with depression using individual symptom data: A comparison of modelling approaches. *Psychological Medicine*, *53*(2), 1–11. https://doi.org/10.1017/S0033291721001616

Burger, J., Epskamp, S., van der Veen, D. C., Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2022). A clinical PREMISE for personalized models: Toward a formal integration of case formulations and statistical networks. *Journal of Psychopathology and Clinical Science*, *131*(8), 906–916. https://doi.org/10.1037/abn0000779

Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., Fried, E. I., Borsboom, D., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods*. https://doi.org/10.1037/met0000471

Burnette, C. B., Luzier, J. L., Bennett, B. L., Weisenmuller, C. M., Kerr, P., Martin, S., Keener, J., & Calderwood, L. (2022). Concerns and recommendations for using Amazon MTurk for eating

disorder research. *The International Journal of Eating Disorders*, *55*(2), 263–272. https://doi.org/10.1002/eat.23614
Button, K. S., Wiles, N. J., Lewis, G., Peters, T. J., & Kessler, D. (2012). Factors associated with differential response to online cognitive behavioural therapy. *Social Psychiatry and Psychiatric Epidemiology*, *47*(5), 827–833. https://doi.org/10.1007/s00127-011-0389-1
Chandler, J., Sisso, I., & Shapiro, D. (2020). Participant carelessness and fraud: Consequences for clinical research and potential solutions. *Journal of Abnormal Psychology*, *129*(1), 49–55. https://doi.org/10.1037/abn0000479
Charova, E., Dorstyn, D., Tully, P., & Mittag, O. (2015). Web-based interventions for comorbid depression and chronic illness: A systematic review. *Journal of Telemedicine and Telecare*, *21*(4), 189–201. https://doi.org/10.1177/1357633X15571997
Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170. https://doi.org/10.1002/wps.20882
Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X
Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759–771. https://doi.org/10.1093/biomet/asn034
Chen, C., Wang, X., Xu, H., & Li, Y. (2023). Effectiveness of digital psychological interventions in reducing perinatal depression: A systematic review of meta-analyses. *Archives of Women's Mental Health*, *26*(4), 423–439. https://doi.org/10.1007/s00737-023-01327-y
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
Chien, I., Enrique, A., Palacios, J., Regan, T., Keegan, D., Carter, D., Tschiatschek, S., Nori, A., Thieme, A., Richards, D., Doherty, G., & Belgrave, D. (2020). A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Network Open*, *3*(7), e2010791. https://doi.org/10.1001/jamanetworkopen.2020.10791
Chmielewski, M., & Kucker, S. C. (2020). An Mturk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149
Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression: Systematic review. *Journal of Medical Internet Research*, *11*(2), e1194. https://doi.org/10.2196/jmir.1194
Christensen, H., & Mackinnon, A. (2006). The law of attrition revisited. *Journal of Medical Internet Research*, *8*(3), e20. https://doi.org/10.2196/jmir.8.3.e20
Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J. P. T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J. P. A., & 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. *Lancet*, *391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7
Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, *14*(1), 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833
Cohen, A. K., Nussbaum, J., Weintraub, M. L. R., Nichols, C. R., & Yen, I. H. (2020). Association of adult depression with educational attainment, aspirations, and expectations. *Preventing Chronic Disease*, *17*, E94. https://doi.org/10.5888/pcd17.200098
Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. https://doi.org/10.2307/2136404
Cohen, Z. D., & DeRubeis, R. J. (2018). Annual Review of Clinical Psychology Treatment Selection in Depression. *Annu. Rev. Clin. Psychol*, *14*, 209–245. https://doi.org/10.1146/annurev-clinpsy-050817
Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, *88*(2), 71–83. https://doi.org/10.1159/000497425

Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.-M., Spiers, H., & Hornberger, M. (2019). Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *116*(19), 9285–9292. https://doi.org/10.1073/pnas.1901600116

Cramer, A. O. J., Borsboom, D., Aggen, S. H., & Kendler, K. S. (2012). The pathoplasticity of dysphoric episodes: Differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological Medicine*, *42*(5), 957–965. https://doi.org/10.1017/S003329171100211X

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE*, *11*(12), e0167490. https://doi.org/10.1371/journal.pone.0167490

Cramer, A. O. J., Waldorp, L. J., Van Der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150. https://doi.org/10.1017/S0140525X09991567

Crown, W. H., Finkelstein, S., Berndt, E. R., Ling, D., Poret, A. W., Rush, A. J., & Russell, J. M. (2002). The impact of treatment-resistant depression on health care utilization and costs. *The Journal of Clinical Psychiatry*, *63*(11), 963–971. https://doi.org/10.4088/JCP.V63N1102

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Cuijpers, P., Karyotaki, E., de Wit, L., Ebert, D.D., (2020a). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research*. 30, 279–293.

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

Cuijpers, P., Karyotaki, E., Eckshtain, D., Ng, M. Y., Corteselli, K. A., Noma, H., Quero, S., & Weisz, J. R. (2020b). Psychotherapy for depression across different age groups: A systematic review and meta-analysis. *JAMA Psychiatry*, *77*(7), 694–702. https://doi.org/10.1001/jamapsychiatry.2020.0164

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Papola, D., Ebert, D., & Karyotaki, E. (2023). Psychological treatment of depression: A systematic overview of a 'Meta-Analytic Research Domain'. *Journal of Affective Disorders*, *335*, 141–151. https://doi.org/10.1016/j.jad.2023.05.011

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. https://doi.org/10.1001/jamapsychiatry.2019.0268

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020c). 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 (WPA)*, *19*(1), 92–107. https://doi.org/10.1002/wps.20701

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021b). 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 (WPA)*, *20*(2), 283–293. https://doi.org/10.1002/wps.20860

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

Cuijpers, P., van Straten, A., & Andersson, G. (2008). Internet-administered cognitive behavior therapy for health problems: A systematic review. *Journal of Behavioral Medicine*, *31*(2), 169–177. https://doi.org/10.1007/s10865-007-9144-1

Curtiss, J. E., Wallace, B., Fisher, L. B., Nyer, M., Jain, F., Cusin, C., & Pedrelli, P. (2021). Change processes in cognitive behavioral therapy and motivational interviewing for depression and heavy alcohol use: A network approach. *Journal of Affective Disorders Reports*, *3*, 100040. https://doi.org/10.1016/j.jadr.2020.100040

Dai, Q., & Smith, G. D. (2023). Resilience to depression: Implication for psychological vaccination. *Frontiers in Psychiatry*, *14*, 1071859.

Das-Munshi, J., Stewart, R., Ismail, K., Bebbington, P. E., Jenkins, R., & Prince, M. J. (2007). Diabetes, common mental disorders, and disability: Findings from the UK National Psychiatric Morbidity Survey. *Psychosomatic Medicine*, *69*(6), 543–550. https://doi.org/10.1097/PSY.0b013e3180cc3062

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

De Vos, S., Wardenaar, K. J., Bos, E. H., Wit, E. C., Bouwmans, M. E. J., & de Jonge, P. (2017). An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLOS ONE*, *12*(6), e0178586. https://doi.org/10.1371/journal.pone.0178586

Dear, B. F., Zou, J. B., Ali, S., Lorian, C. N., Johnston, L., Sheehan, J., Staples, L. G., Gandy, M., Fogliati, V. J., Klein, B., & Titov, N. (2015). Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety: A randomized controlled trial. *Behavior Therapy*, *46*(2), 206–217. https://doi.org/10.1016/j.beth.2014.09.007

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. *Psychological Science*, *27*(6), 848–858. https://doi.org/10.1177/0956797616639301

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z. D., DeRubeis, R. J., & Barkham, M. (2022). Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry*, *79*(2), 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539

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

Delgadillo, J., Overend, K. J., Lucock, M., & Mcmillan, D. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Article in Behaviour Research and Therapy*. https://doi.org/10.1016/j.brat.2017.09.011

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0083875

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy vs. medications for depression: Treatment outcomes and neural mechanisms. *Nature Reviews. Neuroscience*, *9*(10), 788–796. https://doi.org/10.1038/nrn2345

Dickens, C., Katon, W., Blakemore, A., Khara, A., McGowan, L., Tomenson, B., Jackson, J., Walker, L., & Guthrie, E. (2012). Does depression predict the use of urgent and unscheduled care by people with long term conditions? A systematic review with meta-analysis. *Journal of Psychosomatic Research*, *73*(5), 334–342. https://doi.org/10.1016/J.JPSYCHORES.2012.08.018

DiMatteo, M. R., Lepper, H. S., & Croghan, T. W. (2000). Depression is a risk factor for noncompliance with medical treatment: Meta-analysis of the effects of anxiety and depression on patient adherence. *Archives of Internal Medicine*, *160*(14), 2101–2107. https://doi.org/10.1001/ARCHINTE.160.14.2101

Dinga, R., Marquand, A. F., Veltman, D. J., Beekman, A. T. F., Schoevers, R. A., van Hemert, A. M., Penninx, B. W. J. H., & Schmaal, L. (2018). Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: A machine learning approach. *Translational Psychiatry*. https://doi.org/10.1038/s41398-018-0289-1

Dobson, K. S., & Dozois, D. J. A. (2021). *Handbook of Cognitive-Behavioral Therapies, Fourth Edition*. Guilford Publications.

Domhardt, M., Steubl, L., Boettcher, J., Buntrock, C., Karyotaki, E., Ebert, D. D., Cuijpers, P., & Baumeister, H. (2021). Mediators and mechanisms of change in internet- and mobile-based interventions for depression: A systematic review. *Clinical Psychology Review*, *83*, 101953. https://doi.org/10.1016/j.cpr.2020.101953

Donegan, K. R., Brown, V., Price, R. B., Pringle, A., Gallagher, E., Hanlon, A., & Gillan, C. (2023). *Using smartphones to optimise and scale-up the assessment of model-based planning.* PsyArXiv. https://doi.org/10.31234/osf.io/hpm4s

Donegan, K. R., & Gillan, C. M. (2022). New principles and new paths needed for online research in mental health: Commentary on Burnette et al. (2021). *International Journal of Eating Disorders*, *55*(2), 278–281. https://doi.org/10.1002/eat.23670

Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of Medical Internet Research*, *13*(3), e1772. https://doi.org/10.2196/jmir.1772

Dossa, A., Glickman, M. E., & Berlowitz, D. (2011). Association between mental health conditions and rehospitalization, mortality, and functional outcomes in patients with stroke following inpatient rehabilitation. *BMC Health Services Research*, *11*(1), 1–10. https://doi.org/10.1186/1472-6963-11-311/TABLES/6

Dunlop, B. W., Binder, E. B., Cubells, J. F., Goodman, M. M., Kelley, M. E., Kinkead, B., Kutner, M., Nemeroff, C. B., Newport, D. J., Owens, M. J., Pace, T. W. W., Ritchie, J. C., Rivera, V. A., Westen, D., Craighead, W. E., & Mayberg, H. S. (2012). Predictors of remission in depression to individual and combined treatments (PReDICT): Study protocol for a randomized controlled trial. *Trials*, *13*, 1–18. https://doi.org/10.1186/1745-6215-13-106

Dunlop, B. W., Cha, J., Choi, K. S., Rajendra, J. K., Nemeroff, C. B., Craighead, W. E., & Mayberg, H. S. (2023). Shared and unique changes in brain connectivity among depressed patients after remission with pharmacotherapy versus psychotherapy. *American Journal of Psychiatry*, *180*(3), 218–229. https://doi.org/10.1176/appi.ajp.21070727

Edmonds, M., Hadjistavropoulos, H. D., Schneider, L. H., Dear, B. F., & Titov, N. (2018). Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? Predictors of symptom change and dropout. *Journal of Anxiety Disorders*, *54*, 24–32. https://doi.org/10.1016/J.JANXDIS.2018.01.003

Egede, L. E., & Ellis, C. (2010). Diabetes and depression: Global perspectives. *Diabetes Research and Clinical Practice*, *87*(3), 302–312. https://doi.org/10.1016/j.diabres.2010.01.024

El Alaoui, S., Ljótsson, B., Hedman, E., Svanborg, C., Kaldo, V., & Lindefors, N. (2016). Predicting outcome in internet-based cognitive behaviour therapy for major depression: A large cohort study of adult patients in routine psychiatric care. *PLOS ONE*, *11*(9), e0161191. https://doi.org/10.1371/journal.pone.0161191

Elliott, H., Jones, P. J., & Schmidt, U. (2020). Central symptoms predict posttreatment outcomes and clinical impairment in anorexia nervosa: A network analysis. *Clinical Psychological Science*, *8*(1), 139–154. https://doi.org/10.1177/2167702619865958

Enrique, A., Mooney, O., Salamanca-Sanabria, A., Lee, C. T., Farrell, S., & Richards, D. (2019a). Assessing the efficacy and acceptability of an internet-delivered intervention for resilience among college students: A pilot randomised control trial protocol. *Internet Interventions*, *17*. https://doi.org/10.1016/j.invent.2019.100254

Enrique, A., Palacios, J. E., Ryan, H., & Richards, D. (2019). Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: Secondary analysis of data from a randomized controlled trial. *J Med Internet Res 2019;21(8):E12775 Https://Www.Jmir.Org/2019/8/E12775*, *21*(8), e12775. https://doi.org/10.2196/12775

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3

Epskamp, S., Borsboom, D., & Fried, E. I. (2018a). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(1), 1–18. https://doi.org/10.18637/jss.v048.i04

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018b). Personalized Network Modeling in Psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science : A Journal of the Association for Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018c). The Gaussian Graphical Model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Ermers, N. J., Hagoort, K., & Scheepers, F. E. (2020). The predictive validity of machine learning models in the classification and treatment of major depressive disorder: State of the art and future directions. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/FPSYT.2020.00472/FULL

Esfahlani, F. Z., Sayama, H., Visser, K. F., & Strauss, G. P. (2017). Sensitivity of the Positive and Negative Syndrome Scale (PANSS) in detecting treatment effects via network analysis. *Innovations in Clinical Neuroscience*, *14*(11), 59–67.

Fiest, K. M., Walker, J. R., Bernstein, C. N., Graff, L. A., Zarychanski, R., Abou-Setta, A. M., Patten, S. B., Sareen, J., Bolton, J. M., Marriott, J. J., Fisk, J. D., Singer, A., & Marrie, R. A. (2016). Systematic review and meta-analysis of interventions for depression and anxiety in persons with multiple

sclerosis. *Multiple Sclerosis and Related Disorders*, *5*, 12–26. https://doi.org/10.1016/J.MSARD.2015.10.004
Fischer, R., Bortolini, T., Karl, J. A., Zilberberg, M., Robinson, K., Rabelo, A., Gemal, L., Wegerhoff, D., Nguyễn, T. B. T., Irving, B., Chrystal, M., & Mattos, P. (2020). Rapid review and meta-meta-analysis of self-guided interventions to address anxiety, depression, and stress during COVID-19 social distancing. *Frontiers in Psychology*, *11*, 563876. https://doi.org/10.3389/FPSYG.2020.563876/FULL
Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment*, *23*(4), 496–506. https://doi.org/10.1177/1073191116638735
Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115
Fisher, L., Hessler, D. M., Polonsky, W. H., & Mullan, J. (2012). When is diabetes distress clinically meaningful?: Establishing cut points for the Diabetes Distress Scale. *Diabetes Care*, *35*(2), 259–264. https://doi.org/10.2337/DC11-1572
Fleming, T., Bavin, L., Lucassen, M., Stasiak, K., Hopkins, S., & Merry, S. (2018). Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *Journal of Medical Internet Research*, *20*(6), e199. https://doi.org/10.2196/jmir.9275
Flygare, O., Enander, J., Andersson, E., Ljótsson, B., Ivanov, V. Z., Mataix-Cols, D., & Rück, C. (2020). Predictors of remission from body dysmorphic disorder after internet-delivered cognitive behavior therapy: A machine learning approach. *BMC Psychiatry*, *20*(1), 247. https://doi.org/10.1186/s12888-020-02655-4
Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485–496. https://doi.org/10.1037/1040-3590.14.4.485
Fonseka, T. M., MacQueen, G. M., & Kennedy, S. H. (2018). Neuroimaging biomarkers as predictors of treatment outcome in Major Depressive Disorder. *Journal of Affective Disorders*, *233*, 21–35. https://doi.org/10.1016/j.jad.2017.10.049
Forand, N. R., Feinberg, J. E., Barnett, J. G., & Strunk, D. R. (2019). Guided internet CBT versus "gold standard" depression treatments: An individual patient analysis. *Journal of Clinical Psychology*, *75*(4), 581–593. https://doi.org/10.1002/jclp.22733
Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., Twose, M., & Dras, M. (2023). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S0033291723002544
Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969–988. https://doi.org/10.1037/abn0000276
Forsell, E., Isacsson, N., Blom, K., Jernelöv, S., Ben Abdesslem, F., Lindefors, N., Boman, M., & Kaldo, V. (2020). Predicting treatment failure in regular care Internet-Delivered Cognitive Behavior Therapy for depression and anxiety using only weekly symptom measures. *Journal of Consulting and Clinical Psychology*, *88*(4), 311–321. https://doi.org/10.1037/ccp0000462
Forsell, E., Jernelöv, S., Blom, K., & Kaldo, V. (2022). Clinically sufficient classification accuracy and key predictors of treatment failure in a randomized controlled trial of Internet-delivered Cognitive Behavior Therapy for Insomnia. *Internet Interventions*, *29*, 100554. https://doi.org/10.1016/j.invent.2022.100554
Forstmeier, W., Wagenmakers, E., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10.1111/brv.12315
Freeman, A., Tyrovolas, S., Koyanagi, A., Chatterji, S., Leonardi, M., Ayuso-Mateos, J. L., Tobiasz-Adamczyk, B., Koskinen, S., Rummel-Kluge, C., & Haro, J. M. (2016). The role of socio-economic status in depression: Results from the COURAGE (aging survey in Europe). *BMC Public Health*, *16*, 1098. https://doi.org/10.1186/s12889-016-3638-0
Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016a). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS ONE*, *9*(2), 90311. https://doi.org/10.1371/journal.pone.0090311

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

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

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

Friedman, E. S., Davis, L. L., Zisook, S., Wisniewski, S. R., Trivedi, M. H., Fava, M., Rush, A. J., & Team, C. S. (2012). Baseline depression severity as a predictor of single and combination antidepressant treatment outcome: Results from the CO-MED Trial. *European Neuropsychopharmacology*, *22*(3), 183. https://doi.org/10.1016/j.euroneuro.2011.07.010

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22. https://doi.org/10.18637/jss.v033.i01

Fu, C. H. Y., Steiner, H., & Costafreda, S. G. (2013). Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiology of Disease*, *52*, 75–83. https://doi.org/10.1016/j.nbd.2012.05.008

Furukawa, T. A., Efthimiou, O., Weitz, E. S., Cipriani, A., Keller, M. B., Kocsis, J. H., Klein, D. N., Michalak, J., Salanti, G., Cuijpers, P., & Schramm, E. (2018). Cognitive-behavioral analysis system of psychotherapy, drug, or their combination for persistent depressive disorder: Personalizing the treatment choice using individual participant data network metaregression. *Psychotherapy and Psychosomatics*, *87*(3), 140–153. https://doi.org/10.1159/000489227

Furukawa, T. A., Suganuma, A., Ostinelli, E. G., Andersson, G., Beevers, C. G., Shumake, J., Berger, T., Boele, F. W., Buntrock, C., Carlbring, P., Choi, I., Christensen, H., Mackinnon, A., Dahne, J., Huibers, M. J. H., Ebert, D. D., Farrer, L., Forand, N. R., Strunk, D. R., … Cuijpers, P. (2021). Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: A systematic review and component network meta-analysis using individual participant data. *The Lancet. Psychiatry*, *8*(6), 500–511. https://doi.org/10.1016/S2215-0366(21)00077-8

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

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1–26. https://doi.org/10.1016/j.pmcj.2018.09.003

Garner, D. M., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*(4), 871–878. https://doi.org/10.1017/S0033291700049163

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Geschwind, N., van Teffelen, M., Hammarberg, E., Arntz, A., Huibers, M. J. H., & Renner, F. (2021). Impact of measurement frequency on self-reported depressive symptoms: An experimental study in a clinical setting. *Journal of Affective Disorders Reports*, *5*, 100168. https://doi.org/10.1016/J.JADR.2021.100168

Gillan, C. M., & Daw, N. D. (2016). Taking psychiatry research online. *Neuron*, *91*(1), 19–23. https://doi.org/10.1016/J.NEURON.2016.06.002

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goaldirected control. *eLife*, *5*(MARCH2016). https://doi.org/10.7554/eLife.11305

Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive–compulsive disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130475. https://doi.org/10.1098/RSTB.2013.0475

Gillan, C. M., & Rutledge, R. B. (2021). Smartphones and the neuroscience of mental health. *Annual Review of Neuroscience*, *44*(1), annurev-neuro-101220-014053. https://doi.org/10.1146/annurev-neuro-101220-014053

Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences*, *18*, 34–42. https://doi.org/10.1016/j.cobeha.2017.07.003

Godlewska, B. R., & Harmer, C. J. (2021). Cognitive neuropsychological theory of antidepressant action: A modern-day approach to depression and its treatment. *Psychopharmacology*, *238*(5), 1265–1278. https://doi.org/10.1007/s00213-019-05448-0

Goldstein-Piekarski, A. N., Staveland, B. R., Ball, T. M., Yesavage, J., Korgaonkar, M. S., & Williams, L. M. (2018). Intrinsic functional connectivity predicts remission on antidepressants: A randomized controlled trial to identify clinically applicable imaging biomarkers. *Translational Psychiatry*, *8*(1), Article 1. https://doi.org/10.1038/s41398-018-0100-3

Gómez Penedo, J. M., Schwartz, B., Giesemann, J., Rubel, J. A., Deisenhofer, A.-K., & Lutz, W. (2022). For whom should psychotherapy focus on problem coping? A machine learning algorithm for treatment personalization. *Psychotherapy Research*, *32*(2), 151–164. https://doi.org/10.1080/10503307.2021.1930242

Gonda, X., Pompili, M., Serafini, G., Carvalho, A. F., Rihmer, Z., & Dome, P. (2015). The role of cognitive dysfunction in the symptoms and remission from depression. *Annals of General Psychiatry*, *14*, 27. https://doi.org/10.1186/s12991-015-0068-9

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., & Kessler, R. C. (2021). The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *PharmacoEconomics 2021 39:6*, *39*(6), 653–665. https://doi.org/10.1007/S40273-021-01019-4

Greenwell, B., M., & Boehmke, B., C. (2020). Variable importance plots—An introduction to the vip package. *The R Journal*, *12*(1), 343. https://doi.org/10.32614/RJ-2020-013

Guthrie, E. A., Dickens, C., Blakemore, A., Watson, J., Chew-Graham, C., Lovell, K., Afzal, C., Kapur, N., & Tomenson, B. (2016). Depression predicts future emergency hospital admissions in primary care patients with chronic physical illness. *Journal of Psychosomatic Research*, *82*, 54–61. https://doi.org/10.1016/J.JPSYCHORES.2014.10.002

Gutiérrez-Rojas, L., Porras-Segovia, A., Dunne, H., Andrade-González, N., & Cervilla, J. A. (2020). Prevalence and correlates of major depressive disorder: A systematic review. *Brazilian Journal of Psychiatry*, *42*(6), 657–672. https://doi.org/10.1590/1516-4446-2019-0650

Hadjistavropoulos, H. D., Pugh, N. E., Hesser, H., & Andersson, G. (2016). Predicting response to therapist-assisted internet-delivered cognitive behavior therapy for depression or anxiety within an open dissemination trial. *Behavior Therapy*, *47*(2), 155–165. https://doi.org/10.1016/j.beth.2015.10.006

Hagan, K. E., Matheson, B. E., Datta, N., L'Insalata, A. M., Onipede, Z. A., Gorrell, S., Mondal, S., Bohon, C. M., Le Grange, D., & Lock, J. D. (2021). Understanding outcomes in family-based treatment for adolescent anorexia nervosa: A network approach. *Psychological Medicine*, 1–12. https://doi.org/10.1017/S0033291721001604

Halford, J., & Brown, T. (2009). Cognitive-behavioural therapy as an adjunctive treatment in chronic physical illness. *Advances in Psychiatric Treatment*, *15*, 306–317. https://doi.org/10.1192/apt.bp.107.003731

Hall, M., Wagner, A. A., Scherner, P., Michael, K. L., Lawyer, G., Lutz, W., & Rubel, J. (2022). *Using* Personalized assessment and network model feedback in psychotherapy: Proof of principle for the TheraNet project. *PsyArXiv*. https://doi.org/10.31234/osf.io/8deyj

Hamaker, E. (2012). Why researchers should think 'within-person': A paradigmatic rationale. In *Handbook of Research Methods For Studying Daily Life* (pp. 43–61).

Harmer, C. J., Duman, R. S., & Cowen, P. J. (2017). How do antidepressants work? New perspectives for refining future treatment approaches. *The Lancet Psychiatry*, *4*(5), 409–418. https://doi.org/10.1016/S2215-0366(17)30015-9

Harrison, P., Walton, S., Fennema, D., Duan, S., Jaeckle, T., Goldsmith, K., Carr, E., Ashworth, M., Young, A. H., & Zahn, R. (2021). Development and validation of the Maudsley Modified Patient Health Questionnaire (MM-PHQ-9). *BJPsych Open*, *7*(4), e123. https://doi.org/10.1192/bjo.2021.953

Haslam, N. (2003). Categorical versus dimensional models of mental disorder: The taxometric evidence. *The Australian and New Zealand Journal of Psychiatry*, *37*(6), 696–704. https://doi.org/10.1080/j.1440-1614.2003.01258.x

Hawgood, S., Hook-Barnard, I. G., O'Brien, T. C., & Yamamoto, K. R. (2015). Precision medicine: Beyond the inflection point. *Science Translational Medicine*, *7*(300). https://doi.org/10.1126/SCITRANSLMED.AAA9970

Herzog, D. P., Beckmann, H., Lieb, K., Ryu, S., & Müller, M. B. (2018). Understanding and predicting antidepressant response: Using animal models to move toward precision psychiatry. *Frontiers in Psychiatry*, *9*, 512. https://doi.org/10.3389/fpsyt.2018.00512

Hicks, O., McInerney, S. J., Lam, R. W., Milev, R. V., Frey, B. N., Soares, C. N., Foster, J. A., Rotzinger, S., Kennedy, S. H., & Harkness, K. L. (2022). Acute and chronic stress predict anti-depressant treatment outcome and naturalistic course of major depression: A CAN-BIND report. *Journal of Affective Disorders*, *313*, 8–14. https://doi.org/10.1016/j.jad.2022.06.058

Hind, D., O'Cathain, A., Cooper, C. L., Parry, G. D., Isaac, C. L., Rose, A., ... & Sharrack, B. (2010). The acceptability of computerised cognitive behavioural therapy for the treatment of depression in people with chronic physical disease: a qualitative study of people with multiple sclerosis. *Psychology and Health*, *25*(6), 699-712.

Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*, 1–28. https://doi.org/10.18637/jss.v042.i08

Hoang, U., Goldacre, M. J., & Stewart, R. (2013). Avoidable mortality in people with schizophrenia or bipolar disorder in England. *Acta Psychiatrica Scandinavica*, *127*(3), 195–201. https://doi.org/10.1111/acps.12045

Hodgson, K., Mufti, S. J., Uher, R., & McGuffin, P. (2012). Genome-wide approaches to antidepressant treatment: Working towards understanding and predicting response. *Genome Medicine*, *4*(6), 52. https://doi.org/10.1186/gm351

Hoffart, A., Langkaas, T. F., Øktedalen, T., & Johnson, S. U. (2019). The temporal dynamics of symptoms during exposure therapies of PTSD: a network approach. *European Journal of Psychotraumatology*, *10*(1), 1618134. https://doi.org/10.1080/20008198.2019.1618134

Hofmann, S. G. (2012). *An Introduction to Modern CBT: Psychological Solutions to Mental Health Problems* (pp. xvii, 218). Wiley-Blackwell.

Hollon, S. D. (2019). Moderation, mediation, and moderated mediation. *World Psychiatry*, *18*(3), 288–289. https://doi.org/10.1002/wps.20665

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, *11*(2), 213–218. https://doi.org/10.1016/0022-3999(67)90010-4

Hornstein, S., Forman-Hoffman, V., Nazander, A., Ranta, K., & Hilbert, K. (2021). Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach: *Https://Doi.Org/10.1177/20552076211060659*, *7*. https://doi.org/10.1177/20552076211060659

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the Personalized Advantage Index approach. *PloS One*, *10*(11), e0140771. https://doi.org/10.1371/journal.pone.0140771

Iniesta, R., Hodgson, K., Stahl, D., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., Dobson, R., Aitchison, K. J., Farmer, A., McGuffin, P., Lewis, C. M., & Uher, R. (2018). Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Scientific Reports 2018 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-23584-z

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., Stahl, D., Dobson, R., Aitchison, K. J., Farmer, A., Lewis, C. M., McGuffin, P., & Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2016.03.016

Isacsson, Nils Hentati, V., Abdesslem, F. B., Forsell, E., Boman, Magnus, & Kaldo, Viktor. (2023). Machine learning predictions of outcome in internet-based cognitive behavioral therapy: Methodological choices and clinical usefulness. *Research Square*.

Jackson, M., Jones, D., Dyson, J., & Macleod, U. (2019). Facilitated group work for people with long-term conditions: A systematic review of benefits from studies of group-work interventions. *British Journal of General Practice*, *69*(682), e363–e372. https://doi.org/10.3399/BJGP19X702233

Jacobs, M., Pradier, M. F., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: The example of the antidepressant selection. *Translational Psychiatry*, *11*(1), 108. https://doi.org/10.1038/s41398-021-01224-x

Jacobson, N. C., & Nemesure, M. D. (2021). Using artificial intelligence to predict change in depression and anxiety symptoms in a digital intervention: Evidence from a transdiagnostic randomized controlled trial. *Psychiatry Research*, *295*, 113618. https://doi.org/10.1016/j.psychres.2020.113618

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

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

Jankowsky, K., Krakau, L., Schroeders, U., Zwerenz, R., & Beutel, M. E. (2022). Predicting treatment response using machine learning: A registered report. *PsyArXiv*. https://doi.org/10.31234/osf.io/fuyjv

Jones, P. J., & Robinaugh, D. R. (2021). An answer to "So What?" Implications of network theory for research and practice. *FOCUS*, *19*(2), 204–210. https://doi.org/10.1176/appi.focus.20200050

Jorm, A. F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. *Psychological Medicine*, *30*(1), 11–22. https://doi.org/10.1017/S0033291799001452

Kang, S.-G., & Cho, S.-E. (2020). Neuroimaging biomarkers for predicting treatment response and recurrence of major depressive disorder. *International Journal of Molecular Sciences*, *21*(6), 2148. https://doi.org/10.3390/ijms21062148

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. G., Furukawa, T. A., Cuijpers, P., Riper, H., Patel, V., Mira, A., Gemmil, A. W., Yeung, A. S., Lange, A., Williams, A. D., Mackinnon, A., Geraedts, A., Van Straten, A., Meyer, B., Björkelund, C., Knaevelsrud, C., … 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. https://doi.org/10.1001/JAMAPSYCHIATRY.2020.4364

Karyotaki, E., Kemmeren, L., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Mackinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J. P., Schröder, J., Bretón-López, J., Scheider, J., Griffiths, K., Farrer, L., … Cuijpers, P. (2018). Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. *Psychological Medicine*, *48*(15), 2456–2466. https://doi.org/10.1017/S0033291718000648

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., MacKinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J. P., Schröder, J., Bretón-López, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M. J. H., … Cuijpers, P. (2017). Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms a meta-analysis of individual participant data. *JAMA Psychiatry*, *74*(4), 351–359. https://doi.org/10.1001/jamapsychiatry.2017.0044

Katon, W. J., Lin, E. H. B., Von Korff, M., Ciechanowski, P., Ludman, E. J., Young, B., Peterson, D., Rutter, C. M., McGregor, M., & McCulloch, D. (2010). Collaborative Care for Patients with Depression and Chronic Illnesses. *New England Journal of Medicine*, *363*(27), 2611–2620. https://doi.org/10.1056/NEJMoa1003955

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *6*(1), 21–37. https://doi.org/10.1177/1745691610393527

Kelley, S. W., Fisher, A. J., Lee, C. T., Gallagher, E., Hanlon, A. K., Robertson, I., & Gillan, C. M. (2023). Elevated emotion network connectivity leaves people vulnerable to fluctuations in depression. *PsyArxiv*.

Kelley, S. W., & Gillan, C. M. (2022). Using language in social media posts to study the network dynamics of depression longitudinally. *Nature Communications 2022 13:1*, *13*(1), 1–11. https://doi.org/10.1038/s41467-022-28513-3

Kessler, R. C. (2018). The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Current Opinion in Psychiatry*, *31*(1), 32–39. https://doi.org/10.1097/YCO.0000000000000377

Kessler, R. C., Davis, C. G., & Kendler, K. S. (1997). Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychological Medicine*, *27*(5), 1101–1119. https://doi.org/10.1017/s0033291797005588

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, *21*(3), 169–184. https://doi.org/10.1002/mpr.1359

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., de Jonge, P., Nierenberg, A. A., Rosellini, A. J., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., & 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

Klipstein, L. von, Borsboom, D., & Arntz, A. (2021). The exploratory value of cross-sectional partial correlation networks: Predicting relationships between change trajectories in borderline personality disorder. *PLOS ONE*, *16*(7), e0254496. https://doi.org/10.1371/journal.pone.0254496

Koenen, K. C., Moffitt, T. E., Roberts, A. L., Martin, L. T., Kubzansky, L., Harrington, H., Poulton, R., & Caspi, A. (2009). Childhood IQ and adult mental disorders: A test of the cognitive reserve hypothesis. *American Journal of Psychiatry*, *166*(1), 50–57. https://doi.org/10.1176/appi.ajp.2008.08030343

Komulainen, K., Airaksinen, J., Savelieva, K., Gluschkoff, K., García Velázquez, R., Elovainio, M., & Jokela, M. (2020). Network dynamics of depressive symptoms in antidepressant medication treatment: Secondary analysis of eight clinical trials. *Molecular Psychiatry*, 1–8. https://doi.org/10.1038/s41380-020-00884-3

Kostev, K., Rex, J., Eith, T., & Heilmaier, C. (2014). Which adverse effects influence the dropout rate in selective serotonin reuptake inhibitor (SSRI) treatment? Results for 50,824 patients. *GMS German Medical Science*, *12*, Doc15. https://doi.org/10.3205/000200

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., & Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: A machine learning approach. *The Lancet. Psychiatry*, *3*(10), 935–946. https://doi.org/10.1016/S2215-0366(16)30171-7

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (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. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, *64*(2), 258–266. https://doi.org/10.1097/00006842-200203000-00008

Kuhn, M. (2008). Building predictive models in *R* using the **caret** package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kumar, V., Sattar, Y., Bseiso, A., Khan, S., & Rutkofsky, I. H. (2017). The effectiveness of internet-based cognitive behavioral therapy in treatment of psychiatric disorders. *Cureus*, *9*(8). https://doi.org/10.7759/CUREUS.1626

Lakey, B., & Cronin, A. (2008). Low social support and major depression: research, theory and methodological issues. In K. S. Dobson & D. J. A. Dozois (Eds.), *Risk Factors in Depression* (pp. 385–408). Elsevier. https://doi.org/10.1016/B978-0-08-045078-0.00017-4

Lam, R. W., Kennedy, S. H., McIntyre, R. S., & Khullar, A. (2014). Cognitive dysfunction in major depressive disorder: Effects on psychosocial functioning and implications for treatment. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *59*(12), 649–654.

Lawler, K., Earley, C., Timulak, L., Enrique, A., & Richards, D. (2021). Dropout from an internet-delivered cognitive behavioral therapy intervention for adults with depression and anxiety: Qualitative study. *JMIR Formative Research*, *5*(11), e26221. https://doi.org/10.2196/26221

Lazar, M. A., Pan, Z., Ragguett, R.-M., Lee, Y., Subramaniapillai, M., Mansur, R. B., Rodrigues, N., & McIntyre, R. S. (2017). Digital revolution in depression: A technologies update for clinicians. *Personalized Medicine in Psychiatry*, *4–6*, 1–6. https://doi.org/10.1016/j.pmip.2017.09.001

Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., Claus, N., Swords, L., O'Keane, V., Stephan, K. E., & Gillan, C. M. (2023). The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry*, *23*(1), 25. https://doi.org/10.1186/s12888-022-04462-5

Lee, E. W., Denison, F. C., Hor, K., & Reynolds, R. M. (2016). Web-based interventions for prevention and treatment of perinatal mood disorders: A systematic review. *BMC Pregnancy and Childbirth*, *16*(1), 38. https://doi.org/10.1186/s12884-016-0831-1

Lee, Y., Ragguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M., Chan, T. C. Y., Fus, D., Park, C., Musial, N., Zuckerman, H., Chen, V. C. H., Ho, R., Rong, C., & McIntyre, R. S. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. *Journal of Affective Disorders*, *241*, 519–532. https://doi.org/10.1016/j.jad.2018.08.073

Lenhard, F., Sauer, S., Andersson, E., Månsson, K. N. T., Mataix-Cols, D., Rück, C., & Serlachius, E. (2018). Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *International Journal of Methods in Psychiatric Research*, *27*(1). https://doi.org/10.1002/mpr.1576

Lépine, J.-P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment*, *7*(Suppl 1), 3. https://doi.org/10.2147/NDT.S19617

Levinson, C. A., Hunt, R. A., Keshishian, A. C., Brown, M. L., Vanzhula, I., Christian, C., Brosof, L. C., & Williams, B. M. (2021). Using individual networks to identify treatment targets for eating disorder treatment: A proof-of-concept study and initial data. *Journal of Eating Disorders*, *9*(1), 147. https://doi.org/10.1186/S40337-021-00504-7/FIGURES/2

Lewis, M. J., Spiliopoulou, A., Goldmann, K., Pitzalis, C., McKeigue, P., & Barnes, M. R. (2023). nestedcv: An R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. *Bioinformatics Advances*, *3*(1), vbad048. https://doi.org/10.1093/bioadv/vbad048

Li, C., Xu, D., Hu, M., Tan, Y., Zhang, P., Li, G., & Chen, L. (2017). A systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy for patients with diabetes and depression. *Journal of Psychosomatic Research*, *95*, 44–54. https://doi.org/10.1016/J.JPSYCHORES.2017.02.006

Li, F., Jörg, F., Merkx, M. J. M., & Feenstra, T. (2023). Early symptom change contributes to the outcome prediction of cognitive behavioral therapy for depression patients: A machine learning approach. *Journal of Affective Disorders*, *334*, 352–357. https://doi.org/10.1016/j.jad.2023.04.111

Liebowitz, M. R. (1987). *Social Phobia*. *22*, 141–173. https://doi.org/10.1159/000414022

Lindner, P., Carlbring, P., Flodman, E., Hebert, A., Poysti, S., Hagkvist, F., Johansson, R., Zetterqvist Westin, V., Berger, T., & Andersson, G. (2016). Does cognitive flexibility predict treatment gains in Internet-delivered psychological treatment of social anxiety disorder, depression, or tinnitus? *PeerJ*, *4*, e1934. https://doi.org/10.7717/peerj.1934

Linn, B. S., Linn, M. W., & Gurel, L. (1968). Cumulative Illness Rating Scale. *Journal of the American Geriatrics Society*, *16*(5), 622–626. https://doi.org/10.1111/j.1532-5415.1968.tb02103.x

Linn, R. L. (1968). Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin*, *69*(1), 69–73. https://doi.org/10.1037/h0025263

Liu, R. T. (2017). Childhood adversities and depression in adulthood: Current findings and future directions. *Clinical Psychology: Science and Practice*, *24*(2), 140–153. https://doi.org/10.1111/cpsp.12190

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*, *157*(2), 98–112. https://doi.org/10.1093/aje/kwf182

Lorig, K. R., & Holman, H. R. (2003). Self-management education: History, definition, outcomes, and mechanisms. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, *26*(1), 1–7. https://doi.org/10.1207/S15324796ABM2601_01

Lorimer, B., Delgadillo, J., Kellett, S., Brown, G., Lorimer, B., Delgadillo, J., Kellett, S., & Brown, G. (2020). Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research*, *30*(5), 650–661.

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science*, *7*(3), 445–461. https://doi.org/10.1177/2167702618815466

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: A validation study of the DSM-IV symptomatic criteria. *Psychological Medicine*, *40*(10), 1679–1690. https://doi.org/10.1017/S0033291709992157

MacGillivray, S., Arroll, B., Hatcher, S., Ogston, S., Reid, I., Sullivan, F., Williams, B., & Crombie, I. (2003). Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: Systematic review and meta-analysis. *BMJ (Clinical Research Ed.)*, *326*(7397), 1014. https://doi.org/10.1136/bmj.326.7397.1014

Maciukiewicz, M., Marshe, V. S., Hauschild, A.-C., Foster, J. A., Rotzinger, S., Kennedy, J. L., Kennedy, S. H., Müller, D. J., & Geraci, J. (2018). GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *Journal of Psychiatric Research*, *99*, 62–68. https://doi.org/10.1016/j.jpsychires.2017.12.009

Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., Demyttenaere, K., McIntyre, R. S., Widiger, T., & Wittchen, H. (2020). The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry*, *19*(3), 269–293. https://doi.org/10.1002/wps.20771

Månsson, K. N. T., Frick, A., Boraxbekk, C. J., Marquand, A. F., Williams, S. C. R., Carlbring, P., Andersson, G., & Furmark, T. (2015). Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Translational Psychiatry*. https://doi.org/10.1038/tp.2015.22

Marathe, P. H., Gao, H. X., & Close, K. L. (2017). American diabetes association standards of medical care in diabetes 2017. *Journal of Diabetes*, *9*(4), 320–324. https://doi.org/10.1111/1753-0407.12524

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Rose Donohue, M., Foran, W., Miller, R. L., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Chen, J., Newbold, D. J., Zheng, A., Seider, N. A., Van, A. N., Laumann, T. O., Thompson, W. K., … Dosenbach, N. (2020). Towards reproducible brain-wide association studies affiliations. *bioRxiv*, *11*, 15–18. https://doi.org/10.1101/2020.08.21.257758

Mariani, N., Cattane, N., Pariante, C., & Cattaneo, A. (2021). Gene expression studies in Depression development and treatment: An overview of the underlying molecular mechanisms and biological processes to identify biomarkers. *Translational Psychiatry*, *11*(1), Article 1. https://doi.org/10.1038/s41398-021-01469-6

Marin, R. S., Biedrzycki, R. C., & Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Research*, *38*(2), 143–162. https://doi.org/10.1016/0165-1781(91)90040-V

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. https://doi.org/10.1037/a0023678

Marsch, L. A., Chen, C.-H., Adams, S. R., Asyyed, A., Does, M. B., Hassanpour, S., Hichborn, E., Jackson-Morris, M., Jacobson, N. C., Jones, H. K., Kotz, D., Lambert-Harris, C. A., Li, Z., McLeman, B., Mishra, V., Stanger, C., Subramaniam, G., Wu, W., & Campbell, C. I. (2022). The feasibility and utility of harnessing digital health to understand clinical trajectories in medication treatment for opioid use disorder: D-TECT study design and methodological considerations. *Frontiers in Psychiatry*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyt.2022.871916

Marwood, L., Wise, T., Perkins, A. M., & Cleare, A. J. (2018). Meta-analyses of the neural mechanisms and predictors of response to psychotherapy in depression and anxiety. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2018.09.022

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research*, *78*(2–3), 293–296. https://doi.org/10.1016/J.SCHRES.2005.06.020

Mazure, C. M., Bruce, M. L., Maciejewski, P. K., & Jacobs, S. C. (2000). Adverse life events and cognitive-personality characteristics in the prediction of major depression and antidepressant response. *The American Journal of Psychiatry*, *157*(6), 896–903. https://doi.org/10.1176/appi.ajp.157.6.896

McElroy, E., Napoleone, E., Wolpert, M., & Patalay, P. (2019). Structure and connectivity of depressive symptom networks corresponding to early treatment response. *EClinicalMedicine*, *8*, 29–36. https://doi.org/10.1016/j.eclinm.2019.02.009

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. *The American Psychologist*, *65*(2), 73–84. https://doi.org/10.1037/a0018121

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

McMahon, F. J. (2014). Prediction of treatment outcomes in psychiatry—Where do we stand ? *Dialogues in Clinical Neuroscience*, *16*(4), 455. https://doi.org/10.31887/DCNS.2014.16.4/FMCMAHON

McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). Mental health and wellbeing in England: Adult Psychiatric Morbidity survey 2014. In *Leeds: NHS Digital*.
McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6515–6518. https://doi.org/10.1073/pnas.1504162112
McNally, R. J. (2020). Network analysis of psychopathology: Controversies and challenges. *Annual Review of Clinical Psychology*, 1–58.
McNamara, M. E., Zisser, M., Beevers, C. G., & Shumake, J. (2022). Not just "big" data: Importance of sample size, measurement error, and uninformative predictors for developing prognostic models for digital interventions. *Behaviour Research and Therapy*, *153*, 104086. https://doi.org/10.1016/J.BRAT.2022.104086
Mehta, S., Peynenburg, V. A., & Hadjistavropoulos, H. D. (2018). Internet-delivered cognitive behaviour therapy for chronic health conditions: A systematic review and meta-analysis. *Journal of Behavioral Medicine*, 1–19. https://doi.org/10.1007/s10865-018-9984-x
Meijer, A., Conradi, H. J., Bos, E. H., Thombs, B. D., van Melle, J. P., & de Jonge, P. (2011). Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: A meta-analysis of 25 years of research. *General Hospital Psychiatry*, *33*(3), 203–216. https://doi.org/10.1016/j.genhosppsych.2011.02.007
Mewton, L., Sachdev, P. S., & Andrews, G. (2013). A naturalistic study of the acceptability and effectiveness of internet-delivered cognitive behavioural therapy for psychiatric disorders in older Australians. *PLoS ONE*, *8*(8), e71825. https://doi.org/10.1371/journal.pone.0071825
Miloseva, L., & Richter, K. (2022). Impact of neurostimulation and psychotherapy on mental health. *Frontiers in Psychology*, *13*, 992634.
Mogoașe, C., Cobeanu, O., David, O., Giosan, C., & Szentagotai, A. (2017). Internet-based psychotherapy for adult depression: What about the mechanisms of change? *Journal of Clinical Psychology*, *73*(1), 5–64. https://doi.org/10.1002/jclp.22326
Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x
Moshe, I., Terhorst, Y., Philippi, P., Domhardt, M., Cuijpers, P., Cristea, I., Pulkki-Råback, L., Baumeister, H., & Sander, L. B. (2021). Digital interventions for the treatment of depression: A meta-analytic review. *Psychological Bulletin*, *147*(8), 749–786. https://doi.org/10.1037/bul0000334
Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in health: Results from the World Health Surveys. *Lancet*, *370*(9590), 851–858. https://doi.org/10.1016/S0140-6736(07)61415-9
Mullarkey, M. C., & Schleider, J. L. (2021). Embracing scientific humility and complexity: Learning "What Works for Whom" in youth psychotherapy research. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *50*(4), 443–449. https://doi.org/10.1080/15374416.2021.1929252
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. *British Journal of Psychiatry*, *180*(5), 461–464. https://doi.org/10.1192/bjp.180.5.461
Muñoz, R. F., Lenert, L. L., Delucchi, K., Stoddard, J., Perez, J. E., Penilla, C., & Pérez-Stable, E. J. (2006). Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *8*(1), 77–87. https://doi.org/10.1080/14622200500431940
Murray, E. (2008). Internet-delivered treatments for long-term conditions: Strategies, efficiency and cost-effectiveness. *Expert Review of Pharmacoeconomics and Outcomes Research*. https://doi.org/10.1586/14737167.8.3.261
Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2017). Just-in-Time Adaptive Interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8
National Collaborating Centre for Mental Health. (2018). *The Improving Access to Psychological Therapies (IAPT) Pathway for People with Long-Term Physical Health Conditions and Medically Unexplained Symptoms*. https://www.england.nhs.uk/publication/the-improving-access-to-

psychological-therapies-iapt-pathway-for-people-with-long-term-physical-health-conditions-and-medically-unexplained-symptoms/
Naylor, C., Parsonage, M., McDaid, D., Knapp, M., Fossey, M., & Galea, A. (2012). *Long-term conditions and mental health: The cost of co-morbidities*. 32.
Nelson, J. C., Mazure, C. M., Jatlow, P. I., Bowers, M. B., & Price, L. H. (2004). Combining norepinephrine and serotonin reuptake inhibition mechanisms for treatment of depression: A double-blind, randomized study. *Biological Psychiatry*, *55*(3), 296–300. https://doi.org/10.1016/j.biopsych.2003.08.007
NICE (2009). Depression in adults with a chronic physical health problem: Recognition and management. Clinical Guideline [CG91]; *National Institute for Health and Care Excellence*; London, U.K: NICE. https://www.nice.org.uk/guidance/cg91
NICE (2011). Generalised anxiety disorder and panic disorder in adults: Management. *National Institute for Health and Care Excellence*; London, U.K: NICE. https://www.nice.org.uk/guidance/cg113
NICE. (2020). Treating depression in adults Your care. *National Institute for Health and Care Excellenc Excellence*; London, U.K: NICE
NICE (2022). Depression in Adults: Treatment and Management (NG222). *London, U.K: National Institute for Care and Excellence;* London, U.K: NICE.
Nierenberg, A. A., Farabaugh, A. H., Alpert, J. E., Gordon, J., Worthington, J. J., Rosenbaum, J. F., & Fava, M. (2000). Timing of onset of antidepressant response with fluoxetine treatment. *Https://Doi.Org/10.1176/Appi.Ajp.157.9.1423*, *157*(9), 1423–1428. https://doi.org/10.1176/APPI.AJP.157.9.1423
Norbury, A., Hauser, T. U., Fleming, S. M., Dolan, R. J., & Huys, Q. J. (2023). Different components of cognitive-behavioural therapy affect specific cognitive mechanisms [Preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/ydct5
Nouretdinov, I., Costafreda, S. G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., & Fu, C. H. Y. (2011). Machine learning classification with confidence: Application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. *NeuroImage*, *56*(2), 809–813. https://doi.org/10.1016/j.neuroimage.2010.05.023
O'Driscoll, C., Buckman, J. E. J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., DeRubeis, R. J., Gilbody, S., Hollon, S. D., Kendrick, T., Kessler, D., Lewis, G., Watkins, E., Wiles, N., & Pilling, S. (2021). The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: Analysis of data from six randomised control trials. *BMC Medicine*, *19*(1), 109. https://doi.org/10.1186/s12916-021-01971-0
Oldham, P. D. (1962). A note on the analysis of repeated measurements of the same subjects. *Journal of Chronic Diseases*, *15*(10), 969–977. https://doi.org/10.1016/0021-9681(62)90116-9
Ophir, Y., Sisso, I., Asterhan, C. S. C., Tikochinski, R., & Reichart, R. (2020). The Turker Blues: Hidden factors behind increased depression rates among Amazon's Mechanical Turkers. *Clinical Psychological Science*, *8*(1), 65–83. https://doi.org/10.1177/2167702619865973
Ormel, J., Hollon, S. D., Kessler, R. C., Cuijpers, P., & Monroe, S. M. (2022). More treatment but no less depression: The treatment-prevalence paradox. *Clinical Psychology Review*, *91*, 102111. https://doi.org/10.1016/j.cpr.2021.102111
Ould Brahim, L., Lambert, S. D., Feeley, N., Coumoundouros, C., Schaffler, J., McCusker, J., Moodie, E. E. M., Kayser, J., Kolne, K., Belzile, E., & Genest, C. (2021). The effects of self-management interventions on depressive symptoms in adults with chronic physical disease(s) experiencing depressive symptomatology: A systematic review and meta-analysis. *BMC Psychiatry*, *21*(1), 1–35. https://doi.org/10.1186/S12888-021-03504-8/TABLES/5
Pabst, A., Löbner, M., Stein, J., Luppa, M., Kersting, A., König, H.-H., & Riedel-Heller, S. G. (2020). Internet-based cognitive behavior therapy only for the young? A secondary analysis of a randomized controlled trial of depression treatment. *Frontiers in Psychiatry*, *11*, 735. https://doi.org/10.3389/fpsyt.2020.00735
Palacios, J., Adegoke, A., Wogan, R., Duffy, D., Earley, C., Eilert, N., ... & Richards, D. (2023). Comparison of outcomes across low-intensity psychological interventions for depression and anxiety within a stepped-care setting: A naturalistic cohort study using propensity score modelling. *British Journal of Psychology*, *114*(2), 299-314.
Palacios, J. E., Enrique, A., Mooney, O., Farrell, S., Earley, C., Duffy, D., Eilert, N., Harty, S., Timulak, L., & Richards, D. (2022). Durability of treatment effects following internet-delivered cognitive behavioural therapy for depression and anxiety delivered within a routine care setting. *Clinical Psychology & Psychotherapy*. https://doi.org/10.1002/cpp.2743

Papini, S., Rubin, M., Telch, M. J., Smits, J. A., & Hien, D. A. (2020). Pretreatment posttraumatic stress disorder symptom network metrics predict the strength of the association between node change and network change during treatment. *Journal of Traumatic Stress*, *33*(1), 64-71.

Patel, V., Maj, M., Flisher, A. J., De Silva, M. J., Koschorke, M., Prince, M., & WPA Zonal and Member Society Representatives. (2010). Reducing the treatment gap for mental disorders: A WPA survey. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *9*(3), 169–176. https://doi.org/10.1002/j.2051-5545.2010.tb00305.x

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015a). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*(2), 292–300. https://doi.org/10.1177/2167702614540645

Pearson, R., Pisner, D., Meyer, B., Shumake, J., & Beevers, C. G. (2019). A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychological Medicine*, *49*(14), 2330–2341. https://doi.org/10.1017/S003329171800315X

Pennebaker, J. W., & Susman, J. R. (1988). Disclosure of traumas and psychosomatic processes. *Social Science and Medicine*, *26*(3), 327–332. https://doi.org/10.1016/0277-9536(88)90397-8

Perlis, R. H. (2016). Abandoning personalization to get to precision in the pharmacotherapy of depression. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *15*(3), 228–235. https://doi.org/10.1002/WPS.20345

Perlman, K., Benrimoh, D., Israel, S., Rollins, C., Brown, E., Tunteng, J. F., You, R., You, E., Tanguay-Sela, M., Snook, E., Miresco, M., & Berlim, M. T. (2019). A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *Journal of Affective Disorders*, *243*, 503–515. https://doi.org/10.1016/J.JAD.2018.09.067

Piazza, J. R., Charles, S. T., & Almeida, D. M. (2007). Living with chronic health conditions: Age differences in affective well-being. *The Journals of Gerontology: Series B*, *62*(6), P313–P321. https://doi.org/10.1093/GERONB/62.6.P313

Pittenger, C., & Bloch, M. H. (2014). Pharmacological treatment of obsessive-compulsive disorder. *The Psychiatric Clinics of North America*, *37*(3), 375. https://doi.org/10.1016/J.PSC.2014.05.006

Rajpurkar, P., Yang, J., Dass, N., Vale, V., Keller, A. S., Irvin, J., Taylor, Z., Basu, S., Ng, A., & Williams, L. M. (2020). Evaluation of a machine learning model based on pretreatment symptoms and electroencephalographic features to predict outcomes of antidepressant treatment in adults with depression: A prespecified secondary analysis of a randomized clinical trial. *JAMA Network Open*, *3*(6), e206653. https://doi.org/10.1001/jamanetworkopen.2020.6653

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior*, *58*, 354–360. https://doi.org/10.1016/J.CHB.2015.12.049

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Reavell, J., Hopkinson, M., Clarkesmith, D., & Lane, D. A. (2018). Effectiveness of cognitive behavioral therapy for depression and anxiety in patients with cardiovascular disease: A systematic review and meta-analysis. *Psychosomatic Medicine*, *80*(8), 742–753. https://doi.org/10.1097/PSY.0000000000000626

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, *170*(1), 59–70. https://doi.org/10.1176/APPI.AJP.2012.12070999/ASSET/IMAGES/LARGE/59F1.JPEG

Richards, D., Enrique, A., Eilert, N., Franklin, M., Palacios, J., Duffy, D., Earley, C., Chapman, J., Jell, G., Sollesse, S., & Timulak, L. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-0293-8

Richards, D., Timulak, L., Doherty, G., Sharry, J., Colla, A., Joyce, C., & Hayes, C. (2014). Internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression: Protocol for a randomized controlled trial. *BMC Psychiatry*, *14*(1), 147. https://doi.org/10.1186/1471-244X-14-147

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008-2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, *125*(6), 747–757.

Rodebaugh, T. L., Tonge, N. A., Piccirillo, M. L., Fried, E., Horenstein, A., Morrison, A. S., Goldin, P., Gross, J. J., Lim, M. H., Fernandez, K. C., Blanco, C., Schneier, F. R., Bogdan, R., Thompson, R. J., & Heimberg, R. G. (2018). Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder? *Journal of Consulting and Clinical Psychology*, *86*(10), 831–844.

Roefs, A., Fried, E. I., Kindt, M., Martijn, C., Elzinga, B., Evers, A. W. M., Wiers, R. W., Borsboom, D., & Jansen, A. (2022). A new science of mental disorders: Using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. *Behaviour Research and Therapy*, *153*, 104096. https://doi.org/10.1016/J.BRAT.2022.104096

Roiser, J. P., Elliott, R., & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology*, *37*(1), Article 1. https://doi.org/10.1038/npp.2011.183

Rost, N., Binder, E. B., & Brückl, T. M. (2023). Predicting treatment outcome in depression: An introduction into current concepts and challenges. *European Archives of Psychiatry and Clinical Neuroscience 2023*, *1*, 1–15. https://doi.org/10.1007/S00406-022-01418-4

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric *Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance*. *Biological Psychiatry*, *84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*(1), 127–141. https://doi.org/10.1002/sim.2331

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., Niederehe, G., Thase, M. E., Lavori, P. W., Lebowitz, B. D., McGrath, P. J., Rosenbaum, J. F., Sackeim, H. A., Kupfer, D. J., Luther, J., & Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *The American Journal of Psychiatry*, *163*(11), 1905–1917. https://doi.org/10.1176/AJP.2006.163.11.1905

Rutjes, A. W. S., Reitsma, J. B., Di Nisio, M., Smidt, N., van Rijn, J. C., & Bossuyt, P. M. M. (2006). Evidence of bias and variation in diagnostic accuracy studies. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *174*(4), 469–476. https://doi.org/10.1503/cmaj.050090

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: Toward clinical applications. *Current Opinion in Neurobiology*, *55*, 152–159. https://doi.org/10.1016/j.conb.2019.02.006

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., Parikh, S. V., Foster, J. A., Turecki, G., Müller, D. J., Strother, S. C., Farzan, F., Kennedy, S. H., & Uher, R. (2021). Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291721003871

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. https://doi.org/10.1037/bul0000102

Saraceno, B., van Ommeren, M., Batniji, R., Cohen, A., Gureje, O., Mahoney, J., Sridhar, D., & Underhill, C. (2007). Barriers to improvement of mental health services in low-income and middle-income countries. *Lancet (London, England)*, *370*(9593), 1164–1174. https://doi.org/10.1016/S0140-6736(07)61263-X

Saragoussi, D., Touya, M., Haro, J. M., Jönsson, B., Knapp, M., Botrel, B., Florea, I., Loft, H., & Rive, B. (2017). Factors associated with failure to achieve remission and with relapse after remission in patients with major depressive disorder in the PERFORM study. *Neuropsychiatric Disease and Treatment*, *13*, 2151–2165. https://doi.org/10.2147/NDT.S136343

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-ii. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/J.1360-0443.1993.TB02093.X

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: Scarcity, inequity, and inefficiency. *The Lancet*, *370*(9590), 878–889. https://doi.org/10.1016/S0140-6736(07)61239-2

Schibbye, P., Ghaderi, A., Ljótsson, B., Hedman, E., Lindefors, N., Rück, C., & Kaldo, V. (2014). Using early change to predict outcome in cognitive behaviour therapy: Exploring timeframe, calculation method, and differences of disorder-specific versus general measures. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0100614

Schleider, J. L., Dobias, M., Sung, J., Mumper, E., & Mullarkey, M. C. (2020). Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. *JMIR Mental Health*, *7*(6), e20513. https://doi.org/10.2196/20513

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology & Laboratory Medicine*, *137*(4), 558–565. https://doi.org/10.5858/arpa.2012-0198-RA

Schumacher, L., Burger, J., Echterhoff, J., & Kriston, L. (2022). Methodological and statistical practices of using symptom networks to evaluate mental health interventions: A systematic review. *PsyArxiv*. https://doi.org/10.31234/osf.io/ea7fu

Schweren, L., Van Borkulo, C. D., Fried, E., & Goodyer, I. M. (2018). Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry*, *75*(1), 98–100. https://doi.org/10.1001/jamapsychiatry.2017.3561

Scodari, B. T., Chacko, S., Matsumura, R., & Jacobson, N. C. (2023). Using machine learning to forecast symptom changes among subclinical depression patients receiving stepped care or usual care. *Journal of Affective Disorders*, *340*, 213–220. https://doi.org/10.1016/j.jad.2023.08.004

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., O'Connell, R., & Gillan, C. M. (2021). Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *Journal of Neuroscience*, *41*(30), 6539–6550. https://doi.org/10.1523/JNEUROSCI.0031-21.2021

Seow, T. X. F., & Gillan, C. M. (2020). Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. *Scientific Reports*, *10*(1), 2883. https://doi.org/10.1038/s41598-020-59646-4

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*, *1*(2), 213-220. https://doi.org/10.1177/2167702612469015

Shim, R. S., Baltrus, P., Ye, J., & Rust, G. (2011). Prevalence, treatment, and control of depressive symptoms in the United States: Results from the National Health and Nutrition Examination Survey (NHANES), 2005-2008. *Journal of the American Board of Family Medicine: JABFM*, *24*(1), 33–38. https://doi.org/10.3122/jabfm.2011.01.100121

Silk, T. J., Malpas, C. B., Beare, R., Efron, D., Anderson, V., Hazell, P., Jongeling, B., Nicholson, J. M., & Sciberras, E. (2019). A network analysis approach to ADHD symptoms: More than the sum of its parts. *PLOS ONE*, *14*(1), e0211053. https://doi.org/10.1371/journal.pone.0211053

Silfvernagel, K., Carlbring, P., Kabo, J., Edström, S., Eriksson, J., Månson, L., & Andersson, G. (2012). Individually tailored internet-based treatment for young adults and adults with panic attacks: Randomized controlled trial. *Journal of Medical Internet Research*, *14*(3), e65. https://doi.org/10.2196/jmir.1853

Silfvernagel, K., Westlinder, A., Andersson, S., Bergman, K., Diaz Hernandez, R., Fallhagen, L., Lundqvist, I., Masri, N., Viberg, L., Forsberg, M.-L., Lind, M., Berger, T., Carlbring, P., & Andersson, G. (2018). Individually tailored internet-based cognitive behaviour therapy for older adults with anxiety and depression: A randomised controlled trial. *Cognitive Behaviour Therapy*, *47*(4), 286–300. https://doi.org/10.1080/16506073.2017.1388276

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: Can we match patients with treatments? *The American Journal of Psychiatry*, *167*(12), 1445–1455. https://doi.org/10.1176/APPI.AJP.2010.09111680

Smits, J. A. J., Minhajuddin, A., Thase, M. E., & Jarrett, R. B. (2012). Outcomes of acute phase cognitive therapy in outpatients with anxious versus nonanxious depression. *Psychotherapy and Psychosomatics*, *81*(3), 153–160. https://doi.org/10.1159/000334909

Spek, V., Nyklícek, I., Smits, N., Cuijpers, P., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: A randomized controlled clinical trial. *Psychological Medicine*, *37*(12), 1797–1806. https://doi.org/10.1017/S0033291707000542

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Consulting Psychologists Press.

Spiller, T. R., Levi, O., Neria, Y., Suarez-Jimenez, B., Bar-Haim, Y., & Lazarov, A. (2020). On the validity of the centrality hypothesis in cross-sectional between-subject networks of psychopathology. *BMC Medicine*, *18*, 297. https://doi.org/10.1186/s12916-020-01740-5

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

Sreeharan, V., Madden, H., Lee, J. T., Millett, C., & Majeed, A. (2013). Improving Access to Psychological Therapies and antidepressant prescribing rates in England: A longitudinal time-series analysis. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, *63*(614), e649-653. https://doi.org/10.3399/bjgp13X671641

Stroud, C. B., Davila, J., & Moyer, A. (2008). The relationship between stress and depression in first onsets versus recurrences: A meta-analytic review. *Journal of Abnormal Psychology*, *117*(1), 206–213. https://doi.org/10.1037/0021-843X.117.1.206

Teepe, G. W., Fonseca, A. D., Kleim, B., Jacobson, N. C., Sanabria, A. S., Car, L. T., Fleisch, E., & Kowatsch, T. (2021). Just-in-time adaptive mechanisms of popular mobile apps for individuals with depression: Systematic app search and literature review. *Journal of Medical Internet Research*, *23*(9), e29412. https://doi.org/10.2196/29412

Teicher, M. H., Gordon, J. B., & Nemeroff, C. B. (2022). Recognizing the importance of childhood maltreatment as a critical factor in psychiatric diagnoses, treatment, research, prevention, and education. *Molecular Psychiatry*, *27*(3), Article 3. https://doi.org/10.1038/s41380-021-01367-9

Terluin, B. (2012). Mathematical coupling does not account for the association between baseline severity and minimally important change values. *Journal of Clinical Epidemiology*, *65*(4), 355–357. https://doi.org/10.1016/j.jclinepi.2011.04.010

Terluin, B., de Boer, M. R., & de Vet, H. C. W. (2016). Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PLOS ONE*, *11*(11), e0155205. https://doi.org/10.1371/journal.pone.0155205

Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P., & Jullagate, S. (2014). Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatric Disease and Treatment*, *10*, 2097–2103. https://doi.org/10.2147/NDT.S72026

Torres, A. R., Prince, M. J., Bebbington, P. E., Bhugra, D., Brugha, T. S., Farrell, M., Jenkins, R., Lewis, G., Meltzer, H., & Singleton, N. (2006). Obsessive-compulsive disorder: Prevalence, Comorbidity, impact, and help-seeking in the British National Psychiatric Morbidity Survey of 2000. *American Journal of Psychiatry*, *163*(11), 1978–1985. https://doi.org/10.1176/AJP.2006.163.11.1978/ASSET/IMAGES/LARGE/R122T4.JPEG

Tozzi, L., Goldstein-Piekarski, A. N., Korgaonkar, M. S., & Williams, L. M. (2020). Connectivity of the cognitive control network during response inhibition as a predictive and response biomarker in major depression: Evidence from a randomized clinical trial. *Biological Psychiatry*, *87*(5), 462–472. https://doi.org/10.1016/j.biopsych.2019.08.005

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., Oquendo, M. A., Bruder, G., Pizzagalli, D., Toups, M., Cooper, C., Adams, P., Weyandt, S., Morris, D. W., Grannemann, B. D., Ogden, R. T., Buckner, R., McInnis, M., Kraemer, H. C., … Weissman, M. M. (2016a). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research*, *78*, 11–23. https://doi.org/10.1016/j.jpsychires.2016.03.001

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., Norquist, G., Howland, R. H., Lebowitz, B., McGrath, P. J., Shores-Wilson, K., Biggs, M. M., Balasubramani, G. K., & Fava, M. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry*, *163*(1), 28–40. https://doi.org/10.1176/appi.ajp.163.1.28

Uher, R., Tansey, K. E., Malki, K., & Perlis, R. H. (2012). Biomarkers predicting treatment outcome in depression: What is clinically significant? *Pharmacogenomics*, *13*(2), 233–240. https://doi.org/10.2217/pgs.11.161

van Ballegooijen, W., Cuijpers, P., van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H., & Riper, H. (2014). Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0100674

van Beugen, S., Ferwerda, M., Hoeve, D., Rovers, M. M., Spillekom-van Koulil, S., van Middendorp, H., & Evers, A. W. (2014). Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: A meta-analytic review. *Journal of Medical Internet Research*, *16*(3), e88. https://doi.org/10.2196/jmir.2777

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA psychiatry*, *72*(12), 1219-1226.

van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. https://doi.org/10.1037/met0000476

van Breda, W., Bremer, V., Becker, D., Hoogendoorn, M., Funk, B., Ruwaard, J., & Riper, H. (2017). Predicting therapy success for treatment as usual and blended treatment in the domain of depression. *Internet Interventions*, *12*, 100–104. https://doi.org/10.1016/j.invent.2017.08.003

van de Velde, S., Bracke, P., & Levecque, K. (2010). Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science & Medicine (1982)*, *71*(2), 305–313. https://doi.org/10.1016/j.socscimed.2010.03.035

van Dooren, F. E. P., Nefs, G., Schram, M. T., Verhey, F. R. J., Denollet, J., & Pouwer, F. (2013). Depression and risk of mortality in people with diabetes mellitus: A systematic review and meta-analysis. *PloS One*, *8*(3). https://doi.org/10.1371/JOURNAL.PONE.0057058

Vieira, S., Liang, X., Guiomar, R., & Mechelli, A. (2022). Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. *Clinical Psychology Review*, 102193. https://doi.org/10.1016/j.cpr.2022.102193

Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience and Biobehavioral Reviews*, *74*(Pt A), 58–75. https://doi.org/10.1016/j.neubiorev.2017.01.002

Vigerland, S., Lenhard, F., Bonnert, M., Lalouni, M., Hedman, E., Ahlen, J., Olén, O., Serlachius, E., & Ljótsson, B. (2016). Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. *Clinical Psychology Review*, *50*, 1–10. https://doi.org/10.1016/j.cpr.2016.09.005

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abdulkader, R. S., Abdulle, A. M., Abebo, T. A., Abera, S. F., Aboyans, V., Abu-Raddad, L. J., Ackerman, I. N., Adamu, A. A., Adetokunboh, O., Afarideh, M., Afshin, A., Agarwal, S. K., Aggarwal, R., … Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2

Wagner, B., Rosenberg, N., Hofmann, L., & Maass, U. (2020). Web-based bereavement care: A systematic review and meta-analysis. *Frontiers in Psychiatry*, *11*, 525. https://doi.org/10.3389/fpsyt.2020.00525

Wallert, J., Boberg, J., Kaldo, V., Mataix-Cols, D., Flygare, O., Crowley, J. J., Halvorsen, M., Ben Abdesslem, F., Boman, M., Andersson, E., Hentati Isacsson, N., Ivanova, E., & Rück, C. (2022). Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Translational Psychiatry*, *12*(1), Article 1. https://doi.org/10.1038/s41398-022-02133-3

Wampold, B. E., Minami, T., Baskin, T. W., & Callen Tierney, S. (2002). A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of Affective Disorders*, *68*(2–3), 159–165. https://doi.org/10.1016/s0165-0327(00)00287-1

Warden, D., Rush, A. J., Trivedi, M. H., Fava, M., & Wisniewski, S. R. (2007). The STAR*D Project results: A comprehensive review of findings. *Current Psychiatry Reports*, *9*(6), 449–459. https://doi.org/10.1007/S11920-007-0061-3

Watkins, L. L., Koch, G. G., Sherwood, A., Blumenthal, J. A., Davidson, J. R. T., O'Connor, C., & Sketch, M. H. (2013). Association of anxiety and depression with all-cause mortality in individuals with coronary heart disease. *Journal of the American Heart Association*, *2*(2). https://doi.org/10.1161/JAHA.112.000068

Watson, D. (2009). Differentiating the mood and anxiety disorders: A quadripartite model. *Annual Review of Clinical Psychology*, *5*, 221–247. https://doi.org/10.1146/annurev.clinpsy.032408.153510

Watts, D., Pulice, R. F., Reilly, J., Brunoni, A. R., Kapczinski, F., & Passos, I. C. (2022). Predicting treatment response using EEG in major depressive disorder: A machine-learning meta-analysis. *Translational Psychiatry*, *12*(1), Article 1. https://doi.org/10.1038/s41398-022-02064-z

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, *88*(1), 25–38. https://doi.org/10.1037/ccp0000451

Webb, C. A., Olson, E. A., Killgore, W. D. S., Pizzagalli, D. A., Rauch, S. L., & Rosso, I. M. (2018). Rostral anterior cingulate cortex morphology predicts treatment response to internet-based cognitive behavioral therapy for depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 255–262. https://doi.org/10.1016/j.bpsc.2017.08.005

Webb, C. A., Rosso, I. M., & Rauch, S. L. (2017). Internet-based cognitive behavioral therapy for depression: current progress & future directions. *Harvard Review of Psychiatry*, *25*(3), 114. https://doi.org/10.1097/HRP.0000000000000139

Weitz, E., Kleiboer, A., Straten, A. van, & Cuijpers, P. (2018). The effects of psychotherapy for depression on anxiety symptoms: A meta-analysis. *Psychological Medicine*, *48*(13), 2140–2152. https://doi.org/10.1017/S0033291717003622

Weitz, E. S., Hollon, S. D., Twisk, J., van Straten, A., Huibers, M. J. H., David, D., DeRubeis, R. J., Dimidjian, S., Dunlop, B. W., Cristea, I. A., Faramarzi, M., Hegerl, U., Jarrett, R. B., Kheirkhah, F., Kennedy, S. H., Mergl, R., Miranda, J., Mohr, D. C., Rush, A. J., … Cuijpers, P. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: An individual patient data meta-analysis. *JAMA Psychiatry*, *72*(11), 1102–1109. https://doi.org/10.1001/jamapsychiatry.2015.1516

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Wichers, M., Wigman, J. T. W., Bringmann, L. F., & de Jonge, P. (2017). Mental disorders as networks: Some cautionary reflections on a promising approach. *Social Psychiatry and Psychiatric Epidemiology*, *52*(2), 143–145. https://doi.org/10.1007/s00127-016-1335-z

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021). A narrative review of network studies in depression: What different methodological approaches tell us about depression. *Frontiers in Psychiatry*, *12*, 1865. https://doi.org/10.3389/FPSYT.2021.719490/BIBTEX

Wickramaratne, P. J., Yangchen, T., Lepow, L., Patra, B. G., Glicksburg, B., Talati, A., Adekkanattu, P., Ryu, E., Biernacka, J. M., Charney, A., Mann, J. J., Pathak, J., Olfson, M., & Weissman, M. M. (2022). Social connectedness as a determinant of mental health: A scoping review. *PLoS ONE*, *17*(10), e0275004. https://doi.org/10.1371/journal.pone.0275004

Williams, L. M., Debattista, C., Duchemin, A. M., Schatzberg, A. F., & Nemeroff, C. B. (2016). Childhood trauma predicts antidepressant response in adults with major depression: Data from the randomized international study to predict optimized treatment for depression. *Translational Psychiatry*, *6*(5), e799–e799. https://doi.org/10.1038/tp.2016.61

Williams, L. M., Korgaonkar, M. S., Song, Y. C., Paton, R., Eagles, S., Goldstein-Piekarski, A., Grieve, S. M., Harris, A. W. F., Usherwood, T., & Etkin, A. (2015). Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *40*(10), 2398–2408. https://doi.org/10.1038/npp.2015.89

Williams, L. M., Rush, A. J., Koslow, S. H., Wisniewski, S. R., Cooper, N. J., Nemeroff, C. B., Schatzberg, A. F., & Gordon, E. (2011). International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: Rationale and protocol. *Trials*, *12*(1), 4. https://doi.org/10.1186/1745-6215-12-4

Wilson, C., & Stock, J. (2019). The impact of living with long-term conditions in young adulthood on mental health and identity: What can help? *Health Expectations*, *22*(5), 1111–1121. https://doi.org/10.1111/HEX.12944

Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., Jablenski, A., Regier, D., & Sartorius, N. (1990). SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, *47*(6), 589–593. https://doi.org/10.1001/archpsyc.1990.01810180089012

World Health Organisation. (2017). *Depression and Other Common Mental Disorders Global Health Estimates*. World Health Organisation. http://apps.who.int/iris/bitstream/handle/10665/254610/WHOMSD?sequence=1

World Health Organization. (2021). *Depression.*

World Health Organization. (2022). *World Mental Health Report: Transforming Mental Health for All*. World Health Organisation. https://www.who.int/publications/i/item/9789240049338

World Health Organization. (2023). *Depression*. https://www.who.int/news-room/fact-sheets/detail/depression
World Health Organization, & Public Health Agency of Canada (Eds.). (2005). *Preventing Chronic Diseases: A Vital Investment*. World Health Organization ; Public Health Agency of Canada.
Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. https://doi.org/10.1037/pas0000685
Wroe, A. L., Rennie, E. W., Sollesse, S., Chapman, J., & Hassy, A. (2018). Is cognitive behavioural therapy focusing on depression and anxiety effective for people with long-term physical health conditions? A controlled trial in the context of type 2 diabetes mellitus. *Behavioural and Cognitive Psychotherapy*, *46*(2), 129–147. https://doi.org/10.1017/S1352465817000492
Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393
Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., Hardy, G., Newbold, A., & 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
Zamani Esfahlani, F., Visser, K., Strauss, G. P., & Sayama, H. (2018). A network-based classification framework for predicting treatment response of schizophrenia patients. *Expert Systems with Applications*, *109*, 152–161. https://doi.org/10.1016/j.eswa.2018.05.005
Zhang, X., Yin, C., Tian, W., Lu, D., & Yang, X. (2020). Effects of cognitive behavioral therapy on anxiety and depression in patients with chronic obstructive pulmonary disease: A meta-analysis and systematic review. *The Clinical Respiratory Journal*, *14*(10), 891–900. https://doi.org/10.1111/CRJ.13226
Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, *52*(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2
Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, *56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007
Zorowitz, S., Solis, J., Niv, Y., & Bennett, D. (2021). Inattentive responding can induce spurious associations between task behavior and symptom measures. *PsyArXiv*. https://doi.org/10.31234/osf.io/rynhk
Zuelke, A. E., Luppa, M., Löbner, M., Pabst, A., Schlapke, C., Stein, J., & Riedel-Heller, S. G. (2021). Effectiveness and feasibility of internet-based interventions for grief after bereavement: Systematic review and meta-analysis. *JMIR Mental Health*, *8*(12), e29661. https://doi.org/10.2196/29661
Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, *12*(1), 63–70. https://doi.org/10.1001/ARCHPSYC.1965.01720310065008

8 Supplementary Materials

8.1 Supplementary Materials for Chapter 2

8.1.1 Supplementary Material 1 for Chapter 2

eTable 8.1.1. Comprehensive sample characteristics of LTC and non-LTC cohorts across intervention groups (iCBT, PGT, GSH).

LTC Status Group Comparisons				
Characteristics	**LTC Status**		**$\chi 2$ / t (df)**	**p**
	LTC (n=4024)	Non-LTC (n=17027)		
Gender (N, %)			2.10 (1)	0.147
Female	2620 (65.10)	11,291 (66.31)		
Male	1404 (34.90)	5736 (33.69)		
Age			-43.14 (21049)	< 0.001***
Mean, SD (Range)	46.05, 15.88 (18-80)	35.79, 12.96 (18-80)		
Employment (N, %)[a]				
Employed	2385 (60.04)	12602 (75.20)		
Unemployed	447 (11.30)	1457 (8.56)		
Student	131 (3.32)	1143 (6.83)		
Retired	563 (14.20)	575 (3.43)		
Other	424 (10.70)	969 (5.79)	982.02 (4)	< 0.001***
Ethnicity (N, %)[b]				
British	3038 (78.70)	12866 (77.90)		
Other White Background	179 (4.64)	1172 (7.10)		
Asian	391 (10.10)	1460 (8.84)		
Caribbean	85 (2.20)	327 (1.98)		
African	55 (1.42)	207 (1.25)		
Other	113 (2.93)	477 (2.89)	36.47 (6)	< 0.001***
Baseline PHQ-9				
Mean, SD	14.46 (6.03)	13.71 (5.85)	-7.31 (21049)	< 0.001***
Baseline GAD-7				
Mean, SD	12.82 (5.17)	12.98 (4.86)	1.89 (21049)	0.059
Baseline WSAS				
Mean, SD	17.83 (9.70)	16.80 (8.88)	-6.52 (21049)	< 0.001***
Treatment Duration (Days)[c]				
Mean, SD	90.47 (58.25)	89.68 (53.63)	-0.75 (17431)	0.452
Number of Appointments				
Mean, SD	5.14 (2.64)	5.10 (2.41)	-0.91 (21049)	0.361
IAPT Clinical Outcomes				
Caseness (N, %)	3636 (90.36)	15516 (91.13)	2.34 (1)	0.126
Recovery (N, %)	1984 (54.57)	8618 (55.54)	1.10 (1)	0.295
Reliable Improvement (N, %)	2364 (65.02)	10334 (66.60)	3.24 (1)	0.072

Reliable Recovery (N, %)	1803 (49.59)	7848 (50.58)	1.12 (1)	0.290

[a]N = 355 missing data for employment information. [b]N = 681 missing data for ethnicity information.
[c]Instances where treatment duration is 0 due to administrative errors were removed from analyses.

8.1.2 Supplementary Material 2 for Chapter 2

eTable 8.1.2.1. Treatment characteristics of LTC and non-LTC cohorts within iCBT intervention group

Characteristics	**LTC Status**		**t (df)**	**p**
	LTC (n=1065)	Non-LTC (n=5792)		
Treatment Duration (Days)[a]				
Mean, SD	93.93 (53.03)	90.19 (49.35)	-2.18 (6855)	0.029*
Number of Appointments				
Mean, SD	5.63 (2.34)	5.54 (2.22)	-1.13 (6855)	0.257

[a]Instances where treatment duration is 0 due to administrative errors were removed from analyses (N=427).

eTable 8.1.2.2. Treatment characteristics of LTC and non-LTC cohorts within PGT intervention group

Characteristics	**LTC Status**		**t (df)**	**p**
	LTC (n=339)	Non-LTC (n=1109)		
Treatment Duration (Days)[a]				
Mean, SD	52.34 (48.50)	48.48 (38.11)	-1.52 (1446)	0.130
Number of Appointments				
Mean, SD	4.92 (2.55)	4.68 (2.08)	-1.80 (1446)	0.072

[a]Instances where treatment duration is 0 due to administrative errors were removed from analyses (N=13).

eTable 8.1.2.3. Treatment characteristics of LTC and non-LTC cohorts within GSH intervention group

Characteristics	**LTC Status**		**t (df)**	**p**
	LTC (n=2620)	Non-LTC (n=10126)		
Treatment Duration (Days)[a]				
Mean, SD	95.07 (59.93)	95.32 (55.86)	0.18 (12744)	0.860
Number of Appointments				
Mean, SD	4.97 (2.74)	4.89 (2.51)	-1.36 (12744)	0.173

[a]Instances where treatment duration is 0 due to administrative errors were removed from analyses (N=3,178).

8.1.3 Supplementary Material 3 for Chapter 2

Intervention-Specific Effectiveness Analyses for LTC patients. For depression, there was a significant main effect for time, indicating that the LTC cohort experienced an overall reduction in PHQ-9 scores from pre-treatment to post-treatment, $F(1, 8041) = 2029.52$, $p < 0.001$. The extent to which LTC patients reduced their PHQ-9 scores from pre-treatment to post-treatment varied across intervention type, as evidenced by a time by intervention group significant interaction, $F(2, 8041) = 8.37$, $p < 0.001$. Results from Tukey post-hoc tests indicated that LTC patients engaged in iCBT exhibited significantly greater improvements in PHQ-9 score (adj. pre-treatment M = 14.22, SE = 0.13; adj. post-treatment M = 7.72, SE = 0.13, d = 1.55) than those in GSH (adj. pre-treatment M = 14.60, SE = 0.08; adj. post-treatment M = 8.67, SE = 0.08, d = 1.41, p = 0.008) and PGT (adj. pre-treatment M = 14.22, SE = 0.23; adj. post-treatment M = 9.17, SE = 0.23, d = 1.20, $p < 0.001$), while GSH was associated with greater improvements than PGT (p = 0.010). A similar pattern was observed for anxiety measured on the GAD-7; a significant main effect of time indicated an overall reduction in GAD-7 scores experienced by the LTC cohort, $F(1, 8041) = 1832.88$, $p < 0.001$. A significant interaction effect between time and intervention indicated that LTC patients across each intervention differed from each other in the extent their anxiety symptoms improved, $F(2, 8041) = 19.22$, $p < 0.001$. LTC patients in iCBT exhibited greater improvement in GAD-7 scores (adj. pre-treatment M = 12.68, SE = 0.12; adj. post-treatment M = 6.65, SE = 0.12, d = 1.56) when compared to PGT (adj. pre-treatment M = 12.43, SE = 0.21; adj. post-treatment M = 8.43, SE = 0.21, d = 1.04, $p < 0.001$) and GSH (adj. pre-treatment M = 12.92, SE = 0.08; adj. post-treatment M = 7.68, SE = 0.08, d = 1.36, p = 0.001), while those in GSH improved more than those in PGT (p = 0.001). For functional impairment, there was a significant main effect of time, $F(1, 8041) = 822.74$, $p < 0.001$, reflecting a general reduction in functional impairment from pre-treatment to post-treatment. However, there was no significant interaction effect between time and intervention group, indicating that reductions in impairment scores did not vary across the intervention groups, $F(2, 8041) = 1.40$, $p = 0.247$.

To explore whether sociodemographic differences such as age and gender of the LTC patients explained any of the observed interaction effects, we added age and gender as covariates to the models for each outcome measure. There was a significant main effect of age in each model (all $p < 0.001$), indicating that older age was generally associated with less severe symptoms. However, each of the observed time by intervention type

interaction effects remained significant. Thus, neither age or gender explained the superiority of iCBT over GSH and PGT, nor the superiority of GSH over PGT, for reducing anxiety and depression symptoms in LTC patients.

8.2 Supplementary Materials for Chapter 3

8.2.1 Variable Directory

Socio-Demographics

Age

Variable type:	Continuous discrete variable
Measurement:	"What is your age?"
Response options:	18-70 (years of age)
Variable name:	age

Sex

Variable type:	Nominal variable
Measurement:	"What is your sex?"
Response options:	Female (0), Male (1), Intersex (2), Male to Female (3), Female to Male (4)
Variable name:	Sex

Country of Residence

Variable type:	Nominal variable
Measurement:	"What is your country of residence?"
Response options:	Free-entry field
Variable name:	Country

Marital Status

Variable type:	Ordinal variable
Measurement:	"What is your marital status?"
Response options:	Single (0), Married (1), In a relationship (2), Separated (3), Divorced (4), Widowed (5)
Variable name:	MStatus

Education Level

Variable type:	Ordinal variable
Measurement:	"What is the highest level of education that you've completed?"
Response options:	No schooling completed (0), Some early primary (1), Completed primary school (2), some secondary education (3), completed secondary education (4), trade/technical/vocational training (5), some undergraduate education (6), completed undergraduate education (7), some postgraduate education (8), master's degree (9), doctorate degree (10)
Variable name:	Education

Employment Status

Variable type:	Ordinal variable
Measurement:	"What best describes your current employment status?"
Response options:	Unemployed not looking (0), unemployed looking (1), full-time employed (2), part-time employed (3), self-employed (4), retired (5)
Variable name:	Employment

Subjective Social Status

Variable type:	Continuous discrete variable

Measurement: MacArthur Scale of Subjective Social Status (Adler et al., 2000)
Response options: 0-10 (the higher the number, the higher the subjective status)
Variable name: SSStatus

Physical Health and Lifestyle

Exercise

Average days of exercise per week
Variable type: Continuous discrete variable
Measurement: "On average, how many days per week do you engage in moderate to strenuous exercise? (e.g. a brisk walk, a run, swimming, cycling, weight lifting, team sports etc.)". Item drawn from the Physical Activity Vital Sign (PAVS)
Response options: 0-7 days
Variable name: Exercise_days

Average minutes of exercise on days of exercising
Variable type: Continuous discrete variable
Measurement: "On these days, on average, how many minutes do you engage in exercise at this level?"
Response options: 0 min, 10 min, 20 min, 30 min, 40 min, 50 min, 60+ min
Variable name: Exercise_min

Diet

Overall diet quality
Variable type: Ordinal variable
Measurement: "In general, how healthy is your overall diet?"
Response options: Very poor (0), Poor (1), Fair (2), Good (3), Very good (4), Excellent (5)
Variable name: Diet_quality

Frequency of eating fish
Variable type: Ordinal variable
Measurement: "How often do you usually eat fresh or canned fish? (NOT including fish and chips)"
Response options: rarely/never (0), once a month (1), twice a month (2), once a week (3), twice a week (4), every second day (5), once a day (6), more than once a day (7)
Variable name: Diet_fish

Frequency of taking diet supplements containing fish
Variable type: Ordinal variable
Measurement: "Do you regularly take diet supplements that contain fish oils or omega 3 fatty acids?"
Response options: rarely/never (0), once a month (1), twice a month (2), once a week (3), twice a week (4), every second day (5), once a day (6), more than once a day (7)
Variable name: Diet_fishsupp

Drug Use

Marijuana use
Variable type: Ordinal variable
Measurement: "Have you ever taken marijuana?"
Response options: No (0), Yes (1)
Variable name: Drugs_1

Marijuana use habits
*Conditional on (1) in Drugs_1
Variable type: Ordinal variable
Measurement: "Have you ever taken marijuana?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_1_group

Frequency of marijuana use in the past (quitters)
*Conditional on (1) in Drugs_1_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_1_past_freq

Age when first started using marijuana (quitters)
*Conditional on (1) in Drugs_1_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_1_past_age1

Age when quitted using marijuana (quitters)
*Conditional on (1) in Drugs_1_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"

Response options: 0-70 years old
Variable name: Drugs_1_past_age2

Frequency of current marijuana use (users)
*Conditional on (2) in Drugs_1_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_1_freq

Age when first started using marijuana (users)
*Conditional on (2) in Drugs_1_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_1_age

Ecstasy/MDMA use
Variable type: Ordinal variable
Measurement: "Have you ever taken ecstasy/MDMA?"
Response options: No (0), Yes (1)
Variable name: Drugs_2

Ecstasy/MDMA use habits
*Conditional on (1) in Drugs_2
Variable type: Ordinal variable
Measurement: "Have you ever taken ecstasy/MDMA?"

Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_2_group

Frequency of ecstasy/MDMA use in the past (quitters)
*Conditional on (1) in Drugs_2_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_2_past_freq

Age when first started using ecstasy/MDMA (quitters)
*Conditional on (1) in Drugs_2_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_2_past_age1

Age when quitted using ecstasy/MDMA (quitters)
*Conditional on (1) in Drugs_2_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_2_past_age2

Frequency of current ecstasy/MDMA use (users)
*Conditional on (2) in Drugs_2_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_2_freq

Age when first started using ecstasy/MDMA (users)
*Conditional on (2) in Drugs_2_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_2_age

Stimulant drug (e.g., cocaine, amphetamines)
Variable type: Ordinal variable
Measurement: "Have you ever taken stimulant drugs (e.g., cocaine/amphetamines)?"
Response options: No (0), Yes (1)
Variable name: Drugs_3

Stimulant drug use habits
*Conditional on (1) in Drugs_3
Variable type: Ordinal variable
Measurement: "Have you ever taken stimulant drugs (e.g., cocaine/amphetamines)?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_3_group

Frequency of stimulant drug use in the past (quitters

*Conditional on (1) in Drugs_3_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_3_past_freq

Age when first started using stimulant drug (quitters)
*Conditional on (1) in Drugs_3_group

Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_3_past_age1

Age when quitted using stimulant drug (quitters)
*Conditional on (1) in Drugs_3_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_3_past_age2

Frequency of current stimulant drug use (users)
*Conditional on (2) in Drugs_3_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_3_freq

Age when first started using stimulant drug (users)
*Conditional on (2) in Drugs_3_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_3_age

Opiates (e.g., on prescription; codeine, oxycodone)
Variable type: Ordinal variable
Measurement: "Have you ever taken opiates on prescription excluding Solpadeine and Neurofen Plus (e.g. codeine, oxycodone)?"
Response options: No (0), Yes (1)
Variable name: Drugs_4

Opiate use habits
*Conditional on (1) in Drugs_4
Variable type: Ordinal variable
Measurement: "Have you ever taken opiates on prescription excluding Solpadeine and Neurofen Plus (e.g. codeine, oxycodone)?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_4_group

Frequency of opiate use in the past (quitters)
*Conditional on (1) in Drugs_4_group
Variable type: Ordinal variable

Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_4_past_freq

Age when first started using opiates (quitters)
*Conditional on (1) in Drugs_4_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_4_past_age1

Age when quitted using opiates (quitters)
*Conditional on (1) in Drugs_4_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_4_past_age2

Frequency of current opiate use (users)
*Conditional on (2) in Drugs_4_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_4_freq

Age when first started using opiates (users)
*Conditional on (2) in Drugs_4_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_4_age

'Street' opiates (e.g., without prescription; heroin/codeine/oxycodone)
Variable type: Ordinal variable
Measurement: "Have you ever taken 'street' opiates, that is, without prescription (e.g. heroine/codeine/oxycodone)?"
Response options: No (0), Yes (1)
Variable name: Drugs_5

Street opiate use habits
*Conditional on (1) in Drugs_5
Variable type: Ordinal variable
Measurement: "Have you ever taken 'street' opiates, that is, without prescription (e.g. heroine/codeine/oxycodone)?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_5_group

Frequency of street opiate use in the past (quitters)
*Conditional on (1) in Drugs_5_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)

Variable name: Drugs_5_past_freq

Age when first started using street opiates (quitters)
*Conditional on (1) in Drugs_5_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_5_past_age1

Age when quitted using street opiates (quitters)
*Conditional on (1) in Drugs_5_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_5_past_age2

Frequency of current street opiate use (users)
*Conditional on (2) in Drugs_5_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_5_freq

Age when first started using street opiates (users)
*Conditional on (2) in Drugs_5_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_5_age

Sedatives / Tranquilizers (e.g., on prescription)
Variable type: Ordinal variable
Measurement: "Have you ever taken sedatives/tranquilizers on prescription?"
Response options: No (0), Yes (1)
Variable name: Drugs_6

Sedative / tranquilizer use habits
*Conditional on (1) in Drugs_6
Variable type: Ordinal variable
Measurement: "Have you ever taken sedatives/tranquilizers on prescription?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_6_group

Frequency of sedative / tranquilizer use in the past (quitters)
*Conditional on (1) in Drugs_6_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_6_past_freq

Age when first started using sedative / tranquilizer (quitters)
*Conditional on (1) in Drugs_6_group
Variable type: Continuous discrete variable

Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_6_past_age1

Age when quitted using sedative / tranquilizer (quitters)
*Conditional on (1) in Drugs_6_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_6_past_age2

Frequency of current sedative / tranquilizer use (users)
*Conditional on (2) in Drugs_6_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_6_freq

Age when first started using sedative / tranquilizer (users)
*Conditional on (2) in Drugs_6_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_6_age

'Street' sedatives / Tranquilizers (e.g., without prescription)
Variable type: Ordinal variable
Measurement: "Have you ever taken street' sedatives / tranquilizers, that is, without prescription?"
Response options: No (0), Yes (1)
Variable name: Drugs_7

Street sedative / tranquilizer use habits
*Conditional on (1) in Drugs_7
Variable type: Ordinal variable
Measurement: "Have you ever taken 'street' sedatives / tranquilizers, that is, without prescription?"
Response options: Once/rarely in the past (0), Frequently in the past but quit (1), Still ongoing (2)
Variable name: Drugs_7_group

Frequency of street sedative / tranquilizer use in the past (quitters)
*Conditional on (1) in Drugs_7_group
Variable type: Ordinal variable
Measurement: "How frequently did you take it in the past?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_7_past_freq

Age when first started using street sedative / tranquilizer (quitters)
*Conditional on (1) in Drugs_7_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_7_past_age1

Age when quitted using street sedative / tranquilizer (quitters)
*Conditional on (1) in Drugs_7_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Drugs_7_past_age2

Frequency of current street sedative / tranquilizer use (users)
*Conditional on (2) in Drugs_7_group
Variable type: Ordinal variable
Measurement: "How frequently do you take it?"
Response options: Less than once a month (0), Every month (1), Every week (2), Daily or almost daily (3), More than once a day (4)
Variable name: Drugs_7_freq

Age when first started using street sedative / tranquilizer (users)
*Conditional on (2) in Drugs_7_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started taking it?"
Response options: 0-70 years old
Variable name: Drugs_7_age

Physical Health Comorbidities (CIRS)
Variable type: Continuous discrete variable
Measurement: Cumulative Illness Rating Scale (CIRS) (LINN et al., 1968)
Response options: A 6-item Likert scale for 13 items: None (0), Mild (1), Moderate (2), Severe (3), Extremely severe (4). A total score is yielded by summing item responses. The higher the score, the higher the impairment.
Variable name: CIRS_total

Pain (PHQ-15)
Variable type: Continuous discrete variable
Measurement: Patient Health Questionnaire (PHQ-15) (Kroenke et al., 2002)
Response options: A 3-item Likert Scale for 5 items. Not bothered at all (0), Bothered a little (1), Bothered a lot (2). A total score is yielded by summing item responses. The higher the score, the higher the pain level.
Variable name: PAIN_total

Smoking

Cigarettes

Ever smoked cigarettes
Variable type: Binary variable
Measurement: "Have you ever smoked cigarettes?"
Response options: No (0), Yes (1)
Variable name: Smoking_ever

Smoking habits (users or quitters)
*Conditional on (1) in Smoking_ever
Variable type: Ordinal variable
Measurement: "Please select the option which best applies to you"
Response options: Once or rare occasions in the past (0), Frequently in the past (1), Ongoing basis (2)
Variable name: Smoking_group

Frequency of smoking in the past (quitters)

*Conditional on (1) in Smoking_group
Variable type: Ordinal variable
Measurement: "How frequently did you smoke in the past? (include instances where you have only taken one or two puffs"
Response options: Less than once per day (0), 1-5 times (1), 5-10 times (2), 10-20 times (3), 20-30 times (4), more than 30 times (5)
Variable name: Smoking_past_freq

Age when first started smoking (quitters)
*Conditional on (1) in Smoking_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started smoking?"
Response options: 0-70 years old
Variable name: Smoking_past_AgeStart

Age when quitted smoking (quitters)
*Conditional on (1) in Smoking_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Smoking_past_AgeQuit

Frequency of current smoking (users)
*Conditional on (2) in Smoking_group
Variable type: Ordinal variable
Measurement: "How frequently do you smoke? (include instances where you have only taken one or two puffs)
Response options: Less than once per day (0), 1-5 times (1), 5-10 times (2), 10-20 times (3), 20-30 times (4), more than 30 times (5)
Variable name: Smoking_present_freq

Age when first started smoking (users)
*Conditional on (2) in Smoking_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started smoking?"
Response options: 0-70 years old
Variable name: Smoking_present_AgeStart

Vaping

Ever vaped (i.e., smoked e-cigarettes)
Variable type: Binary variable
Measurement: "Have you ever vaped (i.e., smoked e-cigarettes)?"
Response options: No (0), Yes (1)
Variable name: Smoking_vape_ever

Vaping habits
*Conditional on (1) on Smoking_vape_ever
Variable type: Ordinal variable
Measurement: "Please select the option which best applies to you:"
Response options: Once or rare occasions in the past (0), Frequently in the past (1), Ongoing basis (2)
Variable name: Smoking_vape_group

Frequency of vaping in the past (quitters)
*Conditional on (1) on Smoking_vape_group
Variable type: Ordinal variable

Measurement: "How frequently did you vape in the past? (account for all instances – even times where you have only taken one or two puffs)
Response options: Less than once per day (0), 1-5 times (1), 5-10 times (2), 10-20 times (3), 20-30 times (4), more than 30 times (5)
Variable name: Smoking_vape_past_freq

Age when first started vaping (quitters)
*Conditional on (1) on Smoking_vape_group

Variable type: Continuous discrete variable
Measurement: "How old were you when you started vaping?"
Response options: 0-70 years old
Variable name: Smoking_vape_past_AgeStart

Age when quitted vaping (quitters)
*Conditional on (1) on Smoking_vape_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you quit?"
Response options: 0-70 years old
Variable name: Smoking_vape_past_AgeQuit

Frequency of current vaping (users)
*Conditional on (2) on Smoking_vape_group
Variable type: Ordinal variable
Measurement: "How frequently do you vape? (account for all instances – even times when you have only taken one or two puffs)
Response options: Less than once per day (0), 1-5 times (1), 5-10 times (2), 10-20 times (3), 20-30 times (4), more than 30 times (5)
Variable name: Smoking_vape_present_freq

Age when first started vaping (users)
*Conditional on (2) on Smoking_vape_group
Variable type: Continuous discrete variable
Measurement: "How old were you when you started vaping?"
Response options: 0-70 years old
Variable name: Smoking_vape_present_AgeStart

Other forms of tobacco
Variable type: Binary variable
Measurement: "Do you consume other forms of tobacco (e.g. chewing tobacco, snuff) on a regular basis?"
Response options: Yes (1), No (0)
Variable name: Smoking_Other

Height and Weight

Weight
Variable type: Continuous discrete variable
Measurement: "What is your weight at present (in lbs) (please give your best estimate)"
Response options: Free numeric entry field
Variable name: currentweight

Height

Height in feet

Variable type: Continuous discrete variable
Measurement: "What is your height (in feet and inches)?"
Response options: 1-8 feet
Variable name: FeetDropdownld

Height in inches
Variable type: Continuous discrete variable
Measurement: "What is your height (in feet and inches)?"
Response options: 0-11 inches
Variable name: InchesDropdownld

Psychosocial

Stressful Life Events (SRRS)
Variable type: Continuous discrete variable
Measurement: Social Readjustment Rating Scale (SRRS) (Holmes & Rahe, 1967)
Response options: False (0), True (1) for 43 items. A total score is yielded by multiplying the 'True' answers by a specific weight. Some life events are thus classes as more stressful than others (e.g., death of a spouse adds 100 points to total, whereas a change in residence adds 20). The higher the summed score, the more stressful events.
Variable name: SRRS_total

Childhood Trauma (CTQ)
Variable type: Continuous discrete variables
Measurement: Childhood Trauma Questionnaire (CTQ) (Pennebaker & Susman, 1988)
Response options: A 5-item Likert scale for 28 items: Never true (0), Rarely true (1), Sometimes true (2), Often true (3), Very often true (4). A total score and 5 subscale scores are yielded by summing respective item responses. The higher the score, the higher the childhood trauma.
Variable name: CTQ_total, CTQ_emabuse, CTQ_physabuse, CTQ_sexabuse, CTQ_emneglect, CTQ_physneglect

Perceived Social Support (MSPSS)
Variable type: Continuous discrete variable
Measurement: Multidimensional Scale of Perceived Social Support (MSPSS) (Zimet et al., 1988)
Response options: A 7-item Likert scale for 12 items: Very strongly disagree (0), Strongly disagree (1), Mildly disagree (2), Neutral (3), Mildly agree (4), Strongly Agree (5), Very strongly Agree (6). A total score is yielded by summing item responses. The higher the score, the higher the perceived support.
Variable name: SOSU_total

Perceived Stress (PSS)
Variable type: Continuous discrete variable
Measurement: Perceived Stress Scale (PSS) (Cohen et al., 1983)
Response options: A 5-item Likert scale for 10 items: Never (0), Almost Never (1), Sometimes (2), Fairly often (3), Very often (4). A total score is yielded by summing item responses. The higher the score, the more perceived stress.
Variable name: PSS_total

Clinical

Chronicity

No. of poor mental health episodes[7]

Variable type:	Continuous discrete variable
Measurement:	"How many times in your life have you experienced an episode of poor mental health?"
Response options:	0-29 times
Variable name:	TFLTE_episodes

Age of onset of mental health episodes

Variable type:	Continuous discrete variable
Measurement:	"What age were you when you experienced your first episode of poor mental health?"
Response options:	0-70 years
Variable name:	TFLTE_ageonset

Onset of current mental health episode[8]

Variable type:	Continuous discrete variable
Measurement:	"When did this current episode of poor mental health start?"
Response options:	DD/MM/YY
Variable name:	TFLTE_currentonset

Psychiatric Diagnoses (Self)

Variable type:	Binary variables
Measurement:	"I currently have a diagnosis of… (Please select all that apply)"
Response options:	Yes (TRUE) and No (FALSE) for 13 diagnoses. A total score is yielded by summing the number of 'True' for all diagnoses.
Variable name:	Diagnoses_total, Diagnosis_Dep, Diagnosis_OCD, Diagnosis_GAD, Diagnosis_PanD, Diagnosis_PTSD, Diagnosis_BPD, Diagnosis_Schiz, Diagnosis_PersD, Diagnosis_SubD, Diagnosis_AnoD, Diagnosis_BulD, Diagnosis_BinD, Diagnosis_TicD

Psychiatric Diagnosis (Family)

Variable type:	Continuous discrete variable
Measurement:	"Do you have any close relatives who ever received a diagnosis of a mental health disorder? (Please only count your biological parents, biological siblings or biological children)"
Response options:	0-10 close relatives
Variable name:	TFLTE_famHist

Miscellaneous Psychiatric Symptoms

Variable type:	Continuous discrete variable
Measurement:	Top predictors in Chekroud and colleagues' 2016 study (Chekroud et al., 2016)
Response options:	Yes (1,) No (0) for 8 items.

- Have you experienced depressed mood most of the day, nearly every day?
- Have you been bothered by aches and pains in many different parts of your body?
- Did reminders of a traumatic event make you shake, break out into a sweat, or have a racing heart?
- Did you try to avoid activities, places, or people that reminded you of a traumatic event?

[7] Due to non-normality of the variable, the responses were binned and converted into a categorical variable 3 distinct groups: 1) <2 lifetime episodes, 2) 2-5 lifetime episodes, and 3) >5 lifetime episodes.

[8] Due to technical glitch in the survey, a small number of participants are missing the specific day / month of the year of onset. For instances missing the day (DD), the first day of the month/year provided was designated as the day of the date (i.e., 1st of the month). For instances missing the day and the month (DD/MM), the first day and first month of the year provided was designated as the day and month of the date (i.e., 1st January).

- Did you have attacks of anxiety that caused you to avoid certain situations or to change your behaviour or normal routine?
- Did standing in long lines make you feel fearful, anxious, or nervous?
- Did driving or riding in a car make you feel fearful, anxious, or nervous?
- Have you ever witnessed a traumatic event such as rape, assault, someone dying in an accident, or any other extremely upsetting incident?

Transdiagnostic Self-Report Questionnaires

Apathy (AES)[9]

Variable type:	Continuous discrete variable
Measurement:	Apathy Evaluation Scale (AES) (Marin et al., 1991)
Response options:	A 4-point Likert scale for 18 items: Not at all (0), Slightly (1), Somewhat (2), A lot (3). A total score is yielded by summing item responses. The higher the score, the higher apathetic.
Variable name:	AES_total

Alcoholism (AUDIT)

Variable type:	Continuous discrete variable
Measurement:	Alcohol Use Disorder Identification Test (AUDIT; Saunders et al., 1993)
Response options:	A 5-point Likert Scale for 8 items (0-4) and a 3-point Likert Scale for 2 items. A total score is yielded by summing item responses. The higher the score, the higher the likelihood of hazardous/harmful alcohol consumption.
Variable name:	AUDIT_total

Impulsivity (BIS)

Variable type:	Continuous discrete variable
Measurement:	Barratt Impulsivity Scale (BIS) (Patton et al., 1995)
Response options:	A 4-point Likert scale for 30 items: Rarely/Never (1), Occasionally (2), Often 3), Almost always/Always (4). A total score is yielded by summing item responses. The higher the score, the higher the level of impulsivity.
Variable name:	BIS_total

Eating disorder (EAT-26)

Variable type:	Continuous discrete variable
Measurement:	Eating Attitudes Test (EAT-26) (Garner et al., 1982)
Response options:	A 6-point Likert scale for 26 items: Never (1), Rarely (2), Sometimes (3), Often (4), Usually (5), Always (6). A total score is yielded by summing item responses. The higher the score, the higher the level of higher level of concern about dieting, body weight, or problematic eating behaviours.
Variable name:	EAT_total

Social anxiety (LSAS)[10]

Variable type:	Continuous discrete variable
Measurement:	Liebowitz Social Anxiety Scale (LSAS) (Liebowitz, 1987)

[9] The AES score labels were abbreviated in our study during administration of the measure. The response options traditionally read (1) Not at all characteristic, (2) Slightly characteristic, (3) Somewhat characteristic, (4) Very characteristic. We simply used (1) Not at all, (2) Slightly, (3) Somewhat, (4) A lot.

[10] In the original version of the scale, participants answered all 24 items twice each time pertaining to how anxious or fearful they feel in the situation respectively. Prior work has shown these correlate with one another near perfectly, so we merged the response option to reduce redundancy; participants answered all 24 items just once with regards to how anxious OR fearful they feel in the situation.

Response options: A 4-point Likert scale for 24 items: None (0), Mild (1), Moderate (2), Severe (3). A total score is yielded by summing item responses. The higher the score, the higher the level of social anxiety.
Variable name: LSAS_total

Schizotypy (SSMS)

Variable type: Continuous discrete variable
Measurement: Short Scales for Measuring Schizotypy (SSMS-R) (Mason et al., 2005)
Response options: No (0) and Yes (1) for 43 items. A total score is yielded by summing 'item responses. The higher the score, the likely the presence of schizotypy.
Variable name: SCZ_total

Depression (SDS)

Variable type: Continuous discrete variable
Measurement: Self-Rating Depression Scale (SDS) (Zung, 1965)
Response options: A 4 Likert scale for 20 items: A little of the time (1), Some of the time (2), Good part of the time (3), Most of the time (4). A total score is yielded by summing item responses. The higher the score, the more depressed the individual.
Variable name: SDS_total

State anxiety (STAI-T)[11]

Variable type: Continuous discrete variable
Measurement: State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1983)
Response options: A 4-point Likert scale for 20 items: Not at all (1), Somewhat (2), Moderately so (3), Very much so (4). A total score is yielded by summing item responses. The higher the score, the higher the level of anxiety.
Variable name: STAI_total

Obsessive-Compulsive disorder (OCI-R)

Variable type: Continuous discrete variable
Measurement: Revised Obsessive-Compulsive Inventory (OCI-R) (Foa et al., 2002)
Response options: A 5-point Likert scale for 18 items: Not at all (0), A little (1), Moderately (2), A lot (3), Extremely (4). A total score is yielded by summing item responses. The higher the score, the more likely the presence of OCD.
Variable name: OCI_total

Clinical Outcome Measures

Depression (QIDS-SR)[12]

Variable type: Continuous discrete variable
Measurement: Quick Inventory of Depressive Symptomatology – Self-Report (QIDS-SR) (Rush et al., 2003)
Response options: A 4-point Likert scale for 16 items (0 to 3). A total score is yielded by summing 9 item responses. The higher the score, the higher the level of depression.
Variable name: QIDS_total

[11] There were errors associated with the administration of the STAI-T measure. 1) The opening instructions were taken from the SDS measure rather than the STAI measure. Thus, instead of asking subjects to rate how they 'generally feel", they were instead asked to rate the response that "best describes how often you felt or behaved this way during the past several days." 2) The STAI response options were taken from the STAI State form ("not at all", "somewhat", moderately so", very much so") instead of STAI Trait Form ("almost never", "sometimes", "often", almost always").

[12] The last item of the QIDS was abbreviated from "I think of suicide or death several times a day in some detail, or I have made specific plans for suicide or have actually tried to take my life" to ("I think of suicide or death several times a day in some detail, or have actually tried to take my life.

Functional Impairment (WSAS)

Variable type:	Continuous discrete variable
Measurement:	Work and Social Adjustment Scale (WSAS; Mundt et al., 2002)
Response options:	A 8-point Likert scale for 5 items: Not at all (0), Slightly (2), Definitely (4), Markedly (6), Very severely (8). A total score is yielded by summing item responses. The higher the score, the higher the level of impairment.
Variable name:	WSAS_total

Treatment

Treatment History

Past antidepressant medication treatment

Variable type:	Binary variable
Measurement:	"Have you ever in the past completed a course of antidepressant medication?"
Response options:	No (0), Yes (1)
Variable name:	TFLTE_ADpast

No. of past antidepressant medication treatment*
*Conditional on TRUE in TFLTE_ADpast

Variable type:	Continuous discrete variable
Measurement:	"How many times?"
Response options:	0-10 times
Variable name:	TFLTE_ADpast_freq

Success of past antidepressant medication treatment*
*Conditional on TRUE in TFLTE_ADpast

Variable type:	Continuous discrete variable
Measurement:	"How many times would you say this kind of treatment has resulted in a significant improvement in your symptoms? "
Response options:	0-10 times
Variable name:	TFLTE_ADpast_helped

Ever taken Sertraline*
*Conditional on TRUE in TFLTE_ADpast

Variable type:	Binary variable
Measurement:	"Have you ever taken an antidepressant called Sertraline (also known as Zoloft/Lustral)?"
Response options:	No (0), Yes (1)
Variable name:	TFLTE_ADpast_sertraline

Blood test for antidepressant medication*
*Conditional on TRUE in TFLTE_ADpast

Variable type:	Binary variable
Measurement:	"Did your prescriber take a blood sample from you before they gave you the prescription for your antidepressant?"
Response options:	No (0), Yes (1)
Variable name:	TFLTE_ADpast_bloodTest

Past psychotherapy treatment

Variable type:	Binary variable
Measurement:	"Have you ever in the past completed a course of psychological therapy?"
Response options:	No (0), Yes (1)

Variable name: TFLTE_PTpast

Types of past psychotherapy treatment*

*Conditional on TRUE in TFLTE_PTpast

Variable type: Binary variables
Measurement: "Please select the type of psychological therapy:"
Response options: Yes (TRUE), No (FALSE) for 10 psychotherapy interventions
Variable name: 10 variables with TFLTE_PTpast the short-handed name of each psychotherapy

- Behavioural therapy (TFLTE_PTpast_BT)
- Cognitive Behavioural Therapy (TFLTE_PTpast_CBT)
- Counselling (TFLTE_PTpast_Counsig)
- Cognitive Therapy (TFLTE_PTpast_CT)
- Family Therapy (TFLTE_PTpast_FT)
- Interpersonal Therapy (TFLTE_PTpast_IPT)
- Mindfulness (TFLTE_PTpast_Mindflns)
- Other (TFLTE_PTpast_Other)
- Psychoanalytic (TFLTE_PTpast_Psychoan)
- Psychodynamic (TFLTE_PTpast_Psychodyn)

Success of past psychotherapy treatment*

*Conditional on TRUE in TFLTE_PTpast

Variable type: Continuous discrete variable
Measurement: "How many times would you say this kind of treatment has resulted in a significant improvement in your symptoms?"
Response options: 0-10 times
Variable name: TFLTE_PTpast_helped

Failure of past psychotherapy treatment*

*Conditional on TRUE in TFLTE_PTpast

Variable type: Continuous discrete variable
Measurement: "How many times would you say this kind of treatment has failed to result in a significant improvement in your symptoms?"
Response options: 0-10 times
Variable name: TFLTE_PTpast_failed

Treatment Expectation

Variable type: Ordinal variable
Measurement: "What point on this 10-point scale best describes your expectations about what is likely to happen as a result of your current antidepressant treatment?"
Response options: 10-point Likert scale from "I don't expect to feel any better" (0) to "I expect to feel completely better" (9)
Variable name: TFLTE_expectations

Antidepressant Medication Type

Variable type: Binary variables
Measurement: "Please select the antidepressant that you have been prescribed from the list below: (Check more than one if you are taking more than one)"
Response options: Yes (TRUE), No (FALSE) for 21 antidepressant medication + 1 free entry field for participants to write in unlisted antidepressant medication type.
Variable name: 22 variables with med in front of medication name

Selective Serotonin Reuptake Inhibitors (SSRIs)

- Citalopram

- Escitalopram
- Fluoxetine
- Fluvoxamine
- Paroxetine
- Sertraline

Serotonin and Norepinephrine Reuptake Inhibitors (SNRIs)

- Desvenlafaxine
- Duloxetine
- Levomilnacipran
- Venlafaxine

Tricyclics

- Amitriptyline
- Clomipramine
- Doxepin
- Dosulepin
- Imipramine
- Trimipramine
- Amoxapine
- Despiramine
- Nortriptyline
- Protriptyline

Atypical Antidepressants

- Agomelatine
- Mirtazapine

Antidepressant Medication Dosage

*Asked for the antidepressant arm only

Variable type:	Binary variables
Measurement:	"Please select the statement that best applies to you"
Response options:	Yes (TRUE), No (FALSE) for 3 binary variables including 1) taking exact dosage, 2) taking less dosage, 3) taking more dosage
Variable name:	Taking_Exact_Dosage, Taking_Less_Dosage, Taking_More_Dosage

Antidepressant Medication Side Effects

*Asked for the antidepressant arm only

Presence of antidepressant side effects

Variable type:	Binary variable
Measurement:	"Have you been bothered by any side-effects?"
Response options:	Yes (TRUE), No (FALSE)
Variable name:	SideEff

Types of antidepressant side effects

Variable type:	Binary variable
Measurement:	"Please tell us the side-effect(s) you have been most bothered by:"
Response options:	Yes (TRUE), No (FALSE) for 10 side effects + 1 free entry field for other unlisted antidepressant side effects.
Variable name:	11 variables with SideEff in front of side effect type

- Difficulty thinking or remembering (SideEff_CogDiff)
- Day-time sleepiness (SideEff_DaySl)
- Gastrointestinal symptoms e.g. nausea, diarrhea, constipation (SideEff_Gastro)
- Migraines/Headaches (SideEff_MigHe)
- Muscle/joint aches (SideEff_MusJoi)
- Night-time sleep disturbance (SideEff_NightSl)

- Personality changes, such as appearing flat, or without emotion (SideEff_PersCh)
- Sexual problems (SideEff_SexPr)
- Suicidal thoughts (SideEff_SuicTh)
- Weight Gain (SideEff_WeightG)
- Other side-effect(s) not listed above (SideEff_Other)

Treatment Adherence

Treatment adherence (antidepressant)

Variable type: Binary variable
Measurement: "Are you still taking an antidepressant medication?"
Response options: Yes (TRUE), No (FALSE)
Variable name: Taking_Antidepressant

Treatment adherence for (iCBT)

Variable type: Binary variable
Measurement: "Are you still taking an online CBT course?"
Response options: Yes (TRUE), No (FALSE)
Variable name: Taking_CBT

Treatment non-adherence reasons (antidepressant)

Variable type: Binary variables
Measurement: "Please tell us why you stopped taking the antidepressant:"
Response options: Yes (TRUE), No (FALSE) for 4 binary variables + 1 free entry field for other unlisted treatment non-adherence reasons
Variable name: 5 variables with non_adhere in front of the treatment non-adherence reason

- My mental health symptom got worse (nonadhere_gotworse)
- The medication was not making me feel better (nonadhere_nothelpful)
- Side effects were too strong (nonadhere_strongsideeff)
- Other reason (nonadhere_other)
- Other reason (free entry field; nonadhere_otherreason)

Treatment non-adherence reasons (iCBT)

Variable type: Binary variables
Measurement: "Please tell us why you stopped:"
Response options: Yes (TRUE), No (FALSE) for 4 binary variables + 1 free entry field for other unlisted treatment non-adherence reasons
Variable name: 5 variables with non_adhere in front of the treatment non-adherence reason

- My mental health symptom got worse (nonadhere_gotworse)
- The CBT was not making me feel better (nonadhere_nothelpful)
- I found the CBT too difficult (nonadhere_toodifficult)
- Other reason (nonadhere_other)
- Other reason (free entry field; nonadhere_otherreason)

Extra Treatment Information

Variable type: Binary variable
Measurement: "Is there any other information you would like us to know that we have not asked?"
Response options: Yes (TRUE), No (FALSE) + 1 free entry field for entering extra information (conditional on TRUE)
Variable name: Extra_WCI_Info (TRUE/FALSE), ExtraInfo (free filling box)

Concurrent Medication Treatment

Variable type:	Binary variable
Measurement:	2 separate questions asked for each treatment arm at 3 weekly check-in timepoints. "Have you started taking any other medications?" (antidepressant) and "Are you currently taking any medication for your mental health?" (iCBT)
Response options:	Yes (TRUE), No (FALSE)
Variable name:	Taking_OtherMed

Concurrent medication treatment type (antidepressant)*

*Conditional on TRUE in Taking_OtherMed (antidepressant)

Variable type:	Categorical string variables
Measurement:	"Please provide details about the medication(s):"
Response options:	3 free entry fields. It was not compulsory for participants to provide this information.
Variable name:	OtherMed_1, OtherMed_2, OtherMed_3

Concurrent medication treatment dosage (antidepressant)*

*Conditional on TRUE in Taking_OtherMed (antidepressant)

Variable type:	Continuous discrete variable
Measurement:	"Please provide details about the medication(s):"
Response options:	3 free entry fields. It was not compulsory for participants to provide this information.
Variable name:	OtherMedDose_1, OtherMedDose_2, OtherMedDose_3

Concurrent medication treatment type (iCBT)*

*Conditional on TRUE in Taking_OtherMed (iCBT)

Variable type:	Binary variables
Measurement:	"Please select the antidepressant that you have been prescribed from the list below: (Check more than one if you are taking more than one)"
Response options:	Yes (TRUE), No (FALSE) for 21 antidepressant medication + 1 free entry field for other unlisted antidepressant medication type. It was not compulsory for participants to provide this information.
Variable name:	22 variables with med in front of medication name (see Antidepressant Medication Type)

Concurrent Psychotherapy Treatment

Concurrent psychotherapy treatment (baseline)

Variable type:	Binary variable
Measurement:	Are you currently receiving regular psychological therapy?
Response options:	Yes (1), No (0)
Variable name:	TFLTE_PTcurrent

Types of concurrent psychotherapy treatment (baseline)*

*Conditional on TRUE in TFLTE_PTcurrent

Variable type:	Binary variables
Measurement:	"Please select the type of psychological therapy:"
Response options:	Yes (TRUE), No (FALSE) for 10 psychotherapy interventions. It was not compulsory for participants to provide this information.
Variable name:	10 variables with PT in front of psychotherapy short-handed name (e.g., TFLTE_PTcurrent_CBT)

- Behavioural therapy (TFLTE_PTcurrent_BT)
- Cognitive Behavioural Therapy (TFLTE_PTcurrent_CBT)
- Counselling (TFLTE_PTcurrent_Counsig)
- Cognitive Therapy (TFLTE_PTcurrent_CT)

- Family Therapy (TFLTE_PTcurrent_FT)
- Interpersonal Therapy (TFLTE_PTcurrent_IPT)
- Mindfulness (TFLTE_PTcurrent_Mindflns)
- Other (TFLTE_PTcurrent_Other)
- Psychoanalytic (TFLTE_PTcurrent_Psychoan)
- Psychodynamic (TFLTE_PTcurrent_Psychodyn)

Concurrent psychotherapy treatment (weekly)

Variable type:	Binary variable
Measurement:	2 separate questions asked for each treatment arm at 3 weekly check-in timepoints. "Are you currently receiving regular psychological therapy?" (antidepressant) and "Have you started any other form of psychological therapy?" (iCBT)
Response options:	Yes (TRUE), No (FALSE)
Variable name:	PT_WCI

Concurrent psychotherapy treatment type (weekly)*

*Conditional on TRUE in concurrent psychotherapy treatment (weekly)

Variable type:	Binary variables
Measurement:	"Please select the type of psychological therapy:"
Response options:	Yes (TRUE), No (FALSE) for 10 psychotherapy interventions. It was not compulsory for participants to provide this information.
Variable name:	10 variables with PT in front of psychotherapy short-handed name

- Behavioural therapy (PT_WCI_BT)
- Cognitive Behavioural Therapy (PT_WCI_CBT)
- Counselling (PT_WCI_Counsig)
- Cognitive Therapy (PT_WCI_CT)
- Family Therapy (PT_WCI_FT)
- Interpersonal Therapy (PT_WCI_IPT)
- Mindfulness (PT_WCI_ Mindflns)
- Other (PT_WCI_Other)
- Psychoanalytic (PT_WCI_Psychoan)
- Psychodynamic (PT_WCI_Psychodyn)

iCBT Treatment Engagement

*Measurement is taken objectively as users engage with the iCBT platform

Total time spent (seconds)

Variable type:	Continuous discrete variable
Measurement:	The combination of the time spent in each session (in seconds) from the first to the last log-in. Interactions lasting longer than 30 minutes are automatically counted as 1 minute, to avoid counting long idle periods when the program is open toward the total count.
Variable name:	total_time_spent_(seconds)

Number of sessions

Variable type:	Continuous discrete variable
Measurement:	The number of times (logins) the user accessed the program. If a specific session has inactivity periods longer than 30 minutes, the next moment of activity will count as a new session.
Variable name:	number_of_sessions

Average session length (seconds)

Variable type:	Continuous variable

Measurement: Dividing the total time on the platform by the number of sessions.
Variable name: avg_session_length_(seconds)

Number of activities

Variable type: Continuous variable
Measurement: Every instance a user interacted actively with the platform, e.g., completed a journal entry, used an interactive tool, downloaded, or played relaxation audios. Participants were able to use these activities as many times as they wished.
Variable name: num_activities

Number of modules in program

Variable type: Continuous discrete ariable
Measurement: The total number of modules available to the user specific to the iCBT program assigned to them.
Variable name: no_of_modules_in_program

Modules completed

Variable type: Continuous discrete variable
Measurement: The number of modules out of the total number of modules available in the program that the user completed.
Variable name: num_activities

Percentage of program viewed

Variable type: Continuous variable
Measurement: Ranging from 0-1, this metric captures the proportion of the program that is completed by the user.
Variable name: percentage_programme_viewed

Activities per session

Variable type: Continuous variable
Measurement: Dividing the total number of activities completed by the total number of sessions.
Variable name: activities_per_session

Modules per session

Variable type: Continuous variable
Measurement: Dividing the total number of modules completed by the total number of sessions.
Variable name: modules_per_session

Number of journal entries

Variable type: Continuous discrete variable
Measurement: The total number of journal entries submitted by the user on the platform.
Variable name: number_journal_entries

Length of journal entries

Variable type: Continuous discrete variable
Measurement: The total number of characters used in the total number of journal entries submitted by the user on the platform.
Variable name: length_journal_entries

Supporter Experience

Variable type: Continuous discrete variable

Measurement: The total number of reviews completed by the clinician prior to their assignment to the user. i.e. the level of experience of the clinician managing the client.
Variable name: supporter_experience

Number of reviews

Variable type: Continuous discrete variable
Measurement: The total number of reviews sent from the assigned clinician to the user so to encourage use of the platform while monitoring and providing feedback about the progress from the last review.
Variable name: num_reviews

Length of reviews

Variable type: Continuous discrete variable
Measurement: The sum of the total number of characters used in all reviews written by the clinician for the user.
Variable name: length_of_reviews

Average review length

Variable type: Continuous discrete variable
Measurement: Dividing the length of reviews by the number of reviews.
Variable name: avg_review_length

Reviews fit versus template

Variable type: Continuous variable
Measurement: Ranging from 0-1, this is a percentage showing how similar the review text is to a set of templates that are used by the service. A lower value means the message was more customised, and a higher value means the closer / more similar to the templates.
Variable name: reviews_fit_vs_template

Number of review notes

Variable type: Continuous discrete variable
Measurement: The number of replies that the user left for their clinician after a review.
Variable name: number_reviewnotes

Review note length

Variable type: Continuous discrete variable
Measurement: The total number of characters in the review notes left by the user for their clinician after a review.
Variable name: note_length

Baseline Depression Score

Variable type: Continuous discrete variable
Measurement: Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001)
Response options: A 4 Likert scale for 9 items: Not at all (0), Several days (2), More than half the days (3), Nearly every day (4). A total score is yielded by summing item responses. The higher the score, the more depressed the individual. This measure is taken at baseline (i.e., when participants first logged onto the platform).
Variable name: phq_time1

Baseline Anxiety Score

Variable type: Continuous discrete variable
Measurement: Generalised Anxiety Disorder Assessment (GAD-7) (Spitzer et al., 2006a)
Response options: A 4 Likert scale for 7 items: Not at all (0), Several days (2), More than half the days (3), Nearly every day (4). A total score is yielded by summing item

responses. The higher the score, the more anxious the individual. This measure is taken at baseline (i.e., when participants first logged onto the platform).
Variable name: gad7_time1

Number of tools used

Variable type: Continuous discrete variable
Measurement: The total number of tools used by the user on the platform. Examples include Mood Monitor and Worry Tree.
Variable name: num_tools_used

Modules viewed

Variable type: Continuous discrete variable
Measurement: The total number of modules within a specific assigned program that were viewed by users on the platform. Partially viewed modules as well as fully viewed modules contribute to this metric.
Variable name: modules_viewed

Mode of Engagement

Variable type: Continuous variable
Measurement: Ranging from 0-1, the percentage the user spent on each digital application to access the platform. Currently, iCBT can be accessed through the phone (mobile), the tablet, the computer (pc), or the SilverCloud app.
Variable name: mobile, tablet, pc, app

Site

Variable type: Categorical variable
Measurement: The mental health service the user registered with to access their iCBT treatment. Participants in the study either registered with the mental health charity Aware in Dublin, Ireland or the Talking Therapies services in Berkshire, London, UK.
Variable name: aware, Berkshire

SilverCloud Program Name

Variable type: Categorical variable
Measurement: The assigned iCBT treatment program to the user by the mental health service based on patients' mental health presentations. SilverCloud has over 40 programs across the spectrum of mental health available to their user. Participants registered with the mental health charity Aware in Dublin, Ireland were only assigned the Life Skills Online program. Below includes a non-exhaustive list of SilverCloud programs available to participants in our study.
Variable name: silvercloud_program_name

- Space from Depression
- Life Skills Online
- Space from GAD
- Space from Depression & Anxiety
- Space from Anxiety
- Space from Stress
- Space from Social Anxiety
- Space for Perinatal Wellbeing
- Space for Resilience
- Space in Chronic Pain from Depression & Anxiety
- Space from OCD
- Space from Health Anxiety

- Space from Phobia
- Space from Panic
- Space for Sleep
- Space in Lung Conditions from Depression & Anxiety
- Space in Diabetes from Depression & Anxiety

Cognitive Performance

Perceptual Decision-Making Task

Variable type: Cognitive test data, numerous derived measures

Measurement: Dot discrimination task based on (Rouault et al., 2018). A perceptual decision-making task which dissociates between decision-formation and two components of metacognitive evaluation – confidence bias (subjective confidence regardless of fluctuations owing to performance) and metacognitive efficiency (the sensitivity of one's confidence reports to correct/incorrect judgements). Gamified as a task where participants guess which sunflower has more seeds (**Figure 8.2.1A**).

Response options: Left, Right (select sides with more dots). Confidence ratings on 6-point scale from guessing to certain.

Variable name: Metacognitive bias, metacognitive efficiency, reaction time, accuracy, stimulus intensity deviations, non-decision time, decision threshold, baseline drift rate, effect of dot difference on drift rate

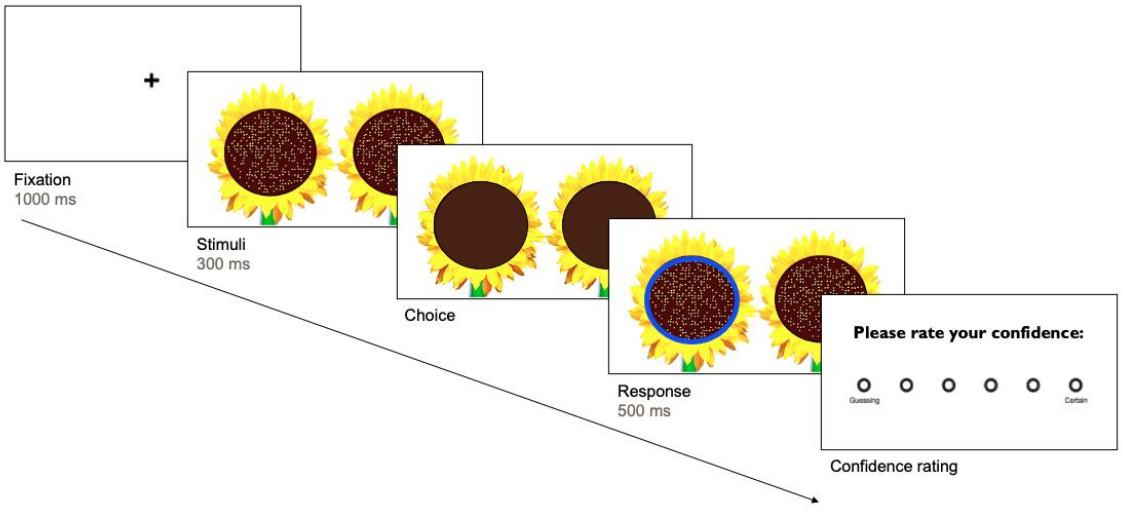


Figure 8.2.1A Dot discrimination task for assessing metacognitive evaluation.

Two-Step Reinforcement-Learning Task

Variable type: Cognitive test data, numerous derived measures

Measurement: Two-step reinforcement-learning task (Daw et al., 2011; Decker et al., 2016). A reinforcement learning task where two sequential decisions are made on each trial with a goal of maximising reward that is probabilistically associated with stimuli in the second stage choice. It allows derivation of model-based and model-free learning estimates per participant (**Figure 8.2.1B**).

Response options: Left, Right (select planets). Left, Right (select aliens).

Variable name: Model-based learning, Model-free learning, Reaction Time, RT sensitivity to Transition Structure, Choice stochasticity, Learning rate, Choice stickiness

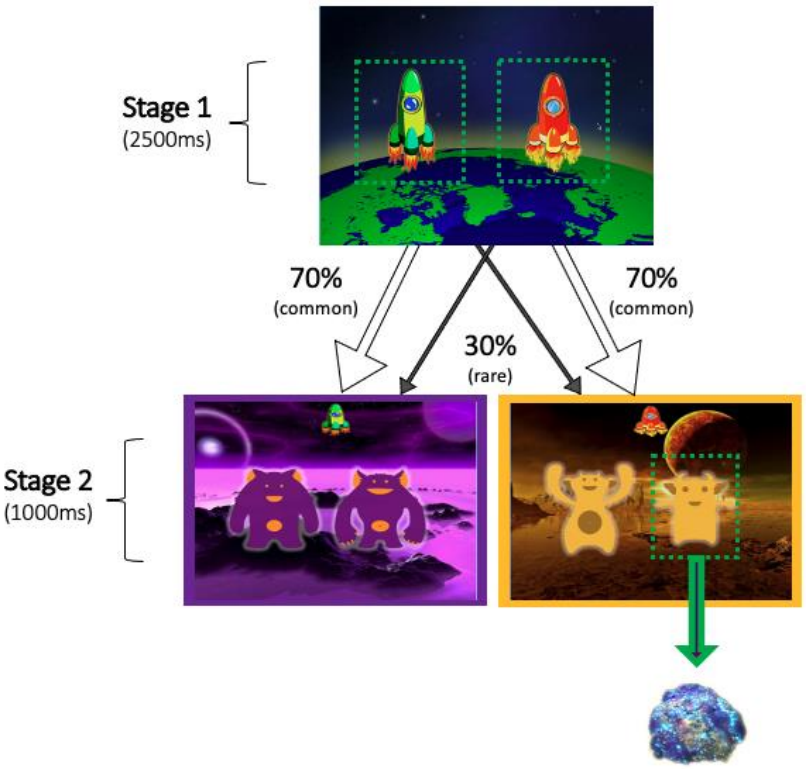


Figure 8.2.1B Reinforcement learning task (N=200 trials) for assessing goal-directed learning.

Learning Under Volatility Task

Variable type:	Cognitive test data, numerous derived measures
Measurement:	An gamified version of an aversive learning task (Behrens et al., 2007) was used to measure how well subjects learning rates adjust to environmental volatility (**Figure 8.2.1C**).
Response options:	Left, Right (select urchin).
Variable name:	Reaction Time, Choice stochasticity, Learning rate, Learning Rate Adjustment based on Volatility, Choice stickiness, Risk preference

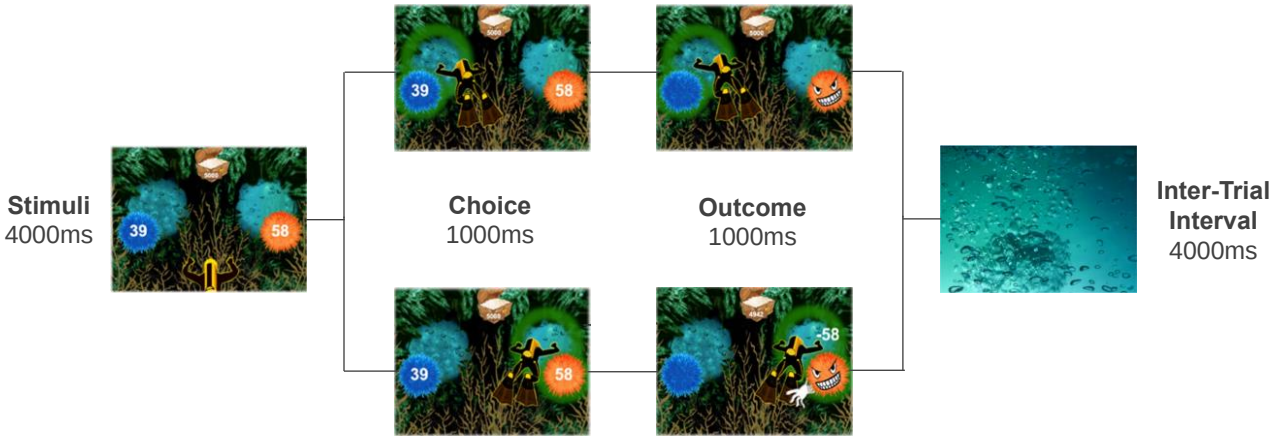


Figure 8.2.1C. Aversive learning task for assessing relative learning rate.

Abstract Reasoning Test[13]

Variable type:	Cognitive test data, numerous derived measures

[13] Due to technical errors, the completion of the Abstract Reasoning Test was not required for participants to progress further with the study. As a result, data on this cognitive task were gathered on a subsample of participants in the study.

Measurement:	Computerized Adaptive Test (CAT) based on a bank of 26 items similar to that used in the Raven's Standard Progressive Matrices (Raven, 2000) previously operationalised in a prior study (Gillan et al., 2016) (**Figure 8.2.1D**).
Response options:	Mouse selection, 1 of 8 options
Variable name:	IQ (adaptive generated theta value); number of correct responses, reaction time.

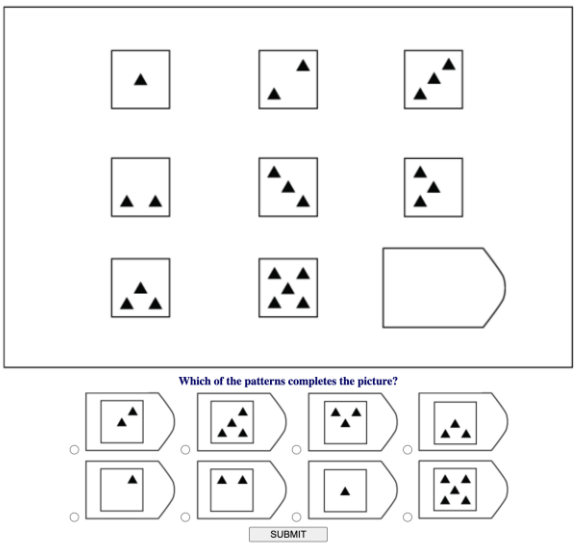


Figure 8.2.1D. An example of trial item of the IQ Computerized Adaptive Test in measuring abstract reasoning

Data Quality Probes

Distraction Probe

Variable type:	Binary variables
Measurement:	"Were you distracted by any of the following while you completed the study?"
Response options:	Self-report checkboxes (TRUE for yes and FALSE for no) for the following distraction types: TV, Phone (e.g., phone call/text message/social media), Something on the computer that you used for the study, interruptions from family/friends, Pet, Knock on the door, background noises, other (free text). If participants did not select any of these options, their response was coded as "not distracted". If they selected any of the above responses, they were coded as 'distracted'

Substance Use Probe

Variable type:	Binary variables
Measurement:	"Did you consume any of the following substances within 5 hours of starting the study?"
Response options:	Self-report checkboxes (TRUE for yes and FALSE for no) for the following substance types: Alcohol, Marijuana, Ecstasy/MDMA, Stimulant drugs (cocaine/amphetamine), Opiates (e.g. codeine, oxycodone, heroin), sedatives/tranquilisers (e.g., Valium/Xanax), Other (free text). If participants did not select any of these options, they response was coded as "no substance use". If participants selected any of the specified substances above, they were coded 'substance use'. In contrast to the distraction probe, if participants selected 'Other', they were coded as 'no substance use'. This is because based on the free text data, participants selected this option if they had taken coffee or prescription medicine, which we do not consider problematic substance use for the purposes of this item.

Attention Check OCI

Variable type:	Data quality probe
Measurement:	“If you are paying attention, select ‘a little’ as your response”
Response options:	A 5-point Likert scale: Not at all (0), A little (1), Moderately (2), A lot (3), Extremely (4).
Variable name:	OCItrap

Attention Check WSAS

Variable type:	Data quality probe
Measurement:	“If you are paying attention to these questions, please select "Not at all" as your answer”
Response options:	A 8-point Likert scale for 5 items: Not at all (0), Slightly (2), Definitely (4), Markedly (6), Very severely (8).
Variable name:	WSAScatch

8.2.2 PIP Supplementary Material

Schedule of Assessments

eTable 8.2.2.1 Measures collected per study stage, from screening, baseline, weekly check-ins to the final assessment.

	Schedule of Assessments					
Measure	Screening	Baseline	WCI 1	WCI 2	WCI 3	Final
Socio-Demographics						
Age	X					
Sex		X				
Country of Residence		X				
Marital Status		X				
Education Level		X				
Employment Status		X				
Subjective Social Status		X				
Physical Health and Lifestyle						
Exercise		X				
Diet		X				
Drug Use		X				
Physical Health Comorbidities (CIRS)		X				
Pain (PHQ-15)		X				
Smoking		X				X
Height & Weight		X				X
Psychosocial						
Stressful Life Events (SRRS)		X				
Childhood Trauma (CTQ)		X				
Perceived Social Support (MSPSS)		X				
Perceived Stress (PSS)		X				X
Clinical						
Chronicity		X				
Psychiatric Diagnoses (Self)		X				
Psychiatric Diagnoses (Family)		X				
Miscellaneous Psychiatric Symptoms		X				
Apathy (AES)		X				X
Alcoholism (AUDIT)		X				X
Impulsivity (BIS)		X				X
Eating Disorder (EAT-26)		X				X
Social Anxiety (LSAS)		X				X
Schizotypy (SSMS)		X				X
Depression (SDS)		X				X

State Anxiety (STAI-T)		X				X
Obsessive Compulsive Disorder (OCI-R)		X	X	X	X	X
Depression (QIDS-SR)		X	X	X	X	X
Functional Impairment (WSAS)	X		X	X	X	X
Treatment						
Antidepressant Medication Type[a]	X					
Treatment History		X				
Treatment Expectation		X				
Treatment Adherence			X	X	X	
Antidepressant Medication Side Effects[a]			X	X	X	
Antidepressant Medication Dosage[a]			X	X	X	
Extra Treatment Information			X	X	X	
Concurrent Medication Treatment	X		X	X	X	
Concurrent Psychotherapy Treatment		X	X	X	X	
iCBT Treatment Engagement[b]		X	X	X	X	X
Cognitive Performance						
Perceptual Decision-Making Task		X				X
Two-Step Reinforcement-Learning Task		X				X
Learning Under Volatility Task		X				X
Abstract Reasoning Test		X				X
Data Quality Probes						
Distraction Probe		X				X
Substance Use Probe		X				X
Attention Check OCI		X				
Attention Check WSAS	X		X	X	X	X

[a]Only collected for the antidepressant arm. [b]Only collected for the iCBT arm

Descriptives of Medication in the Antidepressant Arm

eTable 8.2.2.2. Proportion of participants receiving medication from the different drug classes (N = 92).

Drug Class	Frequency (N)	Percent (%)
Selective serotonin reuptake inhibitor (SSRI)	79	85.87
Serotonin–norepinephrine reuptake inhibitor (SNRI)	12	13.04
Atypical antidepressant	6	6.52
Tricyclic antidepressant (TCAs)	2	2.17
Other non-antidepressant	5	5.43

The total percentage exceeds 100 as participants can be taking more than one medication of different classes.

eTable 8.2.2.3. List of antidepressant medications prescribed (N = 92).

Name	Frequency (N)	Percent (%)	Mean Dose	SD Dose	Range Dose
SSRI					
Sertraline	37	40.22	60.20	29.59	2.5-150
Escitalopram	17	18.48	12.65	5.34	5-20
Fluoxetine	14	15.22	22.86	7.26	20-40
Citalopram	7	7.61	20	5.78	10-30
Paroxetine	2	2.17	20	0	NA
Fluvoxamine	1	1.09	20	0	NA
Vortioxetine	1	1.09	10	0	NA
SNRI					
Venlafaxine	8	8.70	51.56	27.90	37.5-112.5
Duloxetine	3	3.26	50	17.32	30-60
Levomilnacipran	1	1.09	20	0	NA

Tricyclic					
Amitriptyline	1	1.09	10	0	NA
Dosulepin	1	1.09	75	0	NA
Atypical					
Mirtazapine	6	6.52	23.33	12.11	15-45

All participants were taking at least one medication from the list above.

eTable 8.2.2.4. List of other non-antidepressant medication prescribed (N = 92).

Name	Frequency (N)	Percent (%)	Mean Dose	SD Dose	Range Dose
Propranolol[a]	1	1.09	NA	NA	NA
Olanzapine[a]	1	1.09	NA	NA	NA
Zolpidem	1	1.09	10	0	NA
Sumatriptan	1	1.09	50	0	NA
Lamotrigine	1	1.09	150	0	NA

In addition to their antidepressant medications, participants reported other medication they were taking for their brain health, entered in free-response boxes. [a]Missing N=1 dose

Baseline Characteristics of Baseline Completers

eTable 8.2.2.5. Baseline demographics of baseline completers (N = 710).

Sample Characteristics	iCBT			Antidepressant			t / X^2 (df)	p
	N	%	Median (SD)	N	%	Median (SD)		
Sex	600			110			2.58 (3)	0.46
Female	466	77.67		79	71.82			
Male	128	21.33		29	26.36			
Other	6	1.00		2	1.82			
Country	596			110			132.71 (2)	< 0.001
UK	487	81.71		40	36.36			
Ireland	100	16.78		47	42.73			
Other	9	1.51		23	20.91			
Age	598		29 (11.00)	109		26 (9.73)	-2.22 (705)	0.03
Marital Status	600			110			1.86 (5)	0.87
Single	239	39.83		47	42.73			
In a Relationship	178	29.67		36	32.73			
Married	145	24.17		21	19.09			
Divorced	21	3.50		3	2.73			
Separated	16	2.67		3	2.73			
Widowed	1	0.17		0	0.00			
Education Level	600			110			4.47 (2)	0.11
<Third Level	142	23.67		16	14.55			
Some/Complete Third Level	323	53.83		66	60.00			
>Third Level	135	22.50		28	25.45			
Employment Status	502			110			10.16 (2)	0.006
Employed	414	69.00		59	53.64			
Unemployed	180	30.00		50	45.45			
Retired	6	1.00		1.0	0.91			
Subjective Social Status	600		4 (1.69)	110		4 (1.96)	1.44 (708)	0.15

Outliers were not excluded in the descriptive analyses of demographic characteristics.

eTable 8.2.2.6. Baseline clinical characteristics of baseline completers (N = 710).

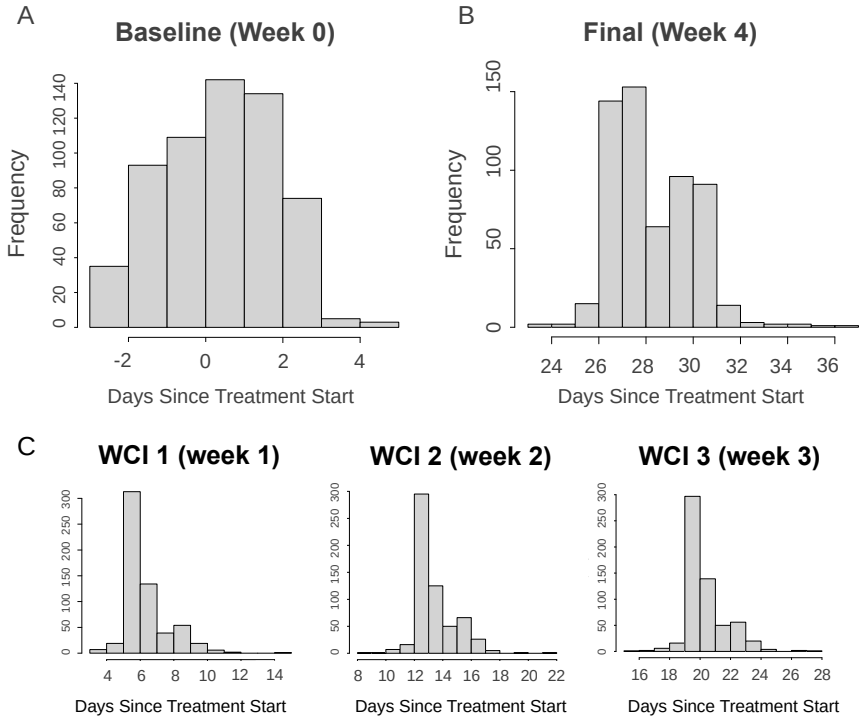
Sample Characteristics	**iCBT**			**Antidepressant**			**t / X^2 (df)**	**p**
	N	%	Median (SD)	N	%	Median (SD)		
No. of Current Diagnosis	600			110			31.37 (2)	< 0.001
None	192	32.00		8	7.27			
One	217	36.17		45	40.91			
>One	191	31.83		57	51.82			
Types of Diagnoses[a]	600			110			9.57 (5)	0.09
None	192	32.00		8	7.27			
Depression	287	47.83		81	73.64			
GAD	245	40.83		63	57.27			
Panic Disorder	30	5.00		5	4.55			
PTSD	25	4.17		13	11.82			
OCD	25	4.17		5	4.55			
Others	46	7.67		13	11.82			
Family with Mental Disorders	600			110			3.47 (3)	0.33
None	251	41.83		37	33.64			
One	187	31.17		35	31.82			
Two	92	15.33		21	19.09			
≥Three	70	11.67		17	15.45			
No. of Lifetime Episodes	592			109			7.50 (2)	0.02
<2	63	10.64		9	8.26			
2-5	292	49.32		41	37.61			
>5	237	40.03		59	54.13			
Age of onset (years)	588			109			12.20 (2)	0.002
Childhood (1-12)	107	18.20		31	28.44			
Teenage (13-17)	251	42.69		53	48.62			
Adulthood (18-70)	230	39.12		25	22.94			
Current episode length (days)	543		187 (2456)	100		202 (2320)	0.01 (641)	0.99
History of Past Treatment	600			110			9.31 (3)	0.03
Never Before	273	45.50		36	32.14			
Psychotherapy & Medication	136	22.67		32	28.57			
Medication only	95	15.83		15	13.39			
Psychotherapy only	96	16.00		27	24.11			
Treatment Expectation (0-9)	600		5 (2.05)	110		5 (1.91)	-1.23 (708)	0.22

[a]Types of Diagnoses: The total number of diagnoses type exceeds the sample size of baseline completers (i.e., participants have the option to pick more than one diagnosis). Outliers were not excluded in the descriptive analyses of clinical characteristics.

Baseline Clinical Score Distribution

eFigure 8.2.2.1. Baseline clinical symptom total score distribution for the iCBT and antidepressant arm.

Study Schedule Compliance



eFigure 8.2.2.2. Distributions of overlapping completions dates of baseline, weekly, and final assessments.

Reliability and Validity of Self-Report Data

eTable 8.2.2.7. Data quality checks on completers (N = 573) at the final assessment.

Data Quality Items	iCBT		Antidepressant		X^2 (df)	p
	N	%	N	%		
Distraction	488	100	85	100	2.82 (1)	0.09
Yes	310	63.52	62	72.94		
No	178	36.48	23	27.06		
Substance Use	488	100	85	100	0.0005 (1)	0.98
Yes	17	3.48	3	3.53		
No	471	96.52	82	96.47		

At final assessment, N = 21 were missing data for distraction and substance use data quality item checks.

eTable 8.2.2.8. Distraction types and substance uses of distracted and/or intoxicated completers at the final assessment.

Data Quality Items	N	%
Distraction Types	372	
Background Noise	125	33.60
Family and Friends	122	32.80
Phone	99	26.61
Others	189	50.81
Substance Uses	20	
Alcohol	13	65.00

Opiates	4	20.00
Marijuana	3	15.00
Others	1	5

eTable 8.2.2.9. Internal consistency of self-report scales at baseline and final assessments (Cronbach's alpha).

Clinical Symptoms	Baseline	Follow-up
	α	α
Depression (QIDS-SR)	0.71	0.81
Impairment (WSAS)[a]	0.72	0.86
Apathy (AES)	0.86	0.89
Alcohol Use (AUDIT)[a]	0.86	0.85
Impulsivity (BIS)	0.82	0.83
Eating Disorder (EAT)	0.89	0.89
Social Anxiety (LSAS)	0.95	0.95
OCD (OCI-R)	0.89	0.92
Schizotypy (SSMS)	0.84	0.87
Depression (SDS)	0.78	0.84
Trait Anxiety (STAI)	0.86	0.92

*At baseline, N = 3 were missing AUDIT symptom score and N = 12 were missing WSAS symptom score.

Between-Group Comparisons

In relation to participant recruitment and attrition, a total sample of N = 594 were recruited for analyses of this study, of which N = 502 (85%) were in the iCBT arm and N = 92 (15%) were in the antidepressant arm. Retention of baseline completers (N = 710) to weekly check-in 3 was excellent for all groups, with no significant differences between iCBT participants (93%) and antidepressant participants (92%), $X2 = 0.06$ (1), $p = 0.80$. For the final assessment, those numbers dropped to 84% for both treatment arms.

In relation to patient demographic and clinical characteristics (**Table 3.1** and **Table 3.2** in the main text), there were no between-group differences in sex, but there was a trend for participants in the iCBT arm to be older than those in the antidepressant arm, $t(590) = -1.78$, $p = 0.08$. Participants in the iCBT arm came from the United Kingdom and Ireland (>98%), whereas the antidepressant arm was more international with 17% coming from other regions around the world, $X2 = 86.93$ (2), $p < 0.001$. Education level, marital status, and subjective social status did not differ across study arms (all $p \geq 0.16$), but there were greater rates of unemployment in the antidepressant arm, $X2 = 12.81$ (2), $p = 0.002$. Most participants reported having one or more mental health diagnoses. This differed across arms; 9% of participants in the antidepressant arm had no formal diagnosis versus 31% of

the iCBT arm, X2 = 21.67 (2), p < 0.001. This was expected as a proportion of the CBT arm were self-referring. There was no difference across study arms in the number of participants reporting having a family member with a mental health condition, X2 = 1.80 (3), p = 0.62. Participants in the antidepressant arm compared to the iCBT arm reported having more lifetime mental health episodes, X2 = 11.09 (2), p = 0.004, as well as earlier age of onset, X2 = 8.68 (2), p = 0.01, but there were no group differences in self-reported duration of current mental health episode, X2 = 0.05, p = 0.96). There was a trend towards more participants in the iCBT arm being treatment-naïve at 45% versus 32% in the antidepressant arm, X2 = 5.95 (3), p = 0.11. To assess expectations about treatment efficacy, participants rated on a scale from 0-9 ("I don't expect to feel any better" to "I expect to feel completely better"). Those in the iCBT arm had a trend towards higher expectations (M = 5.02, SD = 2.04) about treatment success than those in the antidepressant arm (M = 4.58, SD = 1.89), t(592) = -1.95, p = 0.05.

In terms of clinical severity at baseline, participants in the antidepressant arm had a mean QIDS-SR score of 16.51 (SD = 4.17), which is conventionally interpreted as corresponding to severe symptoms of depression (24). QIDS-SR scores were somewhat lower in the iCBT arm with a mean of 13.86 (SD = 4.28), which corresponds to moderate depression symptoms (t(592) = 5.47, p < 0.001. For WSAS, those in the antidepressant arm had a mean of 22.73 (SD = 6.84), which indicates severe impairment in functioning. In the iCBT arm, scores were again lower with a mean of 19.02 (SD = 6.65), falling in the moderate range, t(580) = 4.61, p < 0.001. Between-group comparisons of clinical severity of other symptoms assessed at baseline are shown in **Table 3.3** in the main text, where the trend continued for those in the antidepressant arm to have greater severity.

While comparisons between-groups on pre-post 4-week clinical changes are discussed in the main text, here we report between-group differences in relation to clinical change trajectory. To understand the trajectory of treatment response, we examined the weekly depression scores measured on the QIDS-SR. We carried out a linear regression with QIDS-SR as the dependent variable and assessment week (0-4, within-subject) and treatment group (between-subject) as independent variables. As expected, there was a linear effect of assessment week, β = -1.25, p < 0.001, indicating that symptoms decreased with time. There was a main effect of group, such that symptoms were overall lower in the iCBT group (M = 11.9, SD = 4.84) when compared to the antidepressant group (M = 12.9, SD = 5.40), β = -1.86, p < 0.001. There was a significant interaction

between group and assessment week, $\beta = 0.44$, $p < 0.001$. Tests of simple effects revealed significant differences across the study arms in depression at baseline only, such that as reported earlier, the iCBT group (M = 13.86, SD = 4.28) initiated the study with lower QIDS scores than the antidepressant group (M = 16.51, SD = 4.17), $\beta = -2.65$, $p < 0.001$. There were however no group differences in depression scores at any other timepoint (all $p >= 0.21$). To identify where in treatment the biggest gains occurred and if this differed by group, we tested for significant changes in symptoms from week to week. For depression, there were significant reductions in symptoms week-on-week for the iCBT group (all $p < 0.001$), except for the final interval of week 3 to week 4 ($\beta = -0.1$, $p = 0.45$). Similar trends can be observed for the antidepressant group, where symptoms reduced week-on-week ($p <= 0.006$) up until the final interval of week 3 to week 4 ($\beta = -0.24$, $p = 0.50$). In terms of between-group differences, participants in the antidepressant arm had larger improvements in depression for the baseline to week 1 interval only compared to iCBT ($\beta = 2.00$, $p < 0.001$), while all other week-to-week changes did not differ across groups (all $p > 0.58$).

There were also between-group differences in study schedule compliance for the comprehensive assessments at the baseline and final timepoint. Participants in the antidepressant arm completed the baseline assessment 1.24 days after initiating antidepressant medication (Median = 1, SD = 1.64, range = -3 to +5 days). For the iCBT cohort, this was reduced to 0.77 days (Median = 1, SD = 1.45, range = -2 to +4 days), which was significantly shorter than for the antidepressant arm, $t(591) = 2.80$, $p = 0.005$. For the iCBT arm, the median interval between treatment initiation and final assessment was 28 days (M = 28.70, SD = 1.59, range = 24-36). In the antidepressant arm, the median interval between treatment initiation and final assessment was 29 days (M = 29.15, SD = 2.42, range = 23-37), which was significantly longer than in the iCBT arm, $t(586) = 2.24$, $p = 0.03$. We further compared the two groups in terms of differences in the number of participants who did not complete the baseline and final assessments in one sitting, defined as those who did not take a break exceeding 4 hours between the study sections. At baseline, 9% (N = 47) of participants in the iCBT arm and 23% of participants (N = 21) for the antidepressant arm did not complete the assessment in one single session. At the final assessment, similarly, 9% (N = 43) in the iCBT arm and 16% (N= 15) in the antidepressant arm did not complete it in a single session. Arm differences

were significant at baseline, X2 = 13.90 (1), p < 0.001, and at final, X2 = 5.28 (1), p = 0.02.

In relation to group differences in data quality, at baseline, there was a trend for participants in the antidepressant arm (N = 61, 75%) to be more distracted than those in the iCBT arm (N = 310, 65%), X2 = 3.56 (1), p = 0.06. There were no differences between the iCBT and antidepressant study arms on the proportion who took intoxicating substances during participation, X2 = 0.91 (1), p = 0.34 (see **eTable 8.2.2.7** and **eTable 8.2.2.8** for similar trends in distraction and substance use items at final assessment). There were also no significant between-group differences in the proportion of participants who were inattentive (i.e., failed the catch questions) (iCBT N = 51, 10%; Antidepressant N = 12, 13%), X2 = 0.68 (1), p = 0.41.

iCBT Program Effects

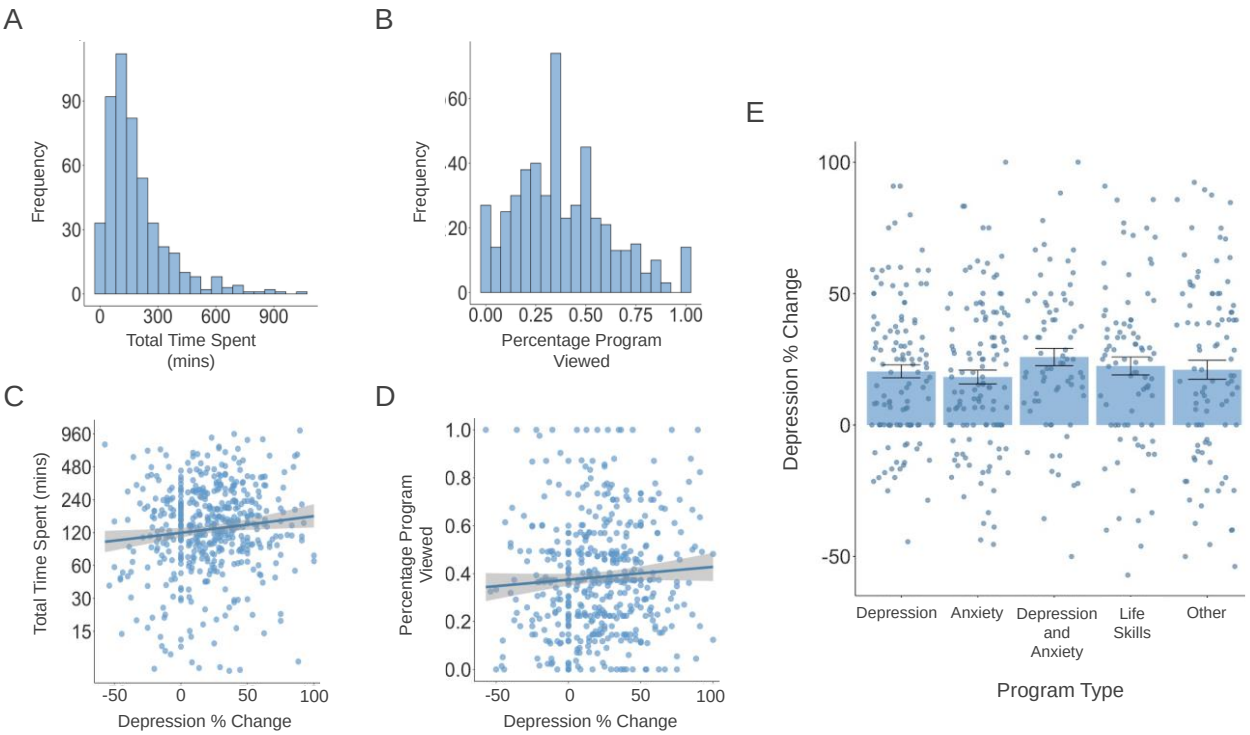
Objective treatment engagement data were available from SilverCloud for almost all participants in the iCBT group (N = 491). Consistent with the self-paced nature of iCBT, we found that engagement varied considerably across participants (see **eTable 8.2.2.10**, **eFigure 8.2.2.3A** and **eFigure 8.2.2.3B**). For example, on average, participants spent just over 2 hours on the platform over the 4-week study period, with values ranging from 2.5 minutes to >17 hours. This corresponded to participants on average viewing 37% of the program they received. On average, participants received two reviews from their supporter and completed 17 activities. Correlational analyses revealed associations between QIDS percent change and total time spent on the platform, r(483) = 0.11, p = 0.01 (**eFigure 8.2.2.3C**), average time per session, r(486) = 0.15, p < 0.001, number of activities, r(486) = 0.11, p = 0.02, and activities per session r(483) = 0.14, p < 0.001.The percentage of the program viewed was not significantly related to percent change, r(486) = 0.07, p = 0.13 (**eFigure 8.2.2.3D**), nor were the number of sessions, the number of reviews by supporters and the number of notes sent from client to supporter (all p > 0.57). Our sample represented users utilising a range of iCBT programs, which include unique tools and modules, but largely overlap in terms of core content. To examine whether the inclusion of a range of programs had effects on treatment response, we compared QIDS percent change across participants undertaking different iCBT program types. Specifically, we compared data from the most utilised programs including Space from Depression, Space from Depression & Anxiety, Space from Anxiety, Space from Generalised Anxiety Disorder (GAD) and Life Skills. Miscellaneous programs were

merged into an 'Others' category (see **Supplementary Materials 8.2.1**). A one-way ANOVA demonstrated no significant effect of program type on depression percent change score measured by the QIDS, $F(4, 483) = 0.79$, $p = 0.53$, indicating that despite there being some variation in the specific iCBT programs participants received, the treatment effect of iCBT on depression was not dependent on program type (**eFigure 8.2.2.3E**).

eTable 8.2.2.10 Summary statistics of iCBT treatment engagement over 4 weeks (N = 491).

	% (N)	**Range**
Program Type		
Space from Depression	24.03 (118)	
Space for Life Skills[a]	16.90(83)	
Space from GAD	15.27 (75)	
Space from Depression & Anxiety	14.87 (73)	
Space from Anxiety	9.57 (47)	
Others	19.35 (95)	
	Median (SD)	
Total Time Spent (minutes)	140.62 (162.38)	2.5-1033.08
Number of Sessions[b]	11 (11.20)	2-133
Average Time per Session (minutes)	13.24 (10.75)	1.16-73.04
Number of Activities[c]	17 (15.02)	1-99
Activities per Session	1.50 (0.98)	0.2-6.71
Percentage of Program Viewed	37.18 (23.54)	0-100
Number of Reviews[d]	2 (0.92)	0-4
Number of Review Notes[e]	1 (2.47)	0-42

[a]All participants recruited through Aware in Ireland was offered Space for Life Skills program. [b]A single session is defined as each individual instance users logged onto the program. If a specific session has inactivity periods longer than 30 minutes, the next moment of activity will count as a new session. [c]An activity is defined as an instance a user interacted actively with the platform, e.g., completed a journal entry, used an interactive tool, downloaded, or played relaxation audios. [d]A review is defined as a message the assigned clinician send to the user so to encourage use of the platform while monitoring and providing feedback about the progress from the last review. [e]A review note is defined as a reply that the user left for their clinician after a review. Treatment engagement data of the iCBT completers cohort were missing for N = 11 due to technical error.



eFigure 8.2.2.3 iCBT treatment engagement data (A) Distribution of total time spent (mins) by participants in the iCBT arm across the first 4 weeks of iCBT treatment. (B) Distribution of the proportion of platform viewed by participants in the iCBT arm across the first 4 weeks of iCBT treatment. (C) Relationship between pre-post 4-week depression (QIDS) percent change and total time spent (mins). Correlation analyses showed a significantly positive association between the two variables, r(483) = 0.11, *p* = 0.01. (D) Relationship between pre-post 4-week depression (QIDS) percent change and proportion of platform viewed. Correlation analyses revealed a nominally positive but not statistically significant association between the two variables, r(483) = 0.07, *p* = 0.11. (E) iCBT program specific effects on pre-post 4-week depression (QIDS) percent change. A one-way ANOVA revealed no significant effect of program type on depression percent change score measured by the QIDS, F(5, 482) = 0.75, *p* = 0.59.

Sensitivity Analyses

Excluding inattentive responders on clinical change

Pre-Post 4-Week Clinical Changes. For participants in the iCBT arm, depression score significantly decreased by an average of 3.05 points (SD = 4.24) (22%), t(449) = 15.25, p < 0.001, d = 0.72. For participants In the antidepressant arm, depression score significantly decreased by an average of 5.38 points (SD = 4.94) (33%) on the QIDS-SR, t(79) = 9.74, p < 0.001, d = 1.10. A two-way ANOVA confirmed this difference was significant, F(1, 528) = 19.46, p < 0.001. Participants in the antidepressant arm

experienced a significantly larger percent reduction in QIDS-SR from baseline than those in the iCBT arm, $t(526) = 3.10$, $p = 0.002$, after controlling for baseline severity, $t(525) = 2.52$, p.adj = 0.01. With regards to response and remission rates, for the iCBT arm, by week 4, 39% of participants have achieved 'early response', 17% of participants have achieved 'response', and 13% of participants have achieved 'remission'. Participants in the antidepressant arm showed significantly higher rate of early response at 54%, $X^2 = 5.51$ (1), $p = 0.02$, and rate of response at 33%, $X^2 = 9.64$ (1), $p = 0.002$, but not in their remission rate of 13%, $X^2 = 0.05$ (1), $p = 0.82$. With regards to general functional impairments (WSAS), participants in the iCBT arm had a significant decrease from baseline to final assessment by an average of 1.69 points (SD = 7.43) (9%), $t(448) = 4.82$, $p < 0.001$, $d = 0.23$, while those in the antidepressant arm reported an average reduction of 3.59 points (SD = 8.29) (16%), $t(69) = 3.62$, $p < 0.001$, $d = 0.43$. These differences across arms were not significant, $t(511) = 1.59$, $p = 0.11$. Furthermore, significant reductions in most other clinical symptoms were observed, in both treatment arms (all $p<0.05$), with the exception of schizotypy ($p = 0.1$) and impulsivity ($p= 0.34$) in the antidepressant arm.

Clinical Change Trajectories. A regression with QIDS-SR as the dependent variable and assessment week (0-4, within-subject) and treatment group (between subject) as independent variables was conducted. There was a linear effect of assessment week, $\beta = -1.32$, $p < 0.001$, indicating that symptoms decreased with time. There was a group main effect where symptoms were overall lower in the iCBT group (M = 11.7, SD = 4.77) than that in the antidepressant group (M = 12.4, SD = 5.25), $\beta = -1.77$, $p < 0.001$. There was a significant interaction between group and assessment week, $\beta = 0.53$, $p < 0.001$. Tests of simple effects revealed significant differences across the study arms in depression at baseline only (iCBT M = 13.60, SD = 4.18; antidepressant M = 16.40, SD = 3.96), $\beta = -2.76$, $p < 0.001$. There were no group differences in depression scores at any other timepoint (all $p >= 0.42$). For depression, there were significant reductions in symptoms week-on-week for the iCBT group (all $p < 0.001$), except for the final interval of week 3 to week 4 ($\beta = -0.05$, $p = 0.72$). Similar trends can be observed for the antidepressant group, where symptoms reduced week-on-week from baseline to the third weekly check-in ($p <= 0.002$). However, there was a slight increase in depression score from week 3 to week 4 where the final assessment took place ($\beta = -0.67$, $p = 0.04$). Participants in the antidepressant arm had larger improvements in depression for the baseline to week 1

interval only compared to iCBT ($\beta = 2.33$, $p < 0.001$), while all other week-to-week changes did not differ across groups (all $p > 0.08$).

Comparison of baseline QIDS-SR symptomatology between dropouts and completers

Sensitivity analyses comparing baseline QIDS-SR total scores between baseline completers (i.e., dropped out subsequent to completing the baseline assessment) and completers (completed all study assessments at all time points) revealed no significant differences between the two cohorts (drop outs M = 14.00, SD = 4.79; completers M = 14.30, SD = 4.37), $t(708) = 0.59$, $p = 0.56$. Similarly, there were no significant differences in baseline WSAS total scores between drop outs (M = 20.70, SD = 7.59) and completers (M = 19.50, SD = 6.79), $t(694) = -1.62$, $p = 0.11$.

Inattentive responders on response consistency indicators

We examined inattentive responders' (N = 63) consistency in self-report height across two timepoints in the study (baseline and final assessments). Results were similar to that of the total sample, where the two height reports were highly correlated across the time points, $r(61) = 94$, $p < 0.001$. We also examined the internal consistency of self-report symptom assessments of inattentive responders. Cronbach's alpha were good for all scales at baseline (i.e., range 0.74-0.96) and at final (i.e., range 0.75-0.96) at the final assessment.

Qualitative Data Analysis

A qualitative content analysis was conducted on four open-ended free-text questions included in the online feedback survey. Content analysis is a qualitative research method used to analyse text-based, qualitative data through subjective interpretation of their content by way of systematic coding, categorising, and identifying themes or patterns. This method was deemed appropriate for the purpose of this analysis as it allows researchers to quantify concepts in the data by counting the number of times these concepts appeared, thus providing descriptive statistics fit for the quantitative reporting of this data. Two researchers reviewed the responses to each four of the survey questions independently and identified categories and sub-categories capturing common themes and ideas. A codebook describing each category/sub-category with inclusion/exclusion criteria was devised independently by the two researchers before they converged to discuss and agree on the list of categories/sub-categories. In the case where multiple themes emerged within one response, each distinct theme was categorised on their own,

meaning respondents may have duplicated responses with specific elements pertaining to separate categories. Using the codebook, one of these researchers subsequently coded the responses and another independent research assistant separately coded 25% of responses to each question. Inter-rater reliability was checked using percentage agreement on the subset of 25% of responses to each question. Inter-rater reliability for each question exceeded 80%.

8.3 Supplementary Materials for Chapter 4

8.3.1 Predictor Importance

eTable 8.3.1.1 Ranked coefficients of retained predictors in the final elastic net regression model employing the total score feature set.

Model with 27 retained predictors (39 times)		**Model with 31 retained predictors (61 times)**	
Variable	*Beta*	*Variable*	*Beta*
QIDS total score	0.349	QIDS total score	0.344
SDS total score	-0.128	SDS total score	-0.134
Treatment expectation	0.096	Treatment expectation	0.098
Anhedonia (QIDS-13)	0.084	Anhedonia (QIDS 13)	0.091
MSPSS total score	0.082	MSPSS total score	0.083
View of self (QIDS-11)	0.062	Sadness (QIDS 05)	0.069
Sadness (QIDS 05)	0.062	View of self (QIDS 11)	0.067
Age of onset (adolescence)	-0.061	Age of onset (adolescence)	-0.066
AES total score	-0.045	AES total score	-0.049
WSAS total score	-0.040	WSAS total score	-0.043
CIRS total score	-0.040	CIRS total score	-0.041
SRRS total score	-0.033	Appetite/weight problems	0.038
Appetite/weight problems	0.033	SRRS total score	-0.040
Diet (fish intake)	0.030	Diet (fish intake)	0.032
Other psychiatric diagnoses	-0.026	Other psychiatric diagnoses	-0.029
LSAS total score	-0.025	LSAS total score	-0.026
BIS total score	-0.020	Anxiety standing in long lines (LP 06)	-0.022
Anxiety standing in long lines (LP 06)	0.020	BIS total score	-0.021
Gender	-0.010	Gender	-0.016
Metacognitive Efficiency	-0.009	Metacognitive Efficiency	-0.012
Model-based index	-0.009	Model-based index	-0.012
Model-free index	-0.008	Fatigue (QIDS 14)	0.010
Fatigue (QIDS 14)	0.005	Model-free index	-0.009
Exercise duration	0.004	Exercise duration	0.007
Trauma avoidance (LP 04)	-0.003	Trauma avoidance (LP 04)	-0.006
Helpfulness of past psychotherapy	0.001	Helpfulness of past psychotherapy	0.004
Diet (fish supplement intake)	-0.001	Diet (fish supplement intake)	-0.003
		Learning rate (stable environment)	0.001
		Risk preference (volatile environment)	0.001
		Body mass index	-0.001
		Body aches and pains (LP 02)	0.0003

eTable 8.3.1.2. Ranked averaged coefficients of retained predictors in the final elastic net regression model employing the total score feature set.

Variable	***Beta***
QIDS total score	0.346
SDS total score	-0.132
Treatment expectation	0.097
Anhedonia (QIDS 13)	0.088
MSPSS total score	0.082
Sadness (QIDS 05)	0.066
View of self (QIDS 11)	0.065
Age of onset (adolescence)	-0.064
AES total score	-0.047

WSAS total score	-0.042
CIRS total score	-0.041
Appetite/weight problems	0.036
SRRS total score	-0.035
Diet (fish intake)	0.031
Other psychiatric diagnoses	-0.028
LSAS total score	-0.025
Anxiety standing in long lines (LP 06)	-0.021
BIS total score	-0.021
Gender	-0.014
Metacognitive Efficiency	-0.011
Model-based index	-0.011
Fatigue (QIDS 14)	0.008
Model-free index	-0.008
Exercise duration	0.006
Trauma avoidance (LP 04)	-0.005
Helpfulness of past psychotherapy	0.003
Diet (fish supplement intake)	-0.002
Learning rate (stable environment)	0.001
Risk preference (volatile environment)	0.001
Body mass index	-0.001
Body aches and pains (LP 02)	0.0003

8.4 Supplementary Materials for Chapter 5

8.4.1 Results on Patients Receiving 8-12 Weeks of iCBT

From our archival dataset, we examined data from patients who received iCBT for depression for 8-12 weeks (N=22,952) to test the generalisability of the main findings (**eFigure 8.4.6**). This analysis captured participants who needed a longer duration to achieve treatment response, but also reduced the sample size as many individuals from the main analysis had stopped using the programme by week 8. This sample therefore also partially overlapped with those in the main analysis.

In line with our main analyses, the full sample Non-Responder network (n=11,476) had greater network connectivity than the full sample Responder network (n=11,476) at baseline (3.08 vs. 2.74, S=0.34, *P*<.001). We found that 7/36 edges were significantly different between-groups (all *P*<.05). As in the main findings, this connectivity differences between Responders and Non-Responders disappeared (2.75 vs. 2.71, S=0.04, *P*=.42) when groups were exact-matched on baseline PHQ-9 sum score mean (n=10,757 per group; mean difference: t[21512]=0, P=1; variance difference: F=1.00, P=1). The symptom 'depressed mood' remained the most central symptom in both the Responder and Non-Responder networks (1.18 vs 1.18, *P*=.95), while sleep (Responder: 0.62 vs. Non-Responder: 0.72, *P*=.02) and 'psychomotor problems' (Responder: 0.62 vs. Non-Responder: 0.73, *P*=.02) were the only symptoms that were significantly different in strength centrality between the cohorts.

Due to the smaller sample size, we were only able to draw 90 non-overlapping samples of n=250 Responders and n=250 Non-Responders, with 45 subsamples per group. We once again found that the 45 networks of Non-Responders had greater network connectivity than that of Responders at baseline (β=-1.07, SE=0.18, *P*<.001), and that the constructs of baseline PHQ-9 sum score mean, PHQ-9 sum score variance, and network connectivity were still positively correlated in the networks of both cohorts (r=0.37-0.70, all *P*<.05). Group differences in the connectivity of these networks, did not survive after controlling for either baseline PHQ-9 sum score mean (β=-0.24, SE=0.26, *P*=.34) nor PHQ-9 sum score variance (β=0.23, SE=0.20, *P*=.25). Results from our parametric analyses of symptom-level data were in line with that of our main analyses; each symptom strength centrality were higher in the Non-Responder networks than that of Responders (all *P*<.001). Baseline PHQ-9 sum score mean remained the most predictive of response

status (β=-1.53, SE=0.14, *P*<.001), followed by PHQ-9 sum score variance (β=-1.50, SE=0.14, *P*<.001), with both non-network metrics being more predictive than network connectivity (β=-1.07, SE=0.18, *P*<.001). Similar to that of our main analyses, the severity of most individual symptoms were more predictive of treatment response than its centrality, with the exception of 'depressed mood' (mean β=-0.60, SE=0.20, *P*=*.003*; strength β=-0.89, SE=0.19, *P*<.001) and worthlessness (mean β=-0.22, SE=0.21, *P*=*.31*; strength β=-0.49, SE=0.21, *P*=.02), and the mean of 6/9 symptoms had higher predictive utility than the strength of 'depressed mood' which was the most central symptom in both networks (all *P*<.001).

8.4.2 Results on Anxiety Symptom Networks (GAD-7)

To investigate the generalisability and specificity of the main findings of our study, we applied the same sampling (**eFigure 8.4.7**) and analytical procedures to anxiety symptom networks constructed using all seven items from the GAD-7 (N=70,620).(Spitzer et al., 2006b) This sample partially overlapped with that of the main analysis, as both include patients receiving the combined depression-anxiety iCBT program. However, it is different in that it also includes those who completed anxiety-only iCBT programs and excluded individuals who completed depression-only programs.

As per our main analyses with depression networks, at baseline, the full sample (n=35,310) non-Responder network had greater connectivity than the full sample (n=35,310) Responder network (2.68 vs 2.42, S=0.26, *P*<.001). Out of 21 estimated edges, seven were significantly different between-groups (all *P*<.05). Subsequent to exact-matching on baseline GAD-7 total scores (n=31,086 per group, mean difference: t[62170]=0, P=1; variance difference: F=1.00, P=1)), this effect was substantially reduced (S=0.26 reduced to S=.04), though connectivity differences remained between Non-Responders and Responders (2.46 vs. 2.42, S=0.03, *P*=.01). With regards to strength centrality, the same symptom that had the highest strength in both groups (i.e., not being able to stop worrying; 1.21 vs 1.20, *P*=.57).

For our parametric analyses, due to the larger full sample, we were able to draw 200 non-overlapping samples of n=350 Responders and n=350 Non-Responders (i.e., 100 subsamples per group). Overall, the 100 networks of Non-Responders were more connected than that of Responders (β=-1.67, SE=0.08, *P*<.001). We again found that GAD-7 sum score mean, GAD-7 sum score variance, and also network connectivity at

baseline were all positively correlated with each other in the networks of both Responders and Non-Responders (r=.027-0.75, all *P*<.01). Between-groups differences in network connectivity at baseline were no longer significant, when baseline GAD-7 sum score variance was accounted for in the model (β=-0.19, SE=0.12, *P*=.13), but not baseline GAD-7 sum score mean (β=-0.59, SE=0.20, *P*=.003), replicating the results of the main depression analysis. In line with our main findings from the parametric analyses on symptom-level data, the centrality of all symptoms were higher in the Non-Responder networks than that of Responders (all *P*<.001). Our parametric analyses in contextualising the effect sizes of these metrics for predicting response revealed that both baseline GAD-7 sum score mean (β=-1.86, SE=0.05, *P*<.001) and baseline GAD-7 sum score variance (β=-1.78, SE=0.06, *P*<.001) were more predictive than baseline network connectivity (β=-1.67, SE=0.08, *P*<.001). The severity of every individual symptom itself was still more predictive of treatment response than its centrality (all *P*<.05), and the mean of all symptoms had higher predictive utility than the most central symptom of not being able to stop worrying (all *P*<.001).

8.4.3 Results on Patients Excluded from Main Analyses

Baseline Descriptive Analyses for Patients who Did Not Meet Response Criteria. The main analyses focused on patients who met caseness for depression (≥10 on the PHQ-9) and received treatment (4-8 weeks clinician-guided, depression-relevant iCBT) (i.e., patients who met Criterion 1 to 4, outlined in **Figure 5.1A** in main manuscript; N=66,970). Further, the sample was filtered to include those who satisfied our response criteria (i.e., Criterion 5), with both Responder and Non-Responder groups matched on treatment days (i.e., Criterion 6), resulting in a final N=40,518). While patient demographics of the study sample was not made available to the authors, we explored some basic descriptive of patients who were excluded based on our response criteria (i.e., did not qualify as a 'Responder' or 'Non-Responder') and were not matched for inclusion in our final sample (N=26,452).

At baseline, excluded patients, on average, presented moderately severe depression scores on the PHQ-9 and severe anxiety scores on the GAD-7. There was no significant difference in baseline depression severity between excluded patients (M=15.78, SD=4.63) and the main study sample (M=15.80, SD=3.83), t(66968)=0.53, *P*=.60. However, excluded patients were significant more anxious at baseline (M=14.65, SD=3.78) than those included (M=14.52, SD=3.67), t(41007)=-3.48, *P*<.001. Both groups on average

underwent approximately 6 weeks of iCBT treatment, with the main study sample undergoing just one more day of treatment (M=44, SD=8.09) compared to excluded patients (M=43, SD=8.19), t(66968)=2.27, *P*=.02).

Results on Patients Receiving 1-4 Weeks of iCBT. Our main analyses focused on patients who received a minimal dose of iCBT (i.e., 4-8 weeks) as the focus of the study was to determine if baseline network characteristics are associated with clinical changes following treatment. Nonetheless, we tested whether our findings generalised to patients who disengaged from treatment early. We re-ran our main analysis on the subset of patients with initial caseness who completed at least 1 week of iCBT but dropped out / disengaged before 4 weeks of treatment. This sample therefore did not overlap with those in the main analysis (N=6552, n=3276 each in Responder and Non-Responder group).

The Non-Responder network of the full sample displayed higher network connectivity at baseline than the Responder network (3.23 vs. 2.39, S=0.84, *P*<.001), which was consistent with the main analyses. Out of the 36 edges examined, 7 were significantly different between-groups (all *P*<.05). When the groups were exactly matched on baseline PHQ-9 sum score mean (n=2639 per group), the connectivity differences between the Responders and Non-Responders vanished (2.46 vs. 2.59, S=0.12, *P*=.30) (mean difference: t[5726]=0, *P*=1; variance difference: F=1.00, *P*=1), again, similar to the primary results. In both Responder and Non-Responder networks, the symptom of 'depressed mood' remained the most central (1.17 vs. 1.18, P=.92). Symptoms that significantly differed in centrality between-groups included 'sleep', 'fatigue', 'concentration', 'appetite', and 'psychomotor problems', all of which exhibited higher centrality in the Non-Responders network.

With a much smaller sample size, we were only able to draw 42 non-overlapping samples of n=150 Responders and n=150 Non-Responders (21 subsamples per group). Baseline PHQ-9 sum score mean, PHQ-9 sum score variance, and network connectivity were all positively correlated in the networks of both cohorts (r=0.16-0.73), but only the correlations between PHQ-9 sum score variance and connectivity were significant for both groups (all *P*<.05), which may attribute to the lack of power. Overall, we found that the network connectivity of the Non-Responders was greater than that of Responders at baseline (β=-1.67, SE=0.17, *P*<.001); this difference disappeared yet again when PHQ-9 sum score variance was included as a covariate in the model (β=0.15, SE=0.33, *P*=.64), not PHQ-9 sum score mean (β=-1.22, SE=0.44, *P*=.009). Each symptom strength

centrality was higher in the Non-Responder networks than in the Responder networks (*P*<.05). Both non-network metrics of baseline PHQ-9 sum score variance (β=-1.83, SE=0.11, *P*<.001) and baseline PHQ-9 sum score mean (β=-1.83, SE=0.12, *P*<.001) were the most predictive of response status (β=-1.83, SE=0.11, *P*<.001) and performed better than network connectivity (β=-1.67, SE=0.17, *P*<.001). The severity of each individual symptoms was also more predictive of treatment response than their strength centrality (all *P*<.001).

8.4.4 Sensitivity Analyses with Alternative Response Definitions

The study operationalised depression treatment response using the IAPT criteria for reliable recovery (i.e., patients who initially met caseness at baseline must transition to non-caseness post-treatment, defined as ≥10 on the PHQ-9, while experiencing ≥6 point reduction from pre- to post-treatment). To test whether our results generalised to other definitions of treatment response, we repeated our core analyses with three levels of percent reduction in PHQ-9 from pre- to post-treatment.

Response Defined as ≥30% Reduction in PHQ-9 Sum Score From Pre- To Post-Treatment. Out of N=66,970 patients who met initial caseness and received a minimal dose of treatment (i.e., 4-8 weeks), n=34,779 were classified as Responders and n=32,191 were classified as Non-Responders. After matching on treatment days, this yielded an equal sample size of n=32,191 in both cohorts. The full-sample baseline Non-Responder network was more connected than Responders (3.33 vs. 3.12, S=0.22, *P*<.001) with 9/36 edges were significantly different between-groups (all *P*<.05). When groups were matched on baseline PHQ-9 sum score (mean difference: t[60454]=0, *P*=1; variance difference: F=1.00, *P*=1), this rendered the connectivity difference between Responders and Non-Responders non-significant (3.16 vs. 3.13, S=0.03, *P*=.11). The most central symptom for both groups was ‘depressed mood’, (Responder: 1.19; Non-Responder: 1.25, *P*=.02). ‘Loss of Interest/Pleasure’ was more central in the Responder network (0.92 vs. 0.85, *P*=.01), while ‘sleep’ (0.78 vs. 0.70, *P*<.001) and ‘motor’ (0.74 vs. 0.69, *P*=.02) were more central in the Non-Responders network. For the parametric analyses, 256 non-overlapping samples of n=250 Responders and n=250 Non-Responders (128 subsamples each group) were drawn. Baseline PHQ-9 sum score mean, PHQ-9 sum score variance, and network connectivity were positively correlated (r=0.40-0.79, all *P*<.001). There were significant group differences in connectivity (β=-1.14, SE=0.10, *P*<.001), but this difference went away when we controlled for PHQ-9 sum score variance in the model

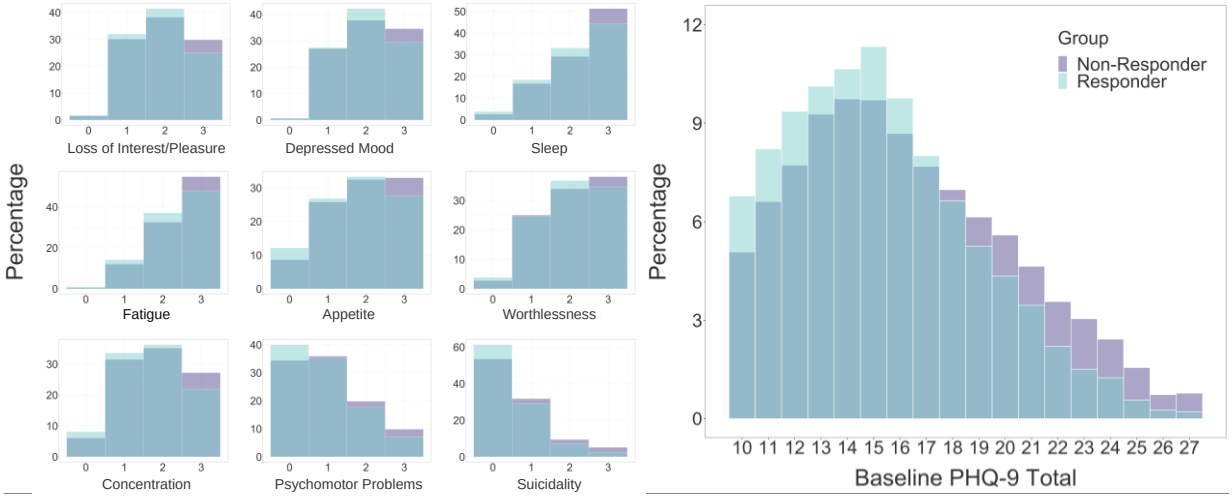
(β=0.13, SE=0.09, *P*=.17), but not PHQ-9 sum score mean (β=-0.61, SE=0.12, *P*<.001). Both PHQ-9 sum score mean (β=-1.24, SE=0.10, *P*<.001) and PHQ-9 sum score variance (β=-1.42, SE=0.09, *P*<.001) predicted treatment response better than network connectivity (β=-1.14, SE=0.10, *P*<.001). The item severity of each symptom was also more predictive of response than their strength centrality, apart from 'depressed mood' (mean: β=-0.27, SE=0.12, *P*=.03; strength: β=-0.62, SE=0.12, *P*<.001) and 'worthlessness' (mean: β=-0.11, SE=0.13, *P*=.34; strength: β=-0.36, SE=0.12, *P*=.003).

Response Defined as ≥50% Reduction in PHQ-9 Sum Score from Pre- To Post-Treatment. Out of N=66,970 patients who met initial caseness and received 4-8 weeks of treatment, n=20,176 were classified as Responders and n=46,794 were classified as Non-Responders. After matching on treatment days, both groups had a sample size of n=20,176. The baseline full-sample non-Responder network was more connected than Responders (3.28 vs. 3.11, S=0.17, *P*<.001) with 9/36 edges significantly different from each other (all *P*<.05). After matching on baseline PHQ-9 sum score (mean difference: t[38050]=0, *P*=1; variance difference: F=1.00, *P*=1), this difference went away (Responder: 3.13 vs. Non-Responder: 3.14, S=0.006, *P*=.78). 'Depressed mood' was the most central symptom for both networks (R: 1.21 vs. NR: 1.22, P=.74). Both 'loss of interest/pleasure' (Responder: 0.94 vs. Non-Responder: 0.86) and 'worthlessness' (Responder: 0.94 vs. Non-Responder: 0.84) were more central for Responders than Non-Responders (all *P*<.05). We drew 160 non-overlapping samples of n=250 Responders and n=250 Non-Responders (80 subsamples each group) for the parametric analyses. Baseline PHQ-9 sum score mean, PHQ-9 sum score variance, and network connectivity were positively correlated (r=0.39-0.77, all *P*<.01). Although there were significant differences in baseline network connectivity between-groups (β=-1.01, SE=0.14, *P*<.001), this difference disappeared after we accounted for PHQ-9 sum score variance (β=-04, SE=0.11, *P*=.71), but not PHQ-9 sum score mean (β=-0.66, SE=0.16, *P*<.001). Both PHQ-9 sum score mean (β=-1.11, SE=0.13, *P*<.001) and PHQ-9 sum score variance (β=-1.20, SE=0.13, *P*<.001) were better predictors of response than network connectivity (β=-1.11, SE=0.13, *P*<.001). The severity of each item was also more predictive of response than their strength.

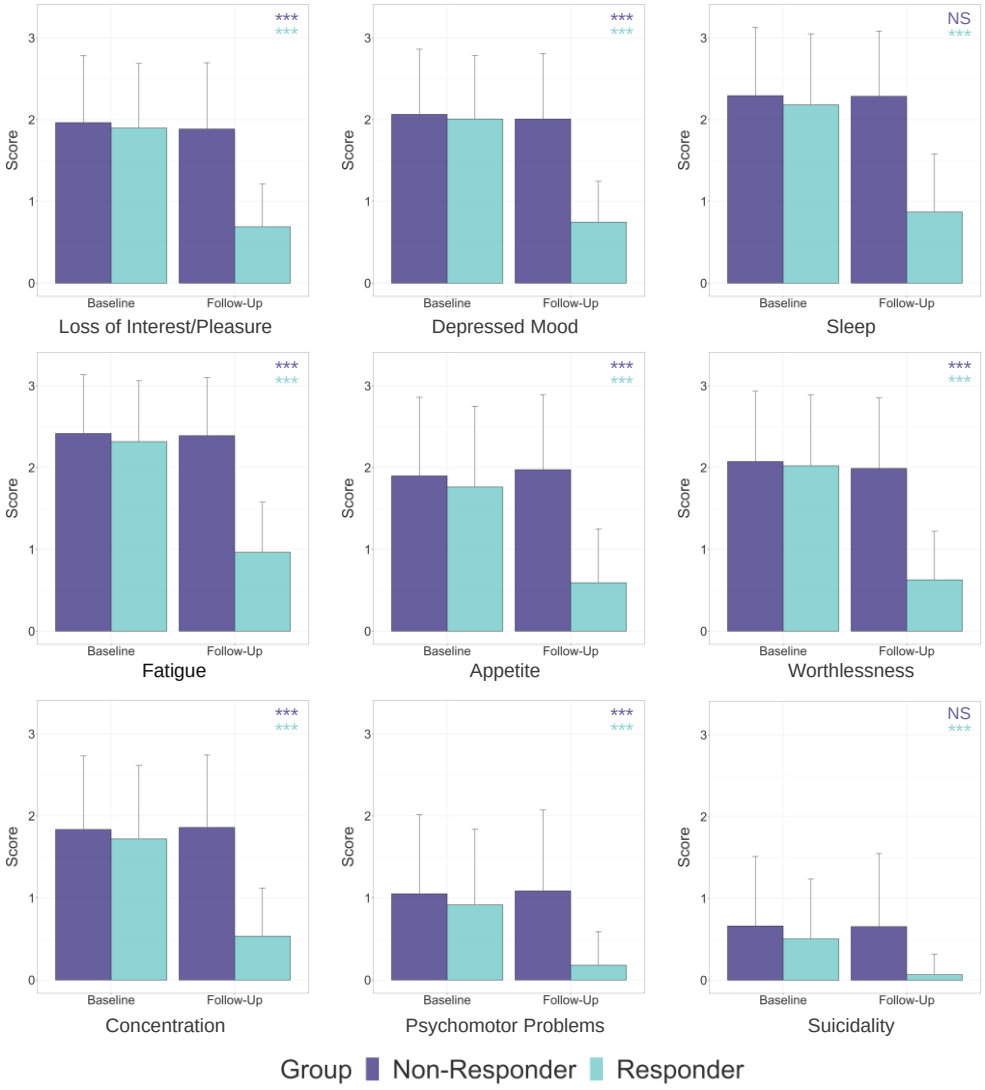
Response Defined as ≥70% Reduction in PHQ-9 Sum Score from Pre- To Post-Treatment. Out of N=66,970 patients who met initial caseness and received 4-8 weeks of treatment, n=8719 were classified as Responders and n=58,251 were classified as Non-

Responders. After matching based on treatment days, both cohorts consisted of n=8719. At baseline the full Non-Responder network was more connected than Responders (3.21 vs. 3.05, S=0.17, *P*<.001). 4/36 edges were significantly different between-groups (all *P*<.05). When groups were matched on baseline PHQ-9 sum score (mean difference: t[16060]=0, *P*=1; variance difference: F=1.00, *P*=1, this rendered the connectivity difference non-significant (Responder: 3.04 vs. Non-Responder: 3.08, S=0.04, *P*=.34). 'Depressed mood' remained the most central symptom for both networks (Responder: 1.19 vs. Non-Responder: 1.21, P=.77), but only 'worthlessness' (Responder: 0.96 vs. Non-Responder: 0.9221, P=.03) was significantly different between-groups. In our parametric analyses, we drew 68 non-overlapping samples of 250 Responders and 250 Non-Responders (34 subsamples each). We found positive correlations between baseline PHQ-9 sum score mean, PHQ-9 sum score variance, and network connectivity (r=0.34-0.76, all *P*<.05). Although group differences in connectivity were significant (β=-1.09, SE=0.20, *P*<.001), controlling for PHQ-9 sum score mean in the model eliminated this difference (β=-1.16, SE=0.34, *P*=.65), but not PHQ-9 sum score variance (β=-0.39, SE=0.17, *P*=0.02). PHQ-9 sum score mean (β=-1.65, SE=0.14, *P*<.001) predicted treatment response better than network connectivity (β=-1.11, SE=0.20, *P*<.001). Item severity was also more predictive of response than strength centrality for each symptom, apart from 'depressed mood' (mean: β=-0.49, SE=0.24, *P*=.04; strength: β=-0.64, SE=0.23, *P*=.008).

eFigure 8.4.1. Distribution of PHQ-9 item scores of Responders and Non-Responders at baseline.

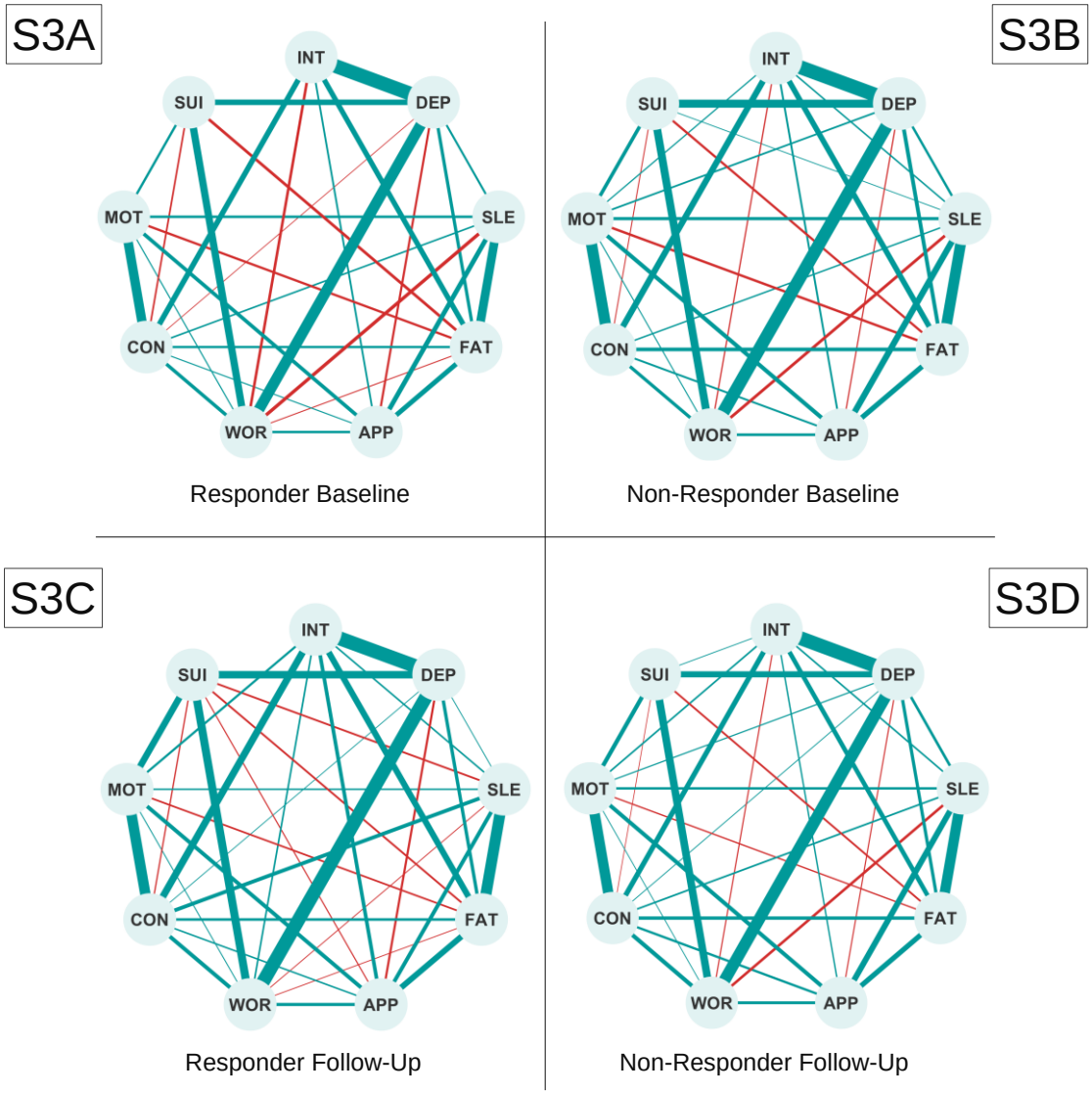


eFigure 8.4.2. Pre-Post treatment PHQ9 item scores by responder group



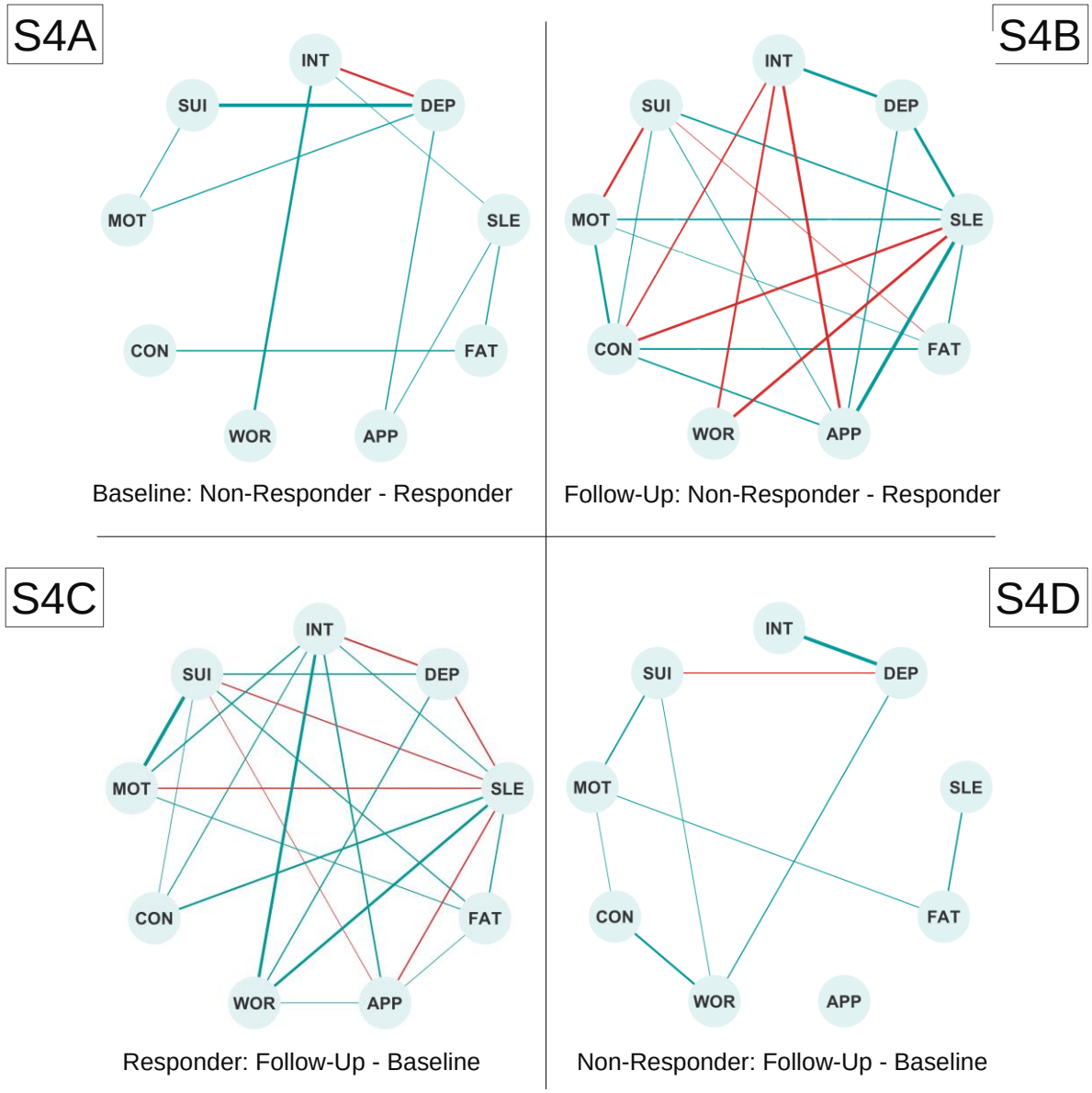
Responders exhibited a significant decrease in all PHQ-9 symptoms from baseline to follow-up. Non-Responders improved on some PHQ-9 items, including 'loss of interest/pleasure', 'depressed mood' , 'fatigue', and 'worthlessness' (all *P.adj*<.001), but worsened in 'appetite', 'concentration', and 'psychomotor problems' (all *P.adj*<.001) and did not change in 'sleep' problems (*P.adj*=.19) nor 'suicidality' (*P.adj*= .13). Error bars denote standard deviations from the mean score.

eFigure 8.4.3. Network visualisations of Responders and Non-Responders at baseline and follow-up



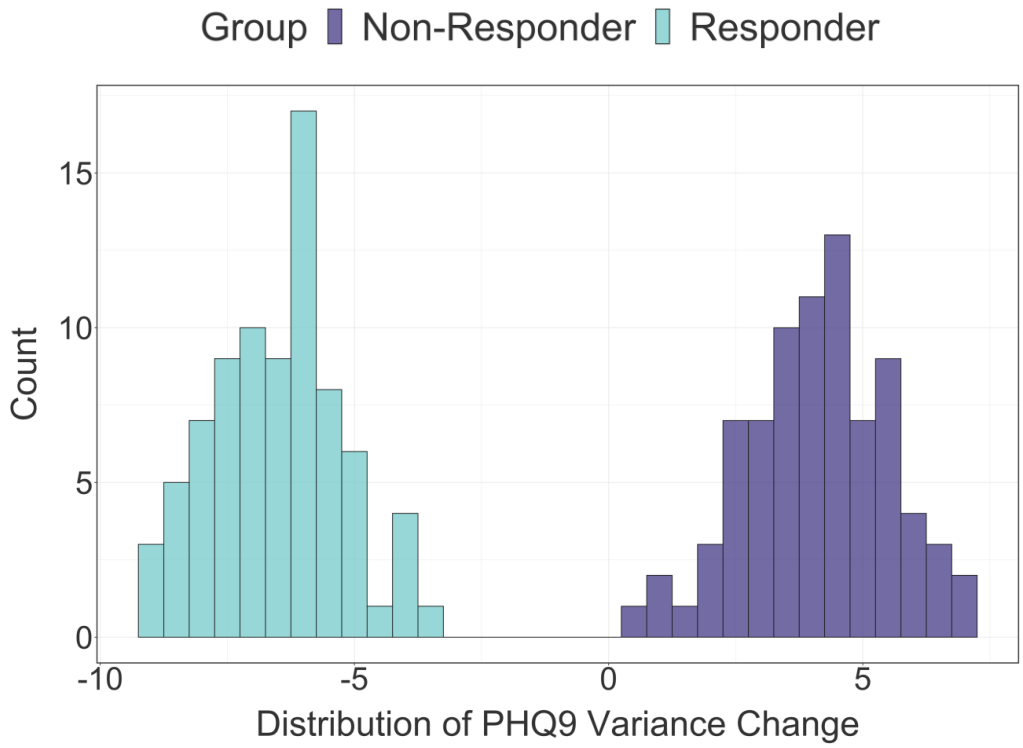
S3A and S3B visualises Responder and Non-Responder networks at baseline, respectively (also reported in the main text; reposted here for visual comparison to follow-up). S3C and S3D visualise Responder and Non-Responder networks at follow-up. Green edges denote positive partial correlations between 2 symptom nodes, while red edges denote negative partial correlations. NOTE: INT = 'loss of interest/pleasure', DEP = 'depressed mood', SLE = 'sleep', FAT = 'fatigue', APP = 'appetite', WOR = 'worthlessness', CON = 'concentration', MOT = 'psychomotor problems', SUI = 'suicidality'.

eFigure 8.4.4. Edge-difference comparison between- and within-groups



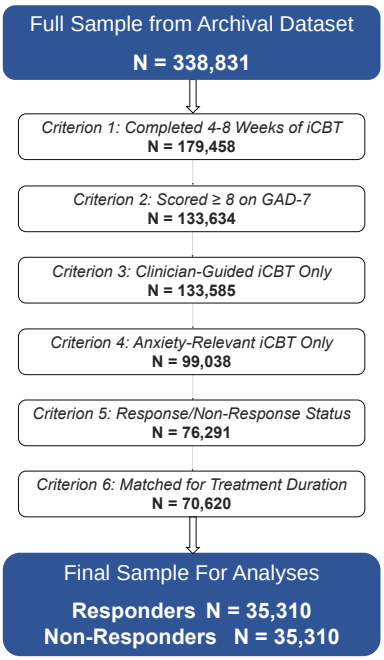
Network of edge-differences showing only significant edges. S4A and S4B visualises edge strength differences between Responder and Non-Responder networks at baseline and follow-up, respectively. Green edges denote stronger edge strength in the Non-Responder network, while red edges denote stronger edge strength in the Responder network; S4C and S4D visualises edge strength differences pre-post network differences for Responders and Non-Responders, respectively. Green edges denote stronger edge strength in the baseline network, while red edges denote stronger edge strength in the follow-up network. NOTE: INT = 'loss of interest/pleasure', DEP = 'depressed mood', SLE = 'sleep', FAT = 'fatigue', APP = 'appetite', WOR = 'worthlessness', CON = 'concentration', MOT = 'psychomotor problems', SUI = 'suicidality'.

eFigure 8.4.5. Distribution of PHQ-9 sum score variance change pre-post treatment by responder group



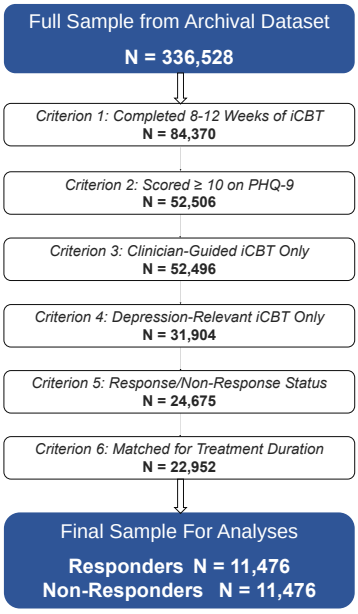
Distributions of pre-post treatment changes in PHQ-9 sum score variance for 160 subsamples, split by responder group. Positive values on the x-axis reflect an increase in PHQ-9 sum score variance pre-post treatment. Conversely, the greater the negative values on the x-axis, the bigger the decrease in PHQ-9 sum score variance pre-post treatment.

eFigure 8.4.6. Sampling procedures for analyses on patients undergoing 8-12 weeks of iCBT.



Final study sample flow chart with inclusion and exclusion criteria for analyses pertaining to patients undergoing 8-12 weeks of iCBT.

eFigure 8.4.7. Sampling procedures for analyses using GAD-7 networks.



Final study sample flow chart with inclusion and exclusion criteria for analyses pertaining to networks constructed using the Generalised Anxiety Disorder-7 (GAD-7).(Spitzer et al., 2006b) The Responder status was determined if patients 1) transitioned from 'caseness' to 'non-caseness' post-treatment, defined by the clinical cut-off of 8 on the measure, and 2) if their score reduction was greater than the Reliable Change Index of ≥4 on the GAD-7.(Jacobson & Truax, 1991) Patients were classified as Non-Responders if they met neither of these criteria, and those who met only one of these criteria were removed from analyses.

eTable 8.4.1. Comparisons of baseline and follow-up PHQ-9 sum and item score means and variances of Responders and Non-Responders.

Group	PHQ-9 Variable	Timepoint		t / F (p.value)
		Baseline	*Follow-Up*	
Responder				
	Loss of Interest/Pleasure			
	Mean	1.90	0.69	203.53 *(P<.001)*
	Variance	0.62	0.28	2.24 *(P<.001)*
	Depressed Mood			
	Mean	2.01	0.75	218.75 *(P<.001)*
	Variance	0.60	0.25	2.41 *(P<.001)*
	Sleep			
	Mean	2.18	0.87	199.87 *(P<.001)*
	Variance	0.75	0.51	1.49 *(P<.001)*
	Fatigue			
	Mean	2.32	0.96	234.76 *(P<.001)*
	Variance	0.56	0.38	1.48 *(P<.001)*

	Appetite			
	Mean	1.76	0.59	176.18 *(P<.001)*
	Variance	0.98	0.43	2.24 *(P<.001)*
	Worthlessness			
	Mean	2.02	0.63	229.91 *(P<.001)*
	Variance	0.75	0.35	2.13 *(P<.001)*
	Concentration			
	Mean	1.72	0.53	192.85 *(P<.001)*
	Variance	0.80	0.34	2.34 *(P<.001)*
	Psychomotor Problems			
	Mean	0.92	0.18	121.39 *(P<.001)*
	Variance	0.85	0.17	5.08 *(P<.001)*
	Suicidality			
	Mean	0.51	0.07	89.63 *(P<.001)*
	Variance	0.54	0.06	8.37 *(P<.001)*
	PHQ-9 Total			
	Mean	15.33	5.27	413.24 *(P<.001)*
	Variance	12.70	6.20	2.05 *(P<.001)*
Non-Responder				
	Loss of Interest/Pleasure			
	Mean	1.96	1.89	12.59 *(P<.001)*
	Variance	0.67	0.66	1.02 *(P=.24)*
	Depressed Mood			
	Mean	2.06	2.01	9.66 *(P<.001)*
	Variance	0.63	0.64	0.99 *(P=.39)*
	Sleep			
	Mean	2.29	2.29	1.32 *(P=.19)*
	Variance	0.70	0.64	1.10 *(P<.001)*
	Fatigue			
	Mean	2.42	2.39	5.03 *(P<.001)*
	Variance	0.52	0.50	1.03 *(P=.06)*
	Appetite			
	Mean	1.90	1.97	-11.82 *(P<.001)*
	Variance	0.92	0.84	1.09 *(P<.001)*
	Worthlessness			
	Mean	2.07	1.99	14.12 *(P<.001)*
	Variance	0.74	0.75	0.99 *(P=.39)*
	Concentration			
	Mean	1.84	1.86	-3.99 *(P<.001)*
	Variance	0.80	0.78	1.03 *(P=.04)*
	Psychomotor Problems			
	Mean	1.05	1.09	-5.86 *(P<.001)*
	Variance	0.93	0.97	0.96 *(P=.004)*
	Suicidality			
	Mean	0.66	0.66	1.53 *(P=.13)*
	Variance	0.72	0.80	0.91 *(P<.001)*
	PHQ-9 Total			
	Mean	16.26	16.14	5.38 *(P<.001)*
	Variance	16.21	20.30	0.80 *(P<.001)*

Note: All p-values indicated above PHQ-9 items have been adjusted for multiple significance testing using the Hochberg method.

eTable 8.4.2. Raw and standardised centrality values of Responders and Non-Responders at baseline.

PHQ-9 Variable	Raw Strength Values		Standardised Strength Values		p
	Responders	Non-Responders	Responders	Non-Responders	
Loss of Interest/Pleasure	0.89	0.84	0.58	0.22	0.047*
Depressed Mood	1.18	1.22	1.94	2.22	0.167
Sleep	0.66	0.74	-0.54	-0.30	0.020*
Fatigue	0.87	0.90	0.46	0.53	0.416
Appetite	0.55	0.57	-1.05	-1.17	0.637
Worthlessness	0.93	0.83	0.75	0.20	0.004**
Concentration	0.71	0.74	-0.32	-0.29	0.244
Psychomotor Problems	0.61	0.71	-0.81	-0.42	0.002**
Suicidality	0.57	0.61	-0.99	-0.99	0.242

eTable 8.4.3. Ranked comparison of significantly different edges between the Responder network and Non-Responder network at baseline

Edge	Local Connectivity		Difference in Local Connectivity	p
	Responders	Non-Responders		
Loss of Interest/Pleasure-Depressed Mood	0.47	0.42	-0.05	< 0.001***
Loss of Interest/Pleasure-Sleep	0.00	0.03	0.03	0.034*
Sleep-Fatigue	0.28	0.31	0.03	0.007***
Depressed Mood-Appetite	-0.05	-0.02	0.03	0.017*
Sleep-Appetite	0.14	0.17	0.03	0.023*
Psychomotor Problems-Suicidality	0.06	0.09	0.03	0.032*
Fatigue-Concentration	0.06	0.09	0.03	0.007**
Depressed Mood-Psychomotor Problems	0.00	0.03	0.03	0.001**
Loss of Interest/Pleasure-Worthlessness	-0.07	-0.02	0.05	< 0.001***
Depressed Mood-Suicidality	0.17	0.23	0.06	< 0.001***

Note: Negative values in column 'Difference in Local Connectivity' indicate the edge shows stronger connectivity in the Responder network, while positive values indicate stronger connectivity in the Non-Responder network.

eTable 8.4.4. Ranked comparison of significantly different edges between baseline and follow-up networks of Responders

Edge	Local Connectivity		Difference in Local Connectivity	p
	Baseline	Follow-Up		

Loss of Interest/Pleasure-Depressed Mood	0.47	0.41	-0.06	< 0.001***
Sleep-Suicidality	0.00	-0.04	-0.04	< 0.001***
Depressed Mood-Sleep	0.05	0.006	-0.04	0.001**
Sleep-Appetite	0.14	0.10	-0.04	0.001**
Sleep-Psychomotor Problems	0.07	0.03	-0.04	0.007**
Appetite-Suicidality	0.00	-0.01	-0.01	0.033*
Sleep-Worthlessness	-0.09	-0.005	0.01	< 0.001***
Fatigue-Suicidality	-0.07	-0.04	0.03	0.007
Fatigue-Psychomotor Problems	-0.06	-0.03	0.03	< 0.001***
Depressed Mood-Worthlessness	0.36	0.39	0.03	< 0.001***
Loss of Interest/Pleasure-Sleep	0.00	0.03	0.03	0.049*
Sleep-Fatigue	0.28	0.32	0.04	0.002**
Depressed Mood-Suicidality	0.17	0.21	0.04	0.009
Loss of Interest/Pleasure-Appetite	0.04	0.09	0.05	< 0.001***
Loss of Interest/Pleasure-Psychomotor Problems	0.00	0.05	0.05	< 0.001***
Loss of Interest/Pleasure-Concentration	0.17	0.23	0.06	< 0.001***
Sleep-Concentration	0.03	0.10	0.07	< 0.001***
Loss of Interest/Pleasure-Worthlessness	-0.07	0.03	0.10	< 0.001***
Psychomotor Problems-Suicidality	0.06	0.18	0.12	< 0.001***

Note: Negative values in column 'Difference in Local Connectivity' indicate the edge shows stronger connectivity in the baseline network, while positive values indicate stronger connectivity in the follow-up network.

eTable 8.4.5. Ranked comparison of significantly different edges between baseline and follow-up networks of Non-Responders

Edge	**Local Connectivity**		**Difference in Local Connectivity**	**p**
	Baseline	**Follow-Up**		
Depressed Mood-Suicidality	0.23	0.20	-0.03	0.008**
Concentration-Psychomotor Problems	0.33	0.35	0.02	0.027*
Depressed Mood-Worthlessness	0.37	0.40	0.03	0.003**
Worthlessness-Suicidality	0.22	0.25	0.03	0.033*
Fatigue-Psychomotor Problems	-0.05	-0.02	0.03	0.036*
Sleep-Fatigue	0.31	0.35	0.04	0.006**
Psychomotor Problems-Suicidality	0.09	0.13	0.04	0.004**
Worthlessness-Concentration	0.08	0.13	0.05	< 0.001***
Loss of Interest/Pleasure-Depressed Mood	0.42	0.48	0.06	< 0.001***

Note: Negative values in column 'Difference in Local Connectivity' indicate the edge shows stronger connectivity in the baseline network, while positive values indicate stronger connectivity in the follow-up network.

eTable 8.4.6. Ranked comparison of significantly different edges between the Responder network and Non-Responder network at follow-up

Edge	Local Connectivity		Difference in Local Connectivity	p
	Responders	Non-Responders		
Sleep-Concentration	0.10	0.04	-0.06	< 0.001***
Sleep-Worthlessness	-0.005	-0.07	-0.06	< 0.001***
Loss of Interest/Pleasure-Appetite	0.09	0.03	-0.06	< 0.001***
Loss of Interest/Pleasure-Worthlessness	0.03	-0.02	-0.05	< 0.001***
Psychomotor Problems-Suicidality	0.18	0.13	-0.05	0.002**
Loss of Interest/Pleasure-Concentration	0.20	0.16	-0.04	0.021*
Concentration-Suicidality	-0.02	-0.001	-0.02	0.003**
Appetite-Suicidality	-0.01	0.00	-0.01	0.007**
Fatigue-Psychomotor Problems	-0.03	-0.02	0.01	0.049*
Fatigue-Suicidality	-0.03	-0.04	0.01	0.017
Depressed Mood-Appetite	-0.04	-0.02	0.02	< 0.001***
Appetite-Concentration	0.03	0.06	0.03	0.021*
Sleep-Psychomotor Problems	0.03	0.06	0.03	0.011*
Sleep-Fatigue	0.32	0.35	0.03	0.018*
Sleep-Suicidality	-0.04	0.00	0.04	< 0.001***
Fatigue-Concentration	0.06	0.10	0.04	0.010*
Concentration-Psychomotor Problems	0.30	0.35	0.05	< 0.001***
Depressed Mood-Sleep	0.006	0.07	0.06	< 0.001***
Loss of Interest/Pleasure-Depressed Mood	0.41	0.48	0.07	< 0.001***
Sleep-Appetite	0.10	0.19	0.09	< 0.001***

Note: Negative values in column 'Difference in Local Connectivity' indicate the edge shows stronger connectivity in the Responder network, while positive values indicate stronger connectivity in the Non-Responder network.