Development of a personalised approach to clinical decision making in psychological treatment services using routine patient data.

Rob Saunders

University College London

PhD

Declaration

I, Rob Saunders, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis is concerned with the development of a personalised treatment approach to aid clinical decisions in psychological interventions provided for common mental health disorders (CMHDs), such as depression and anxiety disorders. It begins with a discussion of personalised medicine in healthcare and its potential for optimising care for CMHDs. The thesis then considers how personalised medicine can be used to inform clinical decision making, specifically clinical decisions in relation to the delivery of treatment in mental health services. This includes a description of the types of clinical decisions required, as well as examples from across healthcare that have used decision support tools (DSTs) to aid clinical judgement. This is followed by a review of patient characteristics that have been associated with outcomes in CMHD treatment. The review is supplemented by an analysis of a large dataset (n=10693) of patients receiving psychological treatment for CMHDs. It explores the associations between routinely available patient characteristics and outcomes. The thesis then reports on the use of latent profile analysis using the patient characteristics to identify statistically distinct sub-groups (profiles) of patients, and considers the variation in treatment outcomes between profiles and by the intensity of treatment. The change in depression and anxiety symptoms, as measured at every treatment session, is statistically modelled to identify different trajectories of change within and between the latent profiles. These trajectories represent differential response to psychological treatment. Information from the identified profiles is combined with the within treatment change methods to develop a personalised treatment approach to decisions about appropriate treatment and also clinical decisions during the course of treatment. The thesis then presents a prototype algorithm that can identify profiles and the likely trajectories of change pre-treatment, before discussing the clinical implications of providing this algorithm in routine care, as well as future directions of research.

Acknowledgements

I would firstly like to thank all the staff and patients from the Improving Access to Psychological Therapies (IAPT) services for their time and participation, as well as to the British Psychological Society for the funding of the work covered in this thesis.

I have had the pleasure of working alongside a number of wonderful colleagues at UCL throughout the duration of this thesis, and I am privileged to have learnt from them all. In particular I would like to thank the contributions of:

Craig Whittington for methodological and statistical advice in the early years of this thesis.

Katherine Clarke and Laura Gibbon for their advice, motivation and help throughout.

Joshua Buckman who has provided a huge amount of advice on elements of this project, and whose comments have been invaluable.

John Cape and Judy Leibowitz for their IAPT-related knowledge and understanding of the IAPT dataset.

Pasco Fearon for his guidance and introduction to the world of structural equation modelling.

And especially to Stephen Pilling for his guidance, encouragement, patience and support throughout this project.

I would also like to take this opportunity to thank my family, especially my parents, for all their unending support and for helping me get to this point.

And lastly to Carly, for your constant understanding, support and advice. I couldn't have completed this without you, and I am looking forward to our next adventure together.

Contents

Figures

Tables

Chapter 1. Personalised medicine: Potential application in mental health disorders.

Personalised medicine

The potential for a more person-centred approach to the provision of healthcare, referred to as ‘personalised medicine’ or ‘stratified medicine’, to provide benefits for patients and treatment services has been increasingly recognised in recent years (Dzau & Ginsburg, 2016; Khoury & Galea, 2016). The goal of this approach is to make treatment decisions informed by the patient’s individual characteristics, tailored to their individual needs by selecting the most appropriate treatment to achieve the desired outcome. In the UK, NHS England has made personalised medicine a key priority and is in the process of producing a five year strategy (Keogh, 2015), with the government establishing a ‘personalised medicine all-party parliamentary (APP)’ group.

The personalised medicine initiative has been developed to complement the current evidence-based medicine approach to healthcare. The major benefits of the evidence-based treatment approach are that patients with similar conditions can be provided with treatments that have clear evidence for their effectiveness. For clinicians, best practice guidelines can be used to guide treatment decisions. However, evidence-based practice has been criticised for a ‘one-size fits all’ approach to patient care (Pencina & Peterson, 2016), as although the recommended treatment may be helpful for many patients with a specific condition, for others it is not. Therefore, a bridge between personalised care and evidence based practice has been sought by clinicians and researchers.

Personalised medicine has led to vast improvements in the treatment of many common cancers, specifically by identifying genetic markers of probable treatment response, for example the identification of BRCA-1 gene mutations and response to chemotherapies for breast and ovarian cancers (Kennedy et al., 2004). Specific interventions are therefore targeted for certain patients who are grouped together based on similar characteristics (e.g. genetic markers). This approach has led to improvements in both response and remission times in cancer treatment (Schwaederle et al., 2016).

Researchers investigating treatments for cancers appear to be leading the medical fields in the use of personalised medicine, however those in other fields of medicine have also made significant improvements to treatments. For example, the use of genetic testing in cystic fibrosis has helped guide treatment decisions for many years and has become the ‘de facto’

approach (Ashley, 2015). Pharmacogenetic profiling such as this is able to guide decisions regarding appropriate dosing or specific pharmacological agents that are likely to be most successful based on the patient's genotype (Pirmohamed, 2014). Stratifying patients in this manner and providing this information to clinicians has the potential to be used to inform treatment decisions, reducing risk of negative outcomes and minimising harm to patients.

Much of the extant literature on personalised medicine has included the use of genetic information and other biomarkers which are largely unavailable or as yet unidentified in mental health fields. However, there are other potential patient characteristics that may be used to personalise healthcare. Information which can be gathered at assessment such as knowledge of comorbidities (physical or mental), current symptoms, family history of the condition or related conditions, and demographic information have been used to identify individuals who may benefit from a preventive intervention, for example due to an increased risk of suicide (Barak-Corren et al., 2017) or the short-to-medium term risk of cardiovascular events (Hippisley-Cox et al., 2007; Kumbhani et al., 2013; Lee, Flammer, Lerman, & Lerman, 2012).

Large data and digital technology

The rise in personalised medicine has coincided with the increased availability of large patient datasets which offer an opportunity to explore patient characteristics associated with health conditions and treatment across large cohorts. Such datasets provide more power for statistical modelling and have previously been utilised to develop personalised approaches in healthcare (Abbasi, 2017). The human genome project can be viewed as one of the major landmarks in personalised medicine, and it is argued that the availability of such a huge amount of patient data was critical to the growth of this patient-centred approach (Wilson & Nicholls, 2015). Many of the advances in the personalised treatment of cancers and cardiovascular diseases have been achieved by using large data sets (Hippisley-Cox et al., 2007; Verma & Mukesh, 2012). In other healthcare fields, especially mental health, there has been limited use of large data to inform personalised medicine, although research has been conducted using large datasets to identify characteristics associated with diagnoses. For example, the 23andMe study used data from over 500,000 patients and identified 15 genes associated with depression (Hyde et al., 2016).

The use of digital technology has the potential to further aid the delivery of personalised medicine, and can provide benefits to both patients and services (Richards, Coulter, & Wicks, 2015). Electronic patient management systems (EPMS), are now widely used across healthcare settings and in the first instance provide a method of collecting and storing the same specific characteristics and clinical records across large numbers of patients (Menachemi & Collum, 2011). In addition, the use of digital technology such as mobile phone and tablet apps provides opportunities for patients to conveniently store personal

information that can be useful clinically, and are potentially more ecologically valid than information collected within healthcare settings (Proudfoot, 2013).

To enable the findings from statistical modelling of patient datasets to be made available to clinicians in a way that can aid clinical practice, decision support tools (DSTs) and algorithms that can be used within treatment services are increasingly developed for healthcare (Sheehan & Sherman, 2012). These decision aids can take the form of sequenced decision charts such as the NICE stroke and transient ischaemic attack (TIA) algorithms (NICE, 2008) or risk prediction rules, for example the QRISK (Hippisley-Cox et al., 2007), which predicts the risk of cardiovascular events for an individual patient over a ten year period, and have therefore become critical to aiding clinical decisions on whether or not patients should start a statin to reduce their cardiovascular risk. These tools are already being routinely used by GPs and emergency health physicians to aid clinical decisions on preventative treatments.

Personalised medicine in mental health conditions

There are currently few examples of the use of personalised medicine approaches in mental health research and clinical settings (Ozomaro et al., 2013). Whereas genetic markers have been found to reliably aid treatment selection in a number of cancers, there has been limited evidence associating genotyping with improved outcomes in the treatment of common mental health disorders (CMHDs) (Munafò, Zammit, & Flint, 2014; Simon & Perlis, 2010). A number of genetic markers have been investigated in relation to CMHDs, primarily serotonin receptors (Papakostas & Fava, 2008), but the lack of convincing evidence for specific genes involved in the cause of these disorders, let alone such markers being implicated in differential response to treatments, has limited the use and utility of this research for the personalisation of CMHD treatment in routine settings (Cuijpers, 2014; Licinio & Wong, 2011).

As genotyping has had limited success in informing the personalisation of treatments for CMHDs, perhaps information which is more readily available in routine care settings and which is more easily recorded than is genetic data, including information on patient characteristics such as demographics and clinical factors, could be more useful in aiding decision making for these conditions. Furthermore, as the cost and practical complications with acquiring biomarkers in routine treatment services at present precludes their use (Evans, et al., 2006), using self-reported patient data to personalise treatment may have greater practical and clinical utility. This next section introduces CMHDs and discusses the issues and limitations of clinical diagnosis in informing treatment decisions. The outcomes for recommended treatment options for these conditions are discussed along with the current service configuration of mental health services providing psychological interventions in the UK.

Common Mental Health Disorders

This section briefly introduces the common mental health disorders (CMHDs), the focus of this thesis. For more information on formal diagnostic criteria please refer to the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10; WHO, 1992) or the Diagnostic and Statistical Manual of Mental Disorders, 5th revision (American Psychiatric Association, 2013).

Depression

Depression is characterised by a loss of pleasure and interest in most activities, lowered mood and a reduction in energy. Other symptoms of depression include poor concentration, increased tiredness, trouble sleeping, changes in appetite, agitation, irritability, tearfulness, social withdrawal and decreased libido. A reduction in self-esteem and self-confidence, feelings of worthlessness, guilt and suicidal ideation may be present. Due to the varying nature of the symptoms, depression is often considered as occurring on a continuum (Lewinsohn, et al., 2000).

Panic Disorder

Panic attacks can occur in the context of many mental health disorders, especially the anxiety disorders. Panic disorder is characterised by the experience of panic attacks or less severe 'limited symptom attacks' and a persistent worry of experiencing further panic attacks. A panic attack is a short period of intensive fear or anxiety, with symptoms including shaking, sweating, heart palpitations, nausea, chest pain and dizziness. There may be a secondary fear of losing control, going mad, fainting, losing control of the bladder or other bodily functions, or of dying during the panic attack. Panic disorder may or may not be associated with agoraphobia; an intense fear of being in a situation where escape is perceived as difficult or impossible, leading to avoidance or frequently only being able to go through the feared situations when accompanied by a particular trusted individual.

Generalised Anxiety Disorder

The main characteristic of generalised anxiety disorder (GAD) is excessive worry and anxiety that is not restricted to, or predominating in a particular circumstance. The worries are perceived as uncontrollable, and associated symptoms include difficulties with concentration, persistent nervousness, irritability, dizziness, palpitations, restlessness, physical tension in the body and sleep problems.

Social anxiety disorder

Social anxiety disorder is characterised by fear of scrutiny or ill-judgement from others, often resulting in the total avoidance of social situations or the use of implicit or explicit 'safety behaviours' to get through/survive the social situation(s). The fear is often of humiliation or embarrassment and the individual will tend to experience significant anxiety before, during and after social situations. Social anxiety disorder will often affect educational or occupational functioning and is frequently associated with lower self-esteem.

Obsessive-compulsive disorder

The main features of Obsessive-Compulsive Disorder (OCD) are recurrent obsessional thoughts and compulsive behaviours. The obsessions cause distress and functional impairment to daily life as the patient tries to resist them. Obsessions are intrusive thoughts, impulses or images that are unwanted and perceived as uncontrollable. Examples include thoughts of contamination from dirt, germs or bodily fluids, and thoughts of being responsible for harming others, or for the occurrence of unwanted or terrible events. Compulsions are ritualistic behaviours or acts that the individual feels 'compelled' to perform again and again. These may be internalised mental acts that are more difficult to observe such as counting or repeating prayers in one's mind, or clearer overt actions, such as frequently washing hands, checking taps, or hoarding items.

Post-Traumatic Stress Disorder

Post-Traumatic Stress Disorder (PTSD) can develop in response to an event where the patient considers themselves or someone close to them to be at threat of serious harm or death, such as such as an assault, a motor vehicle accident, natural disasters or military combat. Sufferers of PTSD may experience symptoms in three domains – re-experiencing (such as having flashbacks or repeated nightmares about the event), hypervigilance (i.e. being easily startled and on-alert) and avoidance (of reminders of the event). These symptoms frequently affect sleep and concentration, and as with the other CMHDs above to meet diagnostic criteria the symptoms must be accompanied by considerable distress and functional impairment.

Specific phobias

The key feature of a specific phobia is a persistent fear of a specific situation or object, for example certain animals, flying, darkness, sight of blood or needles. The presence of the situation/object provokes an anxiety response that is out of proportion to the actual risk or danger of harm posed by the object. The sufferer will recognise that their response to the

stimuli is out of proportion with the actual threat, but extreme emotional responses are triggered, leading to the avoidance of the situation/object.

Impact and costs

The point prevalence of all CMHDs in the UK is approximately 16% (McManus et al., 2009; Spiers et al., 2016), which would imply that a very large number of individuals may be in need of treatment from mental health services at any given time. Depression was reported by 6.7% of individuals surveyed in the USA in a 12 month period (Kessler, Chiu, Demler, & Walters, 2005) and findings from the UK based Adult Psychiatric Morbidity Survey (Stansfield et al., 2016) suggest approximately 3.3% of adults reported symptoms commensurate with being in a major depressive episode when surveyed. Depression is estimated to account for between 12% and 14% of UK General Practice attendances (King, Nazareth, et al., 2008; Rait et al., 2009). Once treated, the recurrence of depression is common. For example, large cohort studies in the USA and Europe have found that between 60% and 85% of patients treated in specialist mental health settings were found to relapse after treatment (Hardeveld, Spijker, De Graaf, Nolen, & Beekman, 2010; Mueller et al., 1999).

The most common anxiety disorder is GAD, which was reported in 5.9% of the UK population in 2014 (Stansfield et al., 2016), and is estimated to be present in 8% of primary care patients (Wittchen, 2002). PTSD is estimated to affect between 3% and 4% of individuals in the UK (Fear et al., 2016; McManus et al., 2009). Panic disorder and OCD are the least common CMHDs, and estimated to affect 0.6% and 1.3% of the general population respectively (Stansfield et al., 2016).

Considering the prevalence of CMHDs, and the distress and dysfunction associated with them, they impact not only the sufferer but family and carers too. At the population level they result in significant societal costs with impact on occupational functioning and healthcare burden on services. CMHDs can have a major effect on performance in employment and evidence suggests that suffering from a CMHD may adversely affect one's ability to work (Mauskopf et al., 2009). Depression was found to be the fourth most common reason for loss of disability-adjusted life years (DALYS) worldwide and it is predicted that it will be the second most common reason in 2020 (Murray & Lopez, 1997). A report by the King's Fund (McCrone, Dhanasiri, Patel, Knapp, & Lawton-Smith, 2008) highlighted the loss of earning in 2007 due to depression as £5.82 billion and £7.7 billion from anxiety, with these figures predicted to rise to £9.19 and £12.15 billion respectively by 2026. The identification of best available interventions for individual patients could result in a significantly reduced societal burden, increased productivity and reduced psychological distress.

CMHDs also have a significant effect on interpersonal relationships. Depression can negatively affect marital and family relationships, and parental depression can lead to neglect of children, which may contribute to childhood difficulties and disturbances (Ramchandani & Stein, 2003). Suicide attempts, which can have a dramatic effect on social and family relationships, are between 4 and 20 times more likely in individuals with depression compared to the general population (Bostwick & Pankratz, 2000), and the lifetime risk of suicide in depression is estimated at around 15% (Guze & Robins, 1970). CMHDs are also associated with an increased frequency of poor physical health outcomes and reduced life expectancy (Chesney, Goodwin, & Fazel, 2014).

Given the social, occupational and clinical impacts of CMHDs, timely optimised treatments are required to improve the health of sufferers (Habert et al., 2016), but also to best use the available healthcare resources. The cost of depression and anxiety to mental health services in England during 2007 was estimated to be £2.92 billion, and projected to rise to £5.0 billion by 2026 (McCrone et al., 2008).

The negative impact of CMHDs on occupational and personal functioning as well as the costs associated with the treatment of these conditions suggest that more effective and efficient means of delivering treatments are required. The development of personalised medicine approaches to CMHD treatments could help identify groups of patients who are more or less likely to benefit from specific interventions than other patients, and could therefore result in improvements in CMHD treatment outcomes. This patient-centred approach could be used to tailor the delivery of mental health treatment, and potentially support a more efficient health service.

Treatment of CMHDs

Pharmacological interventions

Pharmacological interventions are the most frequently used treatment option for CMHDs, with estimates suggesting 46 million prescriptions for antidepressants were issued in the UK in 2011 (Spence, 2013). The most commonly prescribed antidepressants are selective serotonin reuptake inhibitors (SSRIs), which are prescribed both for depression and for many anxiety disorders, and there is good evidence that SSRIs act on the stress-adaption system of the brain which is involved in the maintenance of both depression and anxiety (Shelton & Brown, 2001).

Clinical trials of antidepressant treatments have frequently found that around 50% of patients show a clinical response to treatment (Papakostas & Fava, 2010; Trivedi et al., 2006). However analysis typically finds that the benefits of antidepressant treatment over those of placebo treatment are usually only seen in patients with more severe levels of symptoms

pre-treatment (Fournier et al., 2010). Although pharmacotherapy is considered a valuable treatment for CMHDs and is usually low cost as most drugs are now out of patent, many patients report unwanted side effects which are associated with poor treatment adherence and early termination of treatment.

Some side effects are common: around 15% of patients report dry mouth, dizziness or nausea while taking SSRIs (Ferguson, 2001). However, surveys have suggested that up to 75% of patients prescribed other classes of antidepressants (e.g. tricyclics) report dry mouth as a side effect of treatment (Uher, Farmer, et al., 2009), and drowsiness is reported by 17% of patients (Hu et al., 2004). If the distress caused by the side effects outweighs the perceived benefits of the treatment, patients may not adhere to their pharmacotherapy regime or may electively terminate the treatment. For this reason psychological interventions may be more desirable for patients as an alternative to pharmacological treatments, and are recommended for CMHDs in national guidelines (NICE, 2011c).

Psychological interventions

Psychological interventions are recommended as a first-line treatment option for many patients with depression and anxiety disorders, and show comparable effectiveness to pharmacological treatments in reducing symptoms and maintaining the positive effects of treatment (NICE, 2011c). For mild-moderate depression, a number of psychological therapy options are recommended including cognitive behaviour therapy (CBT) for which there is the greatest amount of evidence of effectiveness (NICE, 2009), behavioural activation (BA), counselling for depression, behavioural couples therapy, interpersonal psychotherapy (IPT) and short-term psychodynamic therapy (STPT), all of which are supported by the literature, though with less evidence than for CBT (van Hees, Rotter, Ellermann, & Evers, 2013). Mindfulness-based cognitive therapy (MBCT), CBT and IPT have all been found to be effective relapse-prevention interventions for patients at risk of relapse from depression (Clarke, Mayo-Wilson, Kenny, & Pilling, 2015). A combination of both psychological and pharmacological treatment is recommended for moderate to severe depression, or when psychological treatments have shown limited effectiveness (NICE, 2009).

Psychological interventions, particularly CBT are also the recommended first-line treatment option in UK clinical guidance for most anxiety disorders, and pharmacological interventions are only recommended when a patient has declared a preference for drug treatment instead of psychological treatment, or when recommended psychological interventions have already shown limited response (NICE, 2011c, 2013). For mild to moderate presentations of GAD and panic disorder (with or without agoraphobia), both facilitated (guided) and non-facilitated (pure) self-help interventions are recommended, in addition to psychoeducational groups (GAD only) due to the significant evidence-base for these interventions in the treatment of

these disorders. CBT is recommended for moderate to severe presentations of panic disorder or GAD with severe functional impairment.

Patients with mild to moderate presentations of OCD are recommended either individual CBT including exposure-focused therapy which can involve self-help materials or group CBT (NICE, 2011c). CBT or antidepressant medication is recommended for patients with moderate functional impairment associated with OCD, whereas a combination of CBT and antidepressants are recommended when there is severe impairment or significant comorbidity with other CMHDs. For patients with PTSD, trauma-focused CBT or eye movement desensitisation and reprocessing (EMDR) are recommended, whereas individual CBT is recommended as initial treatment for social anxiety disorder (NICE, 2013), with pharmacological treatment for either disorder only recommended when psychological treatment is declined.

A systematic review of patients attending primary care services suggested that psychological over pharmacological treatment are frequently preferred for depression, mainly due to concerns over side effects (Van Schaik et al., 2004), but matching patients to the treatment they prefer did not improve patient outcomes in this study. Although 'side effects' in the biological sense do not exist with psychological interventions, there is a risk of clinical harm and it is reported that approximately 5-10% of patients may display a worsening of clinical symptoms from baseline to the end of treatment, referred to as clinical deterioration (Boisvert & Faust, 2003; Crawford et al., 2016; Rozental, et al., 2017). The monitoring of symptoms during the course of treatment is therefore recommended, as it is with pharmacological interventions, to identify deterioration early and modify the treatment approach if required (NICE, 2011c).

Treatment outcomes

Given the substantial costs in treating CMHDs and long-term projected resource needed to provide care, selecting the most appropriate treatment can not only improve patients' wellbeing, but increase the efficiency of healthcare provision. The consequences of selecting an 'incorrect' treatment can have significant costs for both the individual and wider society due to the continued illness and the associated loss of productivity (McGrath et al., 2013). The use of a personalised treatment approach could inform which type of intervention is likely to result in the best patient outcomes.

Evaluations of psychological interventions delivered in routine care for CMHDs suggest that just over 45% of patients report a level of symptoms below the established clinical cut-off after treatment (HSCIC, 2015). A recent national evaluation of treatment outcomes found that recovery for anxiety disorders was slightly higher than for depression (when diagnosis was recorded in the data) (NHS Digital, 2016). These recovery rates are similar to those reported in controlled trials of psychological and pharmacological treatment for depression

(Luty et al., 2007; Papakostas & Fava, 2010) and anxiety disorders, for example PTSD (Morina et al., 2014), and appear higher than controlled trials for GAD (Fisher & Durham, 1999; Leichsenring et al., 2013). However, this suggests that for over half of patients the treatment has not reduced symptoms below levels that would indicate an absence of a CMHD. The variation in effectiveness of interventions between individual patients indicates that interventions can be effective for some groups of patients but not others, and therefore the potential benefit of personalised medicine is to identify which characteristics might help identify patients likely to experience this differential response.

Head-to-head comparisons of pharmacological and psychological treatments for CMHDs have suggested limited differences in effectiveness in trial populations. A recent individual patient-data meta-analysis reported no differences between CBT and antidepressants in treatment for depression either in terms of remission (scoring below a clinical threshold) or response (symptoms reducing per a pre-established absolute value or percentage) (Weitz et al., 2015). However, attrition is reported to be higher in pharmacological interventions than CBT (Vittengl et al., 2016), and the benefits of treatment have been found to last considerably longer after terminating CBT compared to pharmacotherapy (Cuijpers et al., 2013). In social anxiety disorder CBT has been found to be more efficacious than pharmacotherapy (Mayo-Wilson et al., 2014), and the comparative benefit of CBT over pharmacological treatment has also been found in the treatment of panic disorder (Roshanaei-Moghaddam et al., 2011) and for OCD (Lack, 2012).

Despite the few conditions for which there is evidence of a significant difference in the efficacy or effectiveness of psychological and pharmacological treatments for CMHDs, for most disorders the comparison of different types of psychological treatment (e.g. CBT vs IPT) tend to show equivalence in head-to-head trials. Therefore clinical guidance is often driven by cost-effectiveness analyses (e.g. National Institute for Health and Clinical Excellence, 2011). This has led to the recommendation of less resource intensive self-help and group interventions which have shown approximately equivalent clinical outcomes with high intensity treatments such as CBT for mild to moderate depression and several anxiety disorders (NICE, 2009, 2011c).

There has been increasing research into the use of combination treatments for CMHDs, where both psychological and pharmacological interventions are delivered concurrently. For example, Hollon et al (2014) compared the effectiveness of combined cognitive therapy and antidepressant treatment to antidepressant monotherapy for depression, finding improved effectiveness of the combination treatment. Systematic reviews have supported the use of combination therapy in depression, OCD and panic disorder (Cuijpers, Sijbrandij, et al., 2014). However, the benefits of combined treatments are most often limited to severe presentations of CMHDs (Cuijpers et al., 2014; Pampallona, et al., 2004).

Despite the potential benefits of providing combination treatments for CMHDs, there is limited evidence for the cost-effectiveness of doing so routinely for all adult CMHD patients (NICE, 2011c), due in part to the side effects and harms with both pharmacological and psychological interventions discussed above. Different mechanisms of action between psychological and pharmacological interventions have been suggested (DeRubeis, Siegle, & Hollon, 2008) which may indicate the potential for an interaction between treatments. This interaction may be to the benefit of some patients with particular disorders or presentations, but it may be a hindrance for other patients with differing presentations. Therefore a personalised medicine approach that has the ability to identify which monotherapy or combination therapy patients are most likely to benefit from, or indeed if they are likely to experience harm from a particular course of treatment, would have clear value to individuals and potentially at the service-level by helping improve clinical outcome and the efficiency of mental health treatment services.

Treatment dropout

Patients dropping out of treatment early are likely to have worse outcomes as the amount or dose of treatment they receive will likely be inadequate. Although these individuals may be less likely to take up further treatment in the short term, in the longer-term they may require further treatment and this may be more costly in terms of healthcare resources (Wade & Häring, 2010). A comprehensive review of dropout in controlled psychological intervention studies suggested that just under 20% of patients dropped out of treatment, and that dropout was more likely when the intervention was delivered by a less experienced clinician (Swift & Greenberg, 2012). A more recent systematic review of dropout from CBT found that around 26% of participants from included studies dropped-out during treatment (Fernandez, Salem, Swift, & Ramtahal, 2015). Studies of pharmacotherapy for CMHDs have reported similar rates of dropout: 23.5% of participants receiving Duloxetine compared with 23% of those on SSRIs dropped out from treatment (Gueorguieva, Mallinckrodt, & Krystal, 2011); a systematic review of dropout across antidepressant studies indicated that there has been a decreasing trend in dropout by decade with an average dropout rate of 40% indicated from studies in the 1980's, decreasing to 24% in studies conducted in the most recent decade (Schalkwijk, Undurraga, Tondo, & Baldessarini, 2014).

Dropout rates from routine treatment services are higher than in controlled trials, with over 30% of patients estimated to dropout of out-patient mental health treatment (Wells et al., 2013). A recent analysis of dropout in UK psychological treatment services found that 34% of patients dropped out of treatment, 53% of which occurred by the third treatment session (Saxon, Barkham, Foster, & Parry, 2017). Non-attendance at treatment sessions can not only negatively affect an individual's clinical outcome, but also impacts on the efficiency of healthcare services (Oldham, Kellett, Miles, & Sheeran, 2012). Understanding and predicting

when dropout is more likely and using these predictions to try and reduce dropout rates, therefore has the potential to improve outcomes for patients and contribute to a more effective use of healthcare resource. Although meta-analyses have suggested little moderating effect of patient characteristics on dropout rates (Swift & Greenberg, 2012), analyses from individual studies has suggested that some patient characteristics, for example being male and younger, are associated with an increased risk of dropout in the treatment of CMHD (Henzen, Moeglin, Giannakopoulos, & Sentissi, 2016; Reneses, Muñoz, & López-Ibor, 2009). Investigations of the patient characteristics associated with dropout, especially to different treatments, have been limited to-date, however such investigations could have a considerable impact on treatment decisions. Better knowledge about the probability of dropout with particular treatments would have added value as part of a wider personalised medicine approach to the treatment of CMHDs, and by potentially helping reduce the risk of non-compliance and dropout could further help improve outcomes for patients and services alike (Warden et al., 2009).

Limitations of diagnoses

Diagnostic guidelines for mental health disorders, such as ICD-10 (WHO, 1992) or DSM-5 (American Psychiatric Association, 2013), typically use the presence of particular symptoms over a specified period of time in order to make a formal diagnosis. A clinician can use the checklist of symptoms to inform a decision about a specific diagnosis, yet this alone may not yield a reliable and valid diagnosis (Lieblich et al., 2015). The inter-rater agreement between diagnoses of CMHD has improved between versions of the DSM, although kappa agreement values for diagnoses of GAD and depression were 0.67 (Brown et al., 2001). While this would indicate ‘good’ agreement, the potential level of disagreement indicated would suggest that basing treatment selection decisions on diagnosis alone risks ignoring other potentially important characteristics which could inform the selection of appropriate treatment. The identification of CMHDs in primary care services, where there may not be time for clinicians to use full diagnostic interviews, is found to be particularly poor (Carey et al., 2014). For example, Mitchel and colleagues (2009) found that general practitioners (GPs) interviewing patients were only able to correctly identify depression in 50% of patients who met criteria according to diagnostic tests.

One issue is that the point at which low mood reaches clinical significance and meets diagnostic criteria (for example in major depression) is not always clear. Instead, researchers and clinicians have both argued that these conditions may be better considered dimensionally, with symptoms occurring on a continuum (Ayuso-Mateos, et al, 2010). A further issue with the diagnosis of CMHDs is the considerable overlap between defining symptoms and experiences, such as excessive worry or panic attacks which are both common across many CMHDs, as well as substantial comorbidity in the presentation of

these disorders (Kessler et al., 2005). For example, depression and anxiety are frequently comorbid with up to 50% of patients attending primary care for one condition meeting diagnostic criteria for the other (Hirschfeld, 2001).

The comorbidity between CMHDs may also have important implications for the treatment of these conditions and clinicians may need to consider co-occurring symptoms when planning interventions. Individuals with co-occurring symptoms of depression and anxiety are frequently found to have worse outcomes following treatment than individuals with depression alone (Fava et al., 2008). However, more recent findings have suggested that the pre-treatment levels of anxiety may not affect overall outcome, but may instead influence the speed of response to treatment (Forand & DeRubeis, 2013). It is possible that treatment for depression and for anxiety may treat a common underlying problem to both conditions (Kircanski, LeMoult, Ordaz, & Gotlib, 2017), and trans-diagnostic psychological interventions have been developed (Díaz-García et al., 2017; Mansell, Harvey, Watkins, & Shafran, 2009).

Critics of the current diagnostic framework commonly cite the lack of evidence from biological, genetic or neuroimaging data to underpin distinct diagnostic groups, and argue that mental illnesses should instead be viewed as overlapping across multiple symptom dimensions (Adam, 2013; Cuthbert & Insel, 2013). The considerable variation in outcomes between individuals with the same diagnosis to the same interventions indicates that current diagnostic categories may not easily facilitate the prediction of response to treatment, suggesting that this approach to categorising signs and symptoms may not be capturing underlying 'mechanisms of dysfunction' (Insel et al., 2010). Instead more multi-dimensional nosological systems that classify mental illnesses based on behavioural responses, such as the Research Domain Criteria (RDoC) (Cuthbert & Kozak, 2013) offer a different way of considering CMHDs in the absence of biomarkers that can clearly demarcate one disorder from another. By viewing mental disorders on continua, advocates of these systems suggest a more dynamic and flexible method of assessing mental illness can be created (Cuthbert & Insel, 2013). An alternative multi-factorial approach, that considers personal and social factors may also be more reliable than current nosological systems, and could provide a clinically valuable way of understanding human behaviour and distress, which in turn could lead to more tailored interventions (Yee, Javitt, & Miller, 2015).

Despite the limits with current diagnostic systems for CMHD, alternative systems such as the RDoC have not received sufficient backing to replace the existing ICD and DSM manuals in mental health treatment services. At present, evidence-based practice in the treatment of mental health disorders is diagnosis specific, and one could argue that there is a need to categorise disorders to identify the most relevant clinical guidance to aid clinical decision making. A further benefit is that adapting diagnostic systems reduces the risk of over-medicalisation, such as referring to any teenager who avoids social situations as

suffering from social anxiety, and therefore identifying individuals below clinical cut-offs is equally valuable (Callard, Bracken, David, & Sartorius, 2013).

Whereas diagnosis has been fundamental in the development of evidence-based practice in the treatment of CMHDs, the variation in outcomes between individuals with the same diagnosis could suggest that diagnoses may have a more limited role in a personalised medicine approach (Perna & Nemeroff, 2017). Instead, incorporating a range of patient characteristics, which may include symptoms and demographic factors may have greater potential to identify stratified groups of patients that respond differently to treatments.

Current service configuration

In order to develop a personalised medicine approach to the psychological treatment of CMHDs, understanding the services which would benefit from and utilise such an approach is essential. National Health Service (NHS) mental health treatment services in the UK have adopted a stepped care approach to the treatment of CMHDs (NICE, 2011; see Figure 1.1 for details), with 90% of patients seen in primary care (England, Nash, & Hawthorne, 2017). The first point of contact with health services is usually a GP, who plays a vital role in the detection and treatment of CMHDs. If there is a suspicion that a CMHD may be present then a brief symptom screening could be used, and if such a screening suggests the presence of a CMHD then more detailed questionnaires or clinical interviews may be used (Department of Health, 2011). In some scenarios the identification of a CMHD may be straight forward due to clear presenting symptoms, and therefore screening may be sufficient to confirm a likely CMHD. Part of this overall consideration of whether a patient is suffering with a CMHD is the assessment of the level and intensity of symptoms. In general practice it is common place to ask one or two questions from the Patient Health Questionnaire 9-tems (PHQ-9) (Kroenke, Spitzer, & Williams, 2001) as a brief screen of depression and if the patient endorses symptoms in these questions then the full PHQ-9 is often given. The GP or other primary care worker may then refer the patient to a relevant mental health services or offer a recommended pharmacological intervention (such as an antidepressant) if appropriate. In England, this referral for psychological treatment is increasingly to an Improving Access to Psychological Therapies service (IAPT).

The IAPT program was initiated in 2007 by the Department of Health to tackle the increasing financial and social burden of CMHDs by increasing the availability of evidence-based psychological interventions. The British government agreed to invest over £300 million over three years in IAPT services to deliver such treatments (Mental Health Policy Group, 2006; Richards & Suckling, 2009). This included the delivery of programs to train 3600 therapists by 2010/2011 (Department of Health, 2008). These therapists would then deliver National Institute of Health and Care Excellence (NICE, 2011a) recommended psychological treatments. An additional £400 million was pledged by the UK government to further support

the implementation and development of IAPT services between 2011 and 2015 (Department of Health, 2014). IAPT services now receive over 1.25 million referrals nationally per year (NHS Digital, 2016) yet this is estimated to be meeting the needs of just 15% of the population. A recent Mental Health Taskforce report (Farmer & Dyer, 2016) suggests that services should be expanded to meet at least 25% of the need, allowing 1.5 million patients access to such care each year.

IAPT services are built on the stepped care model of treatment (Figure 1.1). In this model, IAPT services provide Step 2 and Step 3 interventions, with Step 1 delivered by GPs and Step 4 typically provided by specialist units, crisis teams or inpatient care. As most CMHD patients referred to IAPT services will enter treatment in Step 2 or Step 3 of the stepped care model, the focus of this thesis will be on these two levels of interventions. The stepped care model provides a tiered care pathway, such that patients initially receiving Step 2 treatments may be stepped up to Step 3 within services should they require more intensive treatment. Conversely, they may be stepped down to less intensive treatment as appropriate.

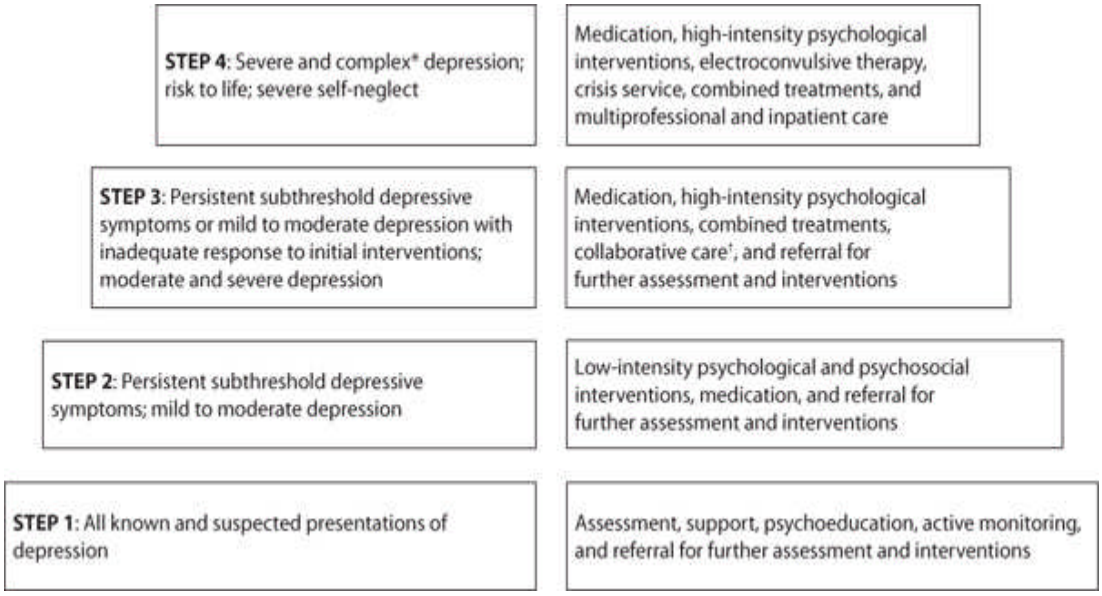

Figure 1.1. The stepped care model (Taken from NICE, 2009)

Step 2 or Low Intensity interventions (referred to as LI interventions from here on) are typically provided by Psychological Wellbeing Practitioners (PWPs) and are primarily offered to clients with mild to moderate difficulties. Some CMHDs have very limited evidence of effectiveness at LI (e.g. PTSD or Social Anxiety Disorder) and so patients with such conditions are typically started at Step 3. LI interventions are mainly cognitive-behavioural in approach, and include guided self-help (facilitated by the PWP), computerised or internet delivered CBT (cCBT), or psychoeducational groups. For depression, other LI interventions based on Behavioural Activation and Structured Exercise are also available in most IAPT

services. Despite the limited direct contact provided as part of LI interventions, meta-analyses have shown them to be effective in the treatment of CMHDs (Bower et al., 2013).

High-Intensity (HI) interventions delivered at Step 3 are designed to treat moderate to severe depression and anxiety disorders within IAPT services, as well as for the full range of severity for disorders with no, or limited evidence for the effectiveness of LI interventions. The types of interventions that may be available include individual or group based CBT, Counselling for Depression, Behavioural Couples Therapy, IPT, Dynamic Interpersonal Psychotherapy (DIT), EMDR, Collaborative Care, and, for relapse-prevention, MBCT (CSIP, 2007).

The healthcare utilisation cost associated with LI interventions is lower than costs of HI interventions, chiefly due to the lower number of sessions delivered as standard with LI interventions (up to 6 sessions on average, compared to 12 for HI interventions; CSIP, 2007), but also due to the lower salaries for LI therapists compared to HI. LI therapists typically have a post-graduate diploma in the delivery of LI interventions compared to HI therapists, who usually have professional qualifications such as doctorates in clinical or counselling psychology, or a professional qualification in the delivery of HI CBT interventions. These costs may be important to treatment selection considerations as all services have limited financial resources available to them. A recent evaluation of the cost of IAPT treatments calculated the cost per session of LI treatment at £99 compared to £177 for HI sessions when all staff and service costs were included (Radhakrishnan et al., 2013). The authors calculated the average cost of complete HI treatment at £1416 per patient compared to £493 for LI. These costs will have important implications at the service-level particularly as services attempt to meet the increased access to care targets, seeing more patients each year than at present. If a patient is likely to benefit from an LI intervention then it would be an overuse of resource to deliver a much costlier treatment when there is no need for it, so personalisation of care in IAPT services will necessarily require the consideration of costs associated with LI compared to HI treatment for any treatment decision aids to be implementable.

Potential for personalised medicine in IAPT

A personalised medicine approach has the potential to optimise the delivery of care in IAPT services, bringing about more rapid improvement in psychological wellbeing during treatment and improved outcomes post-treatment. This could result in more efficient use of healthcare resource, if patient characteristics can be identified that allow for the accurate prediction of treatment outcomes.

All IAPT services in England are mandated to collect the same standardised ‘minimum dataset’ (MDS) of patient characteristics, which are to be collected at referral for all patients.

In addition, services are required to implement routine outcome measure (ROM), the collection of sessional patient-reported outcome measures that provide information about the impact of treatment at each contact with the service (IAPT, 2011). The collection of standardised patient characteristics for all patients provides an opportunity to explore the role of potential factors with treatment outcomes, and understand which characteristics at referral are associated with response to different intensities of IAPT treatment. The use of ROM has the capacity to provide feedback on treatment progress to both the clinician and patient, informing decisions about the progress of treatment. If a lack of symptom improvement has been shown over the first few sessions of treatment then this information could provide an opportunity to consider a change in treatment approach, such as stepping up from LI to HI interventions.

When new referrals are received by IAPT services, a telephone triage assessment is usually performed to gather information about the patient's clinical needs. In many services this triage assessment is typically conducted by a PWP and following this assessment the patient may be discussed with their clinical supervisor before a clinician is allocated and a treatment plan formulated. It should be noted that this process can vary between IAPT services, for example senior clinicians may conduct screenings before deciding on appropriate treatment. There has been an increasing rise in the number of self-referrals to services, and these will normally be followed up with a telephone call before starting treatment. A small number of patients may be referred directly to HI treatments by their GP and will have their initial assessment with a HI therapist rather than a PWP.

There is the potential for a personalised medicine approach to inform treatment choices for new patients and their assessing/treating clinicians at the point of assessment within IAPT services. This could be achieved with the use of a decision support tool, developed to support treatment selection decisions based on patient characteristics recorded during a patient's screening or assessment in the IAPT services. Identifying patient's characteristics that indicate an increased benefit of HI treatment instead of LI, or situations where IAPT delivered treatments would be unlikely to benefit a patient could improve healthcare efficiency and result in more efficient provision of treatment.

Such a decision support tool could reduce the number of patients who are stepped up during treatment, and instead could identify individuals for whom HI would be a more appropriate treatment. A national evaluation of IAPT services suggested that on average over 28% of patients were stepped up (Gyani, Shafran, Layard, & Clark, 2013), and therefore it could be argued that LI treatments were delivered inappropriately for over a quarter of patients. However, as LI treatments are effective for treating CMHDs in a large number of patients (Bower et al., 2013) and are significantly cheaper to deliver (Radhakrishnan et al., 2013), identifying patient characteristics that indicate an increased benefit of HI over LI, and those which suggest equivalent response to the different intensities of intervention could optimise the use of healthcare resource.

The target of IAPT services is to achieve recovery in 50% of patients receiving treatment, levels similar to those achieved in controlled trials (NHS Digital, 2016). Evaluations of service performance nationally have found that many services have achieved or exceeded this target and the nationwide average is now close to reaching this target as 46% of patients recover nationwide (NHS Digital, 2016). However, around 50% of patients referred to IAPT services decline treatment or dropout nationwide, with considerable variation between services (HSCIC, 2015). There will be a number of reasons why patients decline treatment, including the lack of perceived need for treatment as well as preference for alternative treatments other than psychological interventions, and it would therefore be of clinical value to understand the patient characteristics that are associated with treatment dropout. These figures suggest that the nationwide IAPT program has been successful in reducing psychological distress for a large number of patients, although the lack of positive outcomes for over half of patients and the relatively high number of patients not completing treatment may indicate that inappropriate interventions were selected in some circumstances. More consideration of patient characteristics that may be associated with treatment response to psychological interventions and treatment dropout could improve outcomes for patients and services alike.

Due to the vast numbers of patients seen in IAPT services, as well as prevalence of these conditions in the UK, methods of optimising treatment that can increase recovery rates by even a few percentage points will have a dramatic impact on the efficiency of these services (NHS Digital, 2016; Spiers et al., 2016). Telephone triage assessments as well as the increasing number of self-referrals to IAPT provide a large amount of standardised patient characteristics in advance of treatment selection for the majority of patients, which provides an ideal opportunity to incorporate a decision support tool into clinical practice. A personalised medicine approach in IAPT services could therefore be used to identify which intensity of treatment is most appropriate, as well as potential situations where IAPT treatment may not be appropriate due to a lack of predicted benefit.

The next chapter explores the types of clinical decisions that are required in treatment services as well as methods to aid decision making in clinical settings that may inform personalised medicine approaches. This includes a description of some decision support tools that are available in physical as well as mental healthcare, and more recent methods to identify patient characteristics associated with treatment outcomes.

Chapter 2. Clinical decision making & decision support tools

Abstract

This chapter introduces clinical decision making, choosing between different options in patient care, with a focus on decisions that could be supported by personalised treatment approaches for CMHDs. These clinical decisions can include those made during assessment about the nature of the presenting problem (for example possible diagnosis), deciding which treatments will be most appropriate (treatment selection), and decisions made during care (treatment monitoring). Theoretical models of decision making including Prospect Theory (Kahneman & Tversky, 1979), Expected Utility Theory (Schoemaker, 1982) and Bayesian Reasoning (Richardson, 2007) are considered in relation to clinical decisions, as well as factors that can impact on decision making. The final section of this chapter discusses the potential for decision support tools (DSTs), such as clinical prediction algorithms (e.g. the QRISK, Hippesely-Cox, et al., 2007), to support clinical judgement across healthcare. Researchers have used patient characteristics to develop predictive models of treatment response that could inform personalised treatment selection (e.g. DeRubeis et al, 2014), although these have not been translated into available DSTs at present. Systems to aid treatment monitoring decisions have been developed (e.g. Lambert et al, 2001), and these use sessional outcome measurement data to suggest whether the change in symptoms is indicative of a poor treatment prognosis or not. These systems could be further adapted to provide more patient-centred information on treatment progress, for example by identifying groups of patients who respond differently to particular treatments. Currently, there are no routinely available DSTs that can inform both treatment selection and treatment monitoring decisions in CMHDs, but there is potential for such a system to be developed using data collected routinely by IAPT services.

Introduction

Clinical decision making in healthcare can be defined as choosing between alternative options in the care of a patient (Dowding & Thompson, 2003). This can include decisions about which of the available treatments would be most appropriate, whether to stop or continue with the current treatment, as well as decisions about adjusting treatment as it progresses. These judgements are likely to have an effect on a patient's wellbeing and therefore a clinician will usually aim to choose the optimal treatment, both to improve the patient's wellbeing and efficiently use healthcare resources.

How decisions are made can depend both on the complexity of the decisions and the characteristics of the decision maker. For more simple and routine decisions the use of intuition (understanding without the need for conscious reasoning) and heuristics (simple automatic rules used in judgement) may be appropriate, however as decisions become more complex, a more analytical or evidence-based approach may be necessary (Bhugra, 2008). Clinical judgement is developed through training and practice (Kienle & Kiene, 2011), therefore the amount of previous experience with specific clinical situations will contribute to variations in the way that decisions are made. This would suggest that clinicians with less experience will be at higher risk of making incorrect or sub-optimal decisions, which may be evident in audits reporting that less experienced clinicians are associated with higher treatment costs than more experienced colleagues (Mehrotra et al., 2012).

In the UK NHS it is recommended that clinicians adhere to published guidance on evidence-based care, such as the National Institute for Health and Care Excellence (NICE) guidelines, when making clinical decisions across healthcare. The development of treatment guidance is built around the best available evidence to inform and support good clinical decision making and the appropriate choice of treatment (Ioannidis & Lau, 2000). One of the main benefits of evidence-based medicine has been the introduction of more objective and quantifiable estimates of clinical variables into healthcare (Sackett & Rosenberg, 1995), as well as the reduction of uncertainty around clinical decisions, for example due to a lack of previous experience with a particular clinical presentation.

Most healthcare guidelines for UK mental health treatment are diagnosis specific (e.g. depression, social anxiety disorder) and evidence on the effectiveness of treatments is typically gathered from randomised controlled trials (RCTs). RCTs are viewed as the gold standard study design to investigate the efficacy of a given treatment and are therefore critical to the development of the evidence-base about particular treatments. However a common criticism of RCTs is that the participant inclusion criteria of many trials can be restrictive, and so those taking part in such studies may not be fully representative of the population of patients attending routine treatment services (Zimmerman, Mattia, & Posternak, 2002). As a result, clinicians regularly supplement knowledge gained from clinical guidance by using their own specialist knowledge and previous experience to support clinical decision making (Schwartz & Elstein, 2009). This leaves a clear gap for clinicians hoping to not only deliver evidence-based treatments but to offer their patients the ‘best available’ treatment for them as individuals.

Adopting a personalised medicine approach offers the opportunity for clinicians to provide treatments tailored to the individual presentations of their patients and to offer these treatment options based on sound evidence for the likelihood of achieving a desired clinical outcome. This approach will require clinicians to take into account multiple patient characteristics that may impact on clinical outcomes, and might include demographics,

biomarkers (including genetic information), social and family factors, in addition to diagnostic information (Konrad et al., 2015; Pirmohamed, 2014).

Shared decision making

Although decision making has often been considered a clinician based function, increasing value has been placed on the involvement of the patient in clinical decisions (Spatz, Krumholz, & Moulton, 2016). Shared decision making (SDM) can be defined as a collaborative process where the patient and clinician participate in care decisions jointly, discussing the options, potential harms and benefits as well as patients goals (Hoffmann, Montori, & Del Mar, 2014). SDM is a more inclusive approach to treatment, and can increase a patient's understanding of the likely treatment outcomes, which may be an important factor in patient consent to treatment. This collaborative approach can empower patients, and can lead to the selection of treatment that increases the likelihood of outcomes prioritised by the patient (Hargraves & Montori, 2014).

However, all patients are different and despite the potential benefits of SDM, research has found that some patients may not want to engage with the process, and instead may prefer to trust clinician judgement (Eliacin, Salyers, Kukla, & Matthias, 2015). Involvement in SDM can also fluctuate over time and the course of treatment. Patients are usually keen to have all the information initially, but are often happy to allow the decisions to be made by the clinician as treatment progresses (Deber, Kraetschmer, & Irvine, 2014). Research into the association between patient preferences and treatment outcomes has reported significant variation regarding the potential benefits. Williams and colleagues (2016) suggested that patients who expressed preferences were less likely to report that treatment had helped them when their preferences had not been met. Other researchers have found that meeting treatment preferences reduced the risk of treatment dropout, but did not significantly affect treatment outcomes (Dunlop et al., 2017). The variation in findings may be linked to the methods used to report treatment outcome, as Eiring et al (2015) found that often outcomes from studies exploring the impact of preference on response to pharmacological treatment used clinician reported rather than patient elicited symptom measures.

Types of clinical decisions

The types of clinical decisions made during the course of assessment and treatment can broadly be grouped as: i) decisions made about possible diagnoses or clinical problems; ii) decisions regarding the most appropriate treatment to prescribe/allocate, and iii) decisions made in response to treatment progress through monitoring. These are briefly described below:

Assessment and diagnosis

Information gathered as part of the initial assessment is fundamental to understanding the presentation and underlying issues causing the patient discomfort or distress. This will likely include clinical symptoms such as low mood or worry, and in mental health settings will regularly involve assessment of the social and occupational impact of these symptoms. This information may be used to diagnose the presenting problem, or to formulate the problem and consider what is maintaining/preventing it from improving without any treatment. If there is insufficient evidence available in the initial assessment then further assessment may be required, such as the use of specialist diagnostic equipment in physical healthcare (e.g. CT scans) or further psychological assessment (e.g. cognitive assessment) in mental health settings.

Treatment selection

Following the assessment of the presenting problem(s) and maintaining factors, the next clinical decision will concern the selection of appropriate treatment. The personalised medicine approach aims to incorporate patient characteristics such as demographics and clinical symptoms into this decision, alongside relevant clinical guidance.

For some clinical scenarios the decision will be straight forward due to either the nature of the condition, or lack of alternative treatment options. However, some clinical decisions, especially those with more complex presentations or scenarios where a number of alternative options are available, will be more challenging for the decision maker. Identifying patient characteristics that are associated with treatment outcomes could therefore help to determine the most appropriate choice. For example, in the treatment of most breast cancers there are a number of options to be considered, but research has shown that patients with higher expression of human epidermal growth factor (HER-2) respond particularly well to Herceptin (Verma & Mukesh, 2012), which can therefore inform the treatment selection decision. In IAPT services, a personalised treatment approach would aim to help make decisions about which type, or intensity of psychological intervention is the most appropriate given the patient's characteristics at presentation to the service.

Monitoring treatment

Once the selected treatment has been initiated, there will likely be further decisions made in response to information collected during routine monitoring of the patient, and their progress during treatment. This monitoring will provide information about the impact of the treatment on the patient's wellbeing, for example whether treatment is having the desired effect in reducing symptoms, or if there are significant side effects which may indicate that a change in treatment should be considered. The use of routine outcome measurement (ROM), the collection of patient information at multiple points during the course of treatment, is usually vital for treatment monitoring decisions. Should the patient's ROM data suggest that there

has been limited response to treatment, then this can inform decisions about whether the treatment should be continued or if an alternative approach should be considered.

Theoretical models of clinical decision making

Given the fundamental role that decision making plays in healthcare there have been various attempts to theorise about how these decisions are made, although relatively little research has been conducted in the context of mental health treatment (Wills & Holmes-Rovner, 2006). Research has suggested that experienced clinicians will often use initial referral information to form an early hypothesis about diagnoses and or appropriate treatment, before or during their first contact with a patient (Elstein & Schwartz, 2002).

Clinicians will most often consider clinical guidance and evidence-based knowledge when making their decisions, but those with less experience may not be as able to supplement this with clinical judgement. Instead, clinical judgement is thought to improve over time with experience and training, until in some situations judgement may become automatic. These automatic decisions have been referred to as 'affective heuristics' (Slovic, Finucane, Peters, & MacGregor, 2002), and will develop with experience. Although clinician experience has been associated with increased healthcare costs (Mehrotra et al., 2012), research in psychotherapy outcomes suggests that patient outcomes for more experienced clinicians can be worse than outcomes for less experienced clinicians (Goldberg et al., 2016). It may be that more experienced clinicians are more likely to be allocated more complex and difficult to treat patients, but researchers have also suggested a phenomenon known as 'therapist drift', where clinicians do not keep up with the evidence-based compared to more recently trained clinicians (Waller & Turner, 2016), which may reduce the effectiveness of interventions they deliver.

Some prominent theoretical models of clinical decision making are described below:

Bayesian reasoning

As the evidence-based approach to clinical guidance in CMHDs has broadly focused on identifying the best treatments for specific conditions, usually one of the first decisions for a clinician will be a formulation of the patient's problem in the context of their current circumstances. This may include the identification of a potential diagnoses, which could then be used to inform further clinical decisions, for example drawing on knowledge of relevant clinical guidance to support a treatment plan. The decision as to whether or not a specific diagnosis is present can be represented as a probabilistic choice between an event (diagnosis) existing or not, as a model of Bayesian reasoning (Richardson, 2007).

Bayesian reasoning is derived from Bayes theorem (Bayes, 1763), and posits that clinicians use information from both the patient assessment and existing knowledge about the likelihood of the event occurring (pre-test probability) to determine whether a diagnosis is present or not. The pre-test probability is based on the likelihood of that disorder naturally occurring in that clinical environment (i.e. prevalence of the disorder in the service), and therefore relies on either the clinician's previous experience of that diagnosis in the service. Patient information collected at assessment is then used to increase or decrease the clinician's estimated likelihood of the diagnosis (Fahey & Van Der Lei, 2009). As new information is acquired during assessment, the clinician's estimate of the likelihood of diagnosis is updated (Schwartz & Elstein, 2009). The post-test probability of the disorder being the 'correct' decision is therefore a function of the pre-test probability and the strength of the available evidence.

Bayesian diagnostic reasoning has generally been applied to physical healthcare, where there is more availability of objective measurement values (e.g. blood pressure), whereas the measurement of symptoms in mental health relies more on the subjective measurement of psychological distress from symptom scales (Bhugra, 2008). However, Bayesian reasoning could be evident in mental health treatment services where a clinician may combine assessment information with the local prevalence of CMHDs (pre-test probability) then estimate the probability of a specific diagnosis (e.g. panic disorder) being present. If the clinical presentation suggests a number of symptoms common to panic disorder, and there is a sufficient pre-test probability of individuals with panic disorder being referred to the service then the clinician may decide that panic disorder is the likely diagnosis and an appropriate treatment plan can be formulated.

Prototypes

An alternative theory about how clinicians make decisions about both the presentation of the patient and appropriate treatment choices is the use of in-built "prototypes", representations of particular illness/disorders constructed by the clinician (Garb, 2005). These prototypes are internally derived representations of how a typical 'type' of patient, for example with social anxiety disorder, would present to services and the clinician would compare a new patient against their existing prototypes to identify an appropriate match (the prototype which appears most similar).

The use of prototypes to identify stratified groups of patients with similar characteristics complements the aims of personalised medicine approaches to treatment, as certain prototypes may be associated with differential outcomes to treatments in IAPT services. However, as these representations are generated by individual clinicians, they will be highly subjective and heavily biased by previous experience of different patient groups. The use of prototypes may explain some of the variance in the inter-rater reliability often found in

diagnostic studies, as different clinicians are likely to have slightly different prototypes for the same diagnosis (Pies, 2007). There could be some overlap between the patient characteristics common to prototypes developed by two independent clinicians, but a system of grouping patients that is to have utility across services will need a more objective method of stratifying patients that would be common to all clinicians.

Expected Utility Theory

Treatment selection decisions will consider which of the available treatments are most likely to result in the best outcome for the patient, and therefore require some estimate of the potential value of the treatment outcomes to the patient. The most commonly cited normative theory of clinical decision making is Expected Utility Theory (EUT) (Schoemaker, 1982), which is used across healthcare to model decisions (Chapman & Sonnenberg, 2000). In EUT, each possible outcome from each available treatment is given both a likelihood of occurrence (probability) and a value of that outcome to the patient (utility). Utilities are usually given as a range from 1 to 0, with 1 being a perfect state (perfect health) and 0 the worst state (e.g. death), and therefore the best decision is one that results in the highest utility.

These utilities are normally taken from patient and clinician recorded measures of quality of life, and in the context of CMHD treatment could be linked to the level of decreased distress and functional impairment caused by the clinical symptoms. The most appropriate treatment is therefore the one with the highest probability of the best outcome for that patient. The probability of outcomes is derived from either the clinician's prior knowledge (experience) or from published research if it exists for the specific situation. Therefore, the clinician's decision is based on their subjectively judged probability of whether the event will happen or not. EUT also attempts to explain how clinicians consider trade-offs in clinical decisions by weighing up the benefits and costs of certain decisions on eventual outcome (Wills & Holmers-Rovner, 2006).

Prospect Theory

An alternative theory of decision making in the presence of clinical uncertainty is Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) and differs from EUT as it is concerned with describing observed decisions assuming the decision maker has inbuilt descriptive rules, rather than assuming the decision maker is rational and able to estimate perfect accuracy (normative). Prospect theory suggests two phases in decision making; an initial editing phase before a subsequent evaluation stage. In the editing phase, all potential outcomes from the decision are ranked based on certain heuristics of the decision maker, specifically around an outcome they consider a reference point and compared to which all

other outcomes are either losses or gains. Using this method, the probability of events occurring can be compared. In the evaluation phase, values are allocated to each outcome (comparable to utilities in EUT) and the decision is made in light of the outcome with the highest value. The major difference between prospect theory and EUT is that the decision maker in EUT does not compute a reference point and therefore focuses only on improving gains, rather than reducing losses in specific situations. In IAPT services this may be a comparison of the probability of treatment success (utility) for either low intensity or high intensity treatment, or whether there is an increased value of allocating to more resource intensive HI treatment over LI interventions.

The similarities between Bayesian reasoning, Prospect Theory and EUT models of decision making are that they all suggest the decision making will use the expected probability of an event (diagnosis or outcome) to inform the decision. This requires either prior experience of the situation or some reference with which to predict the likelihood of the event. As this information is not always available, decision making in these circumstances is vulnerable to bias resulting in either incorrect or over risk adverse decisions that may not be favourable to the patient.

Challenges in making effective decisions

The models of decision making discussed above (for example EUT) propose that the clinician making the decision requires a reliable estimate of both the presence of a specific disorder and the expected value or benefits of appropriate treatment. However, research has frequently shown that clinician's estimates are prone to biases and errors, in both test and clinical environments. Research from the US has indicated that of closed malpractice cases across all healthcare settings, 64% involved diagnostic error suggesting that incorrect treatments may have been selected due to incorrect estimates from the clinician (Gandhi et al., 2006). It is possible that there is an increased risk of this 'diagnostic error' in mental health services where there is often a lack of objective 'gold standard' measurement tools and systems, compared to physical health settings where many conditions have objective tests (e.g. blood pressure or liver function tests).

Although heuristics can be useful strategies for clinical decision making (Slovic et al., 2002), they also increase the likelihood of systematic biases in judgement (Fahey & Van Der Lei, 2009). As discussed above in relation to Bayesian reasoning, the estimates of pre-test probability (the likelihood of a disorder naturally occurring) and the expected utilities of treatment will be both informed and biased by the level of experience of the clinician as well as the environment in which they work. Clinicians from secondary or tertiary care services will typically see more complex or severe presentations, which are rare in less specialist services and therefore may overestimate the likelihood of more severe illnesses in other settings (e.g. primary care). This may result in a very different set of heuristics for these

clinicians, for example how they judge the risk of suicide in complex patients. One study by McGinn and colleagues (2002) asked a group of junior doctors to estimate the likelihood of a pulmonary embolism after reading a clinical vignette of a patient's symptoms, and found that estimates ranged from 5% to 80%. These estimates may be heavily biased by either experience, perception of risk or clinical specialism.

One challenge with clinical decision making is that individuals can struggle to estimate the probabilities of events from basic available information. Yamagishi (1997) used a simple experiment where undergraduate participants were asked to decide which particular causes of death (e.g. cancer) were the most likely, based on a provided incidence of mortality. Information on the number of deaths were provided in relation to increasing denominators (e.g. 24.14 out of 100, 1,286 out of 10,000) and results showed that participants would frequently rate 1,286 out of 10,000 as higher risk of mortality than 24.14 out of 100 despite 24.14% being higher than 12.86%. It would appear that the absolute number of events, rather than the percentage of events influenced the perceived risk in the sample.

Although this experiment was conducted on non-clinician participants, the findings may be very relevant to shared decision making, as patients may be equally vulnerable to perceived risk and therefore considering the way in which information is presented could reduce potential biases (Jefferies-Sewell et al., 2015). Further research suggests that individuals often underestimate probabilities of difficult to recall or vague events whereas there is an overestimate for more vivid events (Elstein, 1999). Incidents that are widely reported in the media are found to be considered more probable than their prevalence suggests (Elstein & Schwartz, 2002).

A further issue with the estimation of probabilities is the phenomena of 'compression error' where small probabilities are over-weighted and large probabilities under-weighted (Fischhoff, Bostrom, & Quadrel, 1993). The notion is that humans are not comfortable being absolutely certain of an event occurring and therefore the difference between 99% and 100% will be assumed as further in distance than the difference between 40% and 41% for example. This distortion is further exacerbated when the exact probability is not known, and individuals will tend to be more conservative in estimation (Schwartz & Elstein, 2009). This has important implications for treatment selection where a slightly riskier intervention may not be selected as the clinician may not be "100% sure" of its success, despite all information indicating that it is the most appropriate treatment option. This may be true of a psychological intervention that has a higher than average incidence of clinical deterioration but is strongly associated with positive outcomes in a specific sub-group of patients.

Research has suggested that psychologists can struggle to accurately predict patient outcomes, and are likely to be over-optimistic of their own patient's progress (Walfish, McAlister, O'Donnell, & Lambert, 2012). For example, Hannon et al, (2005) asked 48 therapists to use their clinical judgement and predict which patients in their care would

deteriorate during treatment. The clinicians predicted only 3 out of a total of 550 patients would deteriorate, whereas 40 (7.3%) patients reported a worsening of symptoms post-treatment. Although deterioration is less likely than positive outcomes from treatment for most patients, these results suggest that clinicians may struggle to predict treatment response in CMHD psychological treatment from initial assessment, which may make their estimates of expected outcomes and utilities prone to error.

Due to the potential challenges with clinical decision making, there has been increasing research to develop methods of supporting decisions in healthcare. These aids to clinical judgement aim to inform decisions on appropriate care and have the potential to incorporate patient characteristics in decision making, creating a more patient-centred approach to healthcare. The next section of this chapter explores the role of decision support tools in healthcare, and the potential for these aids to inform clinical decisions in IAPT services.

Decision Support tools (DSTs)

One potential method for aiding decision making in clinical settings is the use of 'decision support tools' (DSTs) in routine care settings. Clinical DSTs can refer to any system that provides patient-specific advice, for example whether clinical information suggests more detailed screening is required (e.g. for cancer) or which treatment is recommend for a given condition (Geissbuhler & Miller, 2000). DSTs are designed to support clinical decision making about individual patients in real-time (Berner & Lande, 2007), and many DSTs are now computerised for ease of use and their ability to provide information efficiently. The aim of many DSTs has been to improve patient safety, for example by reducing medical errors, with many DSTs originally designed to simulate human decision making processes (Jia et al., 2016).

One of the biggest areas of improvement gained by the use of DSTs is in medication administration, where a large number of medical errors (such as wrong dosage or drug) have historically occurred (Kohn, Corrigan, & Donaldson, 2000). DSTs in these environments may be set up to replicate current evidence-based treatment guidelines, and therefore prompt clinicians with recommended treatment options including dosage. Additionally, such tools may have the facility to alert clinicians to dangerous drug interactions, thereby minimising errors and potential side effects. These tools can add structure to medical knowledge and provide clinicians with standardised patient information (Saverno et al., 2011). The utility of DSTs is greatly enhanced by incorporating them within local clinical information and electronic patient management systems (EPMS), as this greatly reduces the amount of information having to be shared across isolated systems (Geissbuhler & Miller, 2000).

DSTs are also used to aid decisions on screening options (e.g. cancer screening) and the management of conditions such as diabetes and cardiovascular disease (Sheehan & Sherman, 2012). DSTs can vary in format, for example some tools are presented as a sequential treatment/procedure flow chart (e.g. NCCC, 2009) whereas others are derived from computerised prediction algorithms that use patient information and provide real time clinical recommendations, such as the Fracture Risk Assessment Tool (FRAX) tool for assessing risk of osteoporosis and fracture (Cadarette et al., 2000; Hillier et al., 2011). Some examples of these are presented below:

Flow charts / sequenced decision support tools.

Sequenced DSTs are typically presented as detailed flow diagrams whereby the clinician or patient is asked a question (e.g. “is systolic blood pressure over 150mmHg?”) and the answer determines whether an intervention is recommended (e.g. ‘prescribe statins’), or whether additional questions are needed before a recommendation can be provided. This DST could be presented as a paper diagram which could be easily followed by the user, or it could be hosted via a computer-based system that required the user to answer a series of sequential questions. Each question could be considered to act like a ‘decision node’, with simple sequenced DSTs having just two or three nodes, and more complex having over 10 nodes with questions to consider.

One group of sequenced DSTs recommended for use in UK emergency care are the NICE stroke and transient ischaemic attack (TIA) algorithms (NICE, 2008). These algorithms provide a series of sequenced questions and recommendations that the clinician should consider when managing a suspected stroke or TIA. For example, if a TIA is suspected and neurological symptoms have resolved then the first decision node asks “is history compatible with TIA?”. If yes, then 300mg aspirin is recommend alongside additional statins or lifestyle management, with the next decision node asking the clinician to assess risk of stroke using a validated measure. If the risk of stroke is high then one path is followed, with a series of recommendations for treatment and assessment, whereas if the risk is low then an alternative care path is followed on the flow diagram.

Clinical prediction rule algorithms.

Prediction rule algorithms differ from sequenced DSTs as they are more likely to provide a recommendation or prognosis drawing on a submission of information at one time point, for example information collected at an initial assessment. These DSTs normally have an underlying algorithm that uses a number of patient characteristics (e.g. demographic, family history, clinical symptoms) to generate a treatment recommendation or the risk/likelihood of illness. These algorithms differ from sequenced DSTs as they usually combine a number of

patient characteristics in one calculation to estimate risk or provide a treatment recommendation, rather than needing a number of stages to be completed by the user.

The QRISK (Hippisley-Cox et al., 2007) is a widely used risk prediction algorithm in primary care to provide a 10-year estimate of the risk of cardiovascular events for patients. DSTs such as these can be used to provide a personalised approach to prevention for cardiovascular events, and clinicians can tailor recommendations, such as lifestyle changes, following the information provided by the QRISK. The use of these DSTs has evolved to use genetic information to personalise healthcare, for example pharmacogenetic profiling to inform medication management in older people, and a recently developed DST has been shown to reduce hospitalisation and emergency department visits in this population (Brixner et al., 2016)

Many of these sophisticated prediction algorithms have been developed using large routine patient datasets, rather than those from controlled trials, as these cohorts increase the available statistical power for predictive analyses (Abbasi, 2017). For example, the QRISK algorithm was developed using a dataset of over 1.25 million patients registered at GP practices across the UK (Hippsley-Cox et al., 2007). The growing use of EPMS and outcome measurement in healthcare can provide data on a huge number of patients over a large range of patient characteristics, which has the potential for development of a range of clinical DSTs.

Impact of decision support tools

Reviews evaluating the use of DSTs across medicine have suggested they can deliver improvements in clinician performance as measured by increased incidence of tests or examinations performed (i.e. increased outcome measurement and monitoring), although improvements in patient outcomes have been more varied (Bright et al., 2012; Garg et al., 2005; Hunt, Haynes, Hanna, & Smith, 1998). A systematic review conducted by Kawamoto and colleagues (2005) evaluated which components of DSTs were associated with improved clinical practice. From the 77 included studies of DSTs across healthcare, 68% suggested a significant improvement in clinical practice. Four features of systems were linked to clinical benefit: i) the automatic provision of decision support as part of clinician workflow; ii) provision of support at the time and location of decision making; iii) provision of a recommendation rather than used for assessment only, and iv) computer-based generation of decision support. Of the 32 DSTs with all four of these aspects, 94% were shown to significantly improve clinical practice.

One potential barrier to the uptake of DSTs can arise when clinicians do not agree with the guidance from the new system, which can result in clinicians ignoring recommendations or refusing to use the DST (Keeffe et al., 2005). This would suggest that providing training to

clinicians, as well as seeking their views during the development of the DST that they will be expected to use may be of value and could increase clinician uptake.

DSTs could also provide a more prominent role for patients within their care, by informing them about potential risks and benefits of available treatment options (Moynihan, 2013). This is especially true of DSTs that have user-friendly formats that can engage the patient into the decision-making process. Critics however, suggest that patients tend to make more cautious decisions, which in turn may increase costs to healthcare (McCarthy, 2013; Walsh et al., 2014).

Decision support tools in mental health settings

Whereas physical health has seen a range of developments with DSTs, the development and evaluation of DSTs has been slower in mental healthcare (Sheehan & Sherman, 2012). However, a small number of potential methods have been developed for treatment selection decisions for both pharmacological and psychological interventions, although few have been made available to services or evaluated in clinical practice. These decision support methods can broadly be considered as either treatment selection tools which aim to aid clinicians in deciding which of a number of available interventions will result in 'the best' outcome, or treatment monitoring tools which are designed to provide an indication of whether to continue or change the current course of treatment, e.g. if there is limited or no clinical response. The main work in the mental health field has been for the treatment of depression (e.g. Chekroud et al., 2016; DeRubeis et al., 2014)

Treatment selection decision support tools.

The development of predictive models and methods of using patient characteristics to inform treatment selection decisions has grown in the past years, with a number of recent publications.

Researchers have used datasets collected during RCTs which have compared the effectiveness of psychological and pharmacological interventions, to develop methods of identifying patients who are more likely to benefit from one type of intervention over the other. Wallace, Frank and Kraemer (2013) have created a single combined moderator value (referred to as M*) (using methods described by Kraemer, 2013) from weightings of key patient characteristics and used this composite moderator value to predict which treatment for depression will be most effective. A significant difference was found between treatment outcomes for patients scoring at the higher and the lower values of the combined moderator, suggesting treatment outcome could be predicted in individuals at the extremes of this M* value. DeRubeis and colleagues (2014) developed the 'Personalized Advantage Index' (PAI)

to predict the final symptom score for a given patient under both psychological and pharmacological treatments for depression. This algorithm used a range of characteristics including IQ, employment, life stressors, previous treatment and comorbid personality disorder, and was able to show a significant advantage of one treatment type over the other for 60% of patients in the development sample.

A couple of recent studies (Chekroud et al., 2016; Iniesta et al., 2016) have developed statistical models to predict outcomes from different antidepressant medications. Chekroud et al (2016) used data from nearly 2000 participants from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial (Trivedi et al., 2006) and used a machine learning driven clustering method (hierarchical clustering) to identify a method of predicting response to citalopram, before validating model performance in data from the COMED trial (Rush et al., 2011). From 164 available patient characteristics, the model was reduced to the 25 best characteristics and the final model included individual symptoms across two depression measures as well as demographic variables (e.g. ethnicity, years of education), and the model showed good performance across three of the four drug treatments. Iniesta et al (2016) used data from the Genome-based Therapeutic Drugs for Depression (GENDEP) (Uher, Huezo-Diaz, et al., 2009), to develop a machine learning driven regression model to predict response to escitalopram or nortriptyline from n=793 participants. The dataset included a wide range of patient characteristics, including demographic characteristics such as BMI, marital status, number of children, clinical symptom data across four depression measures, medication history and details about stressful life events. The best model structure varied between treatment type and models could explain 5-10% of the variance in a clinical significant reduction in symptoms, and around 15% of the variance in remission, which was considerably above the authors' benchmark 6.3% of variance explained in order to be clinically important (Uher, Tansey, Malki, & Perlis, 2012).

Although these studies have identified patient characteristics associated with treatment outcomes and used statistical modelling to develop methods of predicting differential response, each of the samples were derived from relatively small clinical trial populations. These methods have not yet been translated into formal DSTs that have been evaluated for effectiveness in routine clinical practice. Clinical trial samples can be highly selected for the specific purposes of the trial and therefore are at risk of not reflecting the typical range of patients attending mental health services (Zimmerman, Walsh, Chelminski, & Dalrymple, 2017), and therefore further evaluation of these models in routine care samples would be recommended.

The Wallace et al (2013) and DeRubeis et al (2014) methods used the available patient characteristics in datasets to group patients into those would were expected to respond better to medication and those for whom psychological therapies were likely to be most beneficial. However, little research has investigated treatment selection between different types of psychological therapies (Fonagy, 2010). As increasing evidence suggests patients

tend to prefer psychological treatment over pharmacotherapy for psychiatric disorders (McHugh, et al., 2013), it would appear that methods to improve treatment selection in psychological therapies would be of benefit, especially for use in IAPT services. Currently two DSTs have been developed and trialled in mental health settings.

One study, from the child mental health literature, used an algorithm to determine whether trauma-focused cognitive-behavioural therapy (TF-CBT) is likely to lead to a positive outcome, based on the current research evidence (Lang, Ford, & Fitzgerald, 2010). Although informative on deciding whether to use the treatment or not, it is limited by an inability to suggest an alternative intervention if TF-CBT is not recommended. An algorithm that could recommend one intervention over another could be of more value to clinical staff and patients.

A second decision support tool has attempted to allocate patients to one of two psychological interventions, either CBT or psychodynamic therapy (Watzke et al., 2010). The decision on treatment selection was made by integrating assessments and clinical interviews, with 86% of decisions made at least partly on the patient's goals of treatment (goals around modifying behaviour and coping with situations recommended to CBT, and goals around understanding unconscious emotions recommended to psychodynamic treatments), and 74% incorporating diagnosis into the decision. However, the evaluation of this DST in clinical practice showed no significant differences in patient outcomes when compared to random treatment allocation.

One potential reason for the lack of a positive effect of using the Watzke et al (2010) DST may be that the decision was made predominantly on just two factors: the patient's goal for treatment and the diagnosis at assessment, with limited use of other information. The multi-dimensional nature of mental illnesses suggests that there are many patient specific characteristics that could be important to account for in treatment selection decisions, and methods such as those employed by DeRubeis et al (2014) or Iniesta et al (2016) suggest a number of patient characteristics can inform differential response.

A very recent study has used a sample from IAPT services in the North of England to explore regression models for predicting a reduction in symptoms, rather than IAPT recovery (Delgadillo, Moreea, & Lutz, 2016). The IAPT MDS for this analysis was supplemented with additional patient characteristics not routinely available in the IAPT dataset, such as patient expectancy of treatment outcome, family history of illness and disability. The final regression model using available characteristics was able to explain approximately 13-15% of variance in patient outcomes, and was developed using a dataset of 1347 patients. A risk weighting called the "Leeds Risk Index" (LRI) was developed from the regression coefficients and has been used to create a 'traffic light' system indicating the risk of poor treatment outcomes, classifying patients as 'low risk' (green), 'moderate risk' (amber) and 'high risk' (red) (J. Delgadillo, personal communication, 27th April, 2016). These groups were associated with

differing likelihoods of achieving significant change in depression and anxiety symptoms, with the low risk group more likely to achieve positive outcomes and the high-risk group the least likely. Outcomes were better for the high-risk group following HI treatments compared to LI for both depression and anxiety symptom reduction, and the LRI could be used to develop a DST to identify high risk patients and recommend high intensity treatment for these patients to increase the likelihood of positive outcomes.

To date the majority of these treatment selection methods have not been trialled in mental health settings despite the potential benefit to patient care. One DST which has been trialled in treatment services (by Watzke et al, 2010) found no benefit of using the treatment selection tool over existing treatment selection methods (usual treatment allocation), but this DST was limited regarding patient characteristics used to inform the treatment selection decision (goal of treatment or diagnosis were the predominant characteristics used). One possible reason for the lack of uptake of these current methods of personalising mental health treatment in routine services is that the majority were developed in controlled trial populations and therefore may not reflect populations receiving routine treatment for CMHDs. It might be argued that using data from local services to develop a DST may increase the likelihood of uptake by treatment services.

Most of the methods considered above have used regression-based analyses to identify patient characteristics predicting treatment outcome. As a result, these models provide evidence for the association of included patient characteristics with the outcome of interest but do not inform on the role of these characteristics with other potentially important outcomes. For example, methods developed by DeRubeis et al (2014) and Iniesta et al (2016) were focused on identifying patient characteristics associated with a reduction in psychological symptoms, but the association between these characteristics and other clinically important outcomes such as treatment dropout cannot be explored in the same model. For treatment services such as IAPT, a number of treatment outcomes can be important, especially if the evaluation of treatment outcome in these services is determined from more than one measure of psychological wellbeing (NHS Digital, 2016). Therefore, alternative approaches to modelling data that can enable multiple outcomes to be considered may have additional benefit for IAPT services.

Treatment monitoring decision support tools

DSTs that focus on treatment monitoring decisions usually require the use of ROM to feedback information about the patient's progress in treatment, which can then be used to inform decision making. Typically the algorithms and DSTs appear to differ between those for pharmacological treatments and psychological interventions. DSTs developed for drug treatments are linked to the sequencing of treatments based on response (or not) to previous drugs (Rush, Crismon, et al., 2003; Stein et al., 2012), and can be viewed as flow charts of decision rules. Monitoring systems for psychological interventions instead focus on

the change in symptoms, and indicate to clinicians when progress during treatment is not following the expected course for the patient (Lambert, 2013).

One major study on the sequencing of pharmacological treatments in CMHDs is the Psychopharmacology Algorithm Project Harvard South Shore Program, which has produced algorithms for anxiety and depressive disorders (Stein, et al. 2012). The algorithm's approach is to recommend pharmacological interventions in a set order (as a flow chart), and suggests a change in treatment following non-response to the previous stage. The algorithm behind the recommendation sequence is drawn from clinical trial evidence of the individual drugs, however past research into the use of sequential methods for treatment delivery has found little evidence for their benefit over standard clinical decision making (Hatcher, 2008). Although some patient characteristics are taken into account within certain algorithms (such as sleep issues in PTSD; (Bajor, Ticlea, & Osser, 2011)), the algorithms function mainly as an ordered list of interventions that the prescriber is recommended to follow.

A second algorithm project, also focused on pharmacological interventions, is the Texas Medication Algorithm Project (TMAP), which has produced decision aids for depression, as well as schizophrenia and bipolar disorder (Crismon et al., 1999; Rush, Crismon, et al., 2003). In a similar fashion to the Harvard South Shore Program described above (Stein, et al. 2012), these algorithms also use sequenced approach to decision making. One key difference between these algorithms, is that the TMAP recommends routine outcome monitoring by the prescribing clinician, with the Quick Inventory of Depressive Symptomatology scale (QIDS-16) (Rush, Trivedi, et al., 2003) completed at regular clinician-patient meetings, whereas the Harvard algorithm does not. The score on this outcome measure then informs the clinician on the level of response or non-response to the intervention, and therefore the need to move onto the next treatment sequence, with the outcome measurement system being vital to the decision to maintain or change treatment.

Algorithms such as those described above are based on a sequence of different interventions, with failure or limited response for one treatment leading to recommendations for the next designated treatment, which is common to recommended treatment algorithms for physical health conditions such as breast cancer (NCCC, 2009). However, these algorithms are designed so that every patient will go through each stage until they respond to an intervention. To be allocated to the eventual 'successful' treatment, patients must first complete an adequate dose of treatment at all previous, non-successful stages, and these ineffective interventions will be costly to both services and patients regarding time and finance, as well as potential side effects.

DSTs for treatment continuation during psychological interventions have not considered the provision of different types of therapeutic model but instead monitoring whether the patient is

responding or not, flagging non-response and providing information to the clinician that more input, or change of approach is needed (Lambert, 2013).

The OQ-45 psychotherapy outcome management system developed by Lambert and colleagues (Lambert, 2001; Lambert et al., 2001) has a large amount of research evidence supporting its clinical utility in improving patient outcomes (Hannan et al., 2005; Shimokawa, Lambert, & Smart, 2010; Slade, et al., 2008). The Outcome Questionnaire-45 (OQ-45) (Lambert, 1983) is a measurement tool designed to capture the range of clinical symptoms for CMHDs, as well as symptoms of additional psychiatric disorders, and the OQ-45 system relies on sessional collection of this outcome measure.

The measure and system were initially developed in university counselling centres, and for each new patient an expected response curve is generated, which is the predicted trajectory of change in OQ-45 score over time if a positive outcome is to be achieved. Research into the trajectory of change in psychotherapy typically finds that change is largest (a greater effect) in the first few sessions of therapy and this levels out as the number of sessions increases (Kopta et al., 1994). The OQ-45 score is completed by the patient at each session and entered into the system, after which advice is returned to the clinician about the patient's current trajectory and whether they are "On track" (expected good outcome) or "Not on track" (predicted poor prognosis). This prompt signals that the clinician may need to consider alternative treatment options in order to get the patient back "on track". Alternative feedback systems have been developed, which provide similar recommendations to those of the OQ-45 systems and a meta-analysis indicated a small effect in favour of using these DST to improve patient outcomes (Knaup, Koesters, Schoefer, Becker, & Puschner, 2009).

Although there are patient benefits associated with the use of the OQ-45 system, with research indicating that flagging patients as 'not on track' can improve outcomes (e.g. Shimokawa et al., 2010), there are some potential limitations to its use in improving decisions in routine treatment services, such as IAPT. Firstly, the OQ-45 measure is relatively long for a ROM, and although the author suggests these 45 items can be completed in five minutes (Lambert, 1983), this may be optimistic as there are suggestions that 10 items can take up to 15 minutes to complete (Marks, 1998), and this lack of brevity may make the assessment tool too time consuming for routine use in highly time pressured services (Clifford, 1998). IAPT services currently use the Patient Health Questionnaire 9 item version (PHQ-9) (Kroenke et al., 2001) and the Generalised Anxiety Disorder scale 7 item version (GAD-7) (Spitzer, Kroenke, Williams, & Löwe, 2006) as sessional outcomes measures. The PHQ-9 and GAD-7 are both estimated to take 2 minutes each to complete (Duffy et al., 2008; Sousa et al., 2015), and therefore it is likely that asking patients to complete an additional 29 items will increase the time burden significantly, and may result in reduced data completion.

An extension of this method of identifying expected treatment response and providing feedback to patients was developed by Lutz and colleagues (Lutz et al., 2006; Lutz, Stulz, Martinovich, Leon, & Saunders, 2009). These researchers used a 'nearest neighbours' modelling approach to identify patients with similar presentations of symptoms to generate expected response curves to psychological treatment. These curves can be used to provide information about whether a current patient receiving treatment is 'on track' or not based on their change in psychological symptoms, like that provided by the OQ-45 system (Lambert, 2001). This nearest neighbour approach is based on an analysis of patients who had previously received treatment and their response to that treatment. The system can identify for new patients entering treatment, which historic patient they most resemble based on presentation to the service (i.e. identify their 'nearest neighbour'). The expected response for the current patient is therefore expected to be similar to that of the nearest neighbour, and the system can flag up situations when the current patient's response to suggests that there are not on track compared to their nearest neighbour's response. The nearest neighbour matching is based on subscale scores on measures of psychological symptoms (e.g. depression, anxiety) collected at assessment. This approach therefore extends to the Lambert (2001) OQ-45 model by considering subscale of symptoms rather than just total psychological symptoms scores. However, it is possible that this approach could be further personalised by including addition patient-specific characteristics, including demographics, which may further inform the expected response to psychological treatments.

Very recently Lutz and colleagues (Lutz, Zimmermann, Müller, Deisenhofer, & Rubel, 2017) have published a protocol paper for a randomised controlled trial of a personalised prediction and feedback tool for psychological treatment. The trial will include a group of patients randomised to receive a treatment selection recommendation based on the identification of their nearest neighbours in an historic database and the treatment that resulted in the best outcomes for these similar patients. Feedback on treatment progress ('on track' or not) will also be provided to clinicians in this intervention group. This group will be compared against a control group were nearest neighbours are not considered and therefore a treatment recommendation is not provided to the clinician. This will likely be the first published study of a method of using both treatment selection and treatment monitoring decision support in the treatment of CMHDs.

DSTs that can aid treatment monitoring decisions in CMHDs have been developed and are able to identify patients who are at risk of poor outcomes, notifying clinicians that a change in treatment approach may be required. The trajectory of change in symptoms that the patient is expected to follow is generated from the OQ-45 total score but does not use additional patient characteristics to personalise this expected response. There have also been recent developments in predicting which treatment is likely to be most beneficial for individual patients, with some models suggesting that differential response can be identified in over 60% of patients (e.g. DeRubeis et al, 2014; Huibers et al, 2015), and recent models

have also been generated using IAPT samples (Delgadillo et al, 2016) to identify groups of patients who are at higher risk of poor outcomes. However, these models have not been translated into DSTs that are routinely available in clinical practice, or evaluated in treatment services. In addition, no DST has been developed that is able to support both treatment selection and treatment monitoring decisions in CMHDs. A DST that is able to use patient characteristics to recommend which treatment is likely to result in the best outcome for a patient, as well as the expected trajectory of change during this treatment, could support a personalised medicine approach for treatment services such as IAPT.

Using data to inform clinical decision making

As described in Chapter 1, one contributing factor to the increase in personalised medicine approaches in healthcare is the availability of large patient datasets which offer the opportunity to explore the association of a number of patient characteristics across cohorts. Using large datasets in this way provides more power for statistical modelling approaches, which have contributed to a large amount of the success of personalised medicine in health care (Abbassi, 2017).

However, mental health generally lacks large standardised patient datasets that allow the exploration of patient characteristics and outcomes (Ozomaro et al., 2013). Most of the current large mental health datasets available have been used to develop predictive models of diagnoses, typically of depression, from either routine patient data (King, Walker, et al., 2008; Wang et al., 2013) or genetic information (Hyde et al., 2016). However, there has been limited use of large mental health datasets in the prediction of treatment response, with a few exceptions using datasets approaching 1500 (Delgadillo et al., 2016) and 2000 patients (Chekroud et al., 2016).

The standardised approach to ROM used by IAPT services and the use of a minimum dataset has resulted in a large national dataset of patients attending IAPT services. Each IAPT service collects the same standardised patient characteristics for every individual referred to, and receiving treatment from the service. As over 1.25 million referrals are received nationally by IAPT services each year (HCSIS, 2015; NHS Digital, 2016), and patient data from all IAPT services is centrally aggregated for NHS Digital's annual reporting of service outcomes, there is a huge potential dataset if this information were made available to researchers. The average number of referrals to each IAPT service was just over 5000 patients by each clinical commissioning group (CCG) between 2012 and 2013 (HSCIC, 2014), therefore a dataset derived from individual services rather than the whole of the national programme would itself provide a large dataset, and would be considerably larger than most RCTs conducted in CMHDs.

Developing a DST for IAPT services

Research on the use of DSTs in healthcare has shown improvements in clinician performance by incorporating these systems into care, as well as potential benefits in patient outcomes (Bright et al., 2012). There is currently limited research on the use of DSTs to aid treatment selection decisions in CMHDs, despite several potential methods developed in datasets (e.g. Wallace et al., 2013). Given the high prevalence and health burden of CMHDs (Spiers et al., 2016), DSTs could provide significant improvements for both patients and services. Considering the high volume of patients referred to IAPT as well as the potential availability of large patient datasets from these service, there is great potential for the development of a DST to support personalised treatment in IAPT. The growing social and economic costs of treating CMHDs in the UK (McCrone et al., 2008) indicate an increasing need for healthcare resource, and methods to optimise care could help reduce this burden. The development of a predictive DST that is able to support both treatment selection and treatment monitoring decisions in IAPT could be used to allocate patients, based on relevant characteristics, to treatments more effectively and efficiently (Goldburger et al., 2013).

A number of the predictive modelling approaches that have been developed to inform treatment selection in CMHD have focused on specific diagnoses, such as depression, in the included samples (e.g. DeRubeis et al., 2014; Iniesta et al., 2016; Chekroud et al., 2016). However, a DST developed for use in routine treatment services would more beneficial if it were developed for patients across a number of potential diagnoses rather than just one, not only due to the potential limitations of diagnoses (Cuthbert & Insel, 2013) but also due to the high level of comorbidity between CMHDs (Mueller et al, 1999). A further issue with the use of a diagnosis-driven DST in IAPT is the relatively low availability of diagnostic information for patients attending some of these services. National reports indicate that a ‘primary problem’ (either formal diagnosis, or disorder most representing the presenting symptoms) is only available for half of IAPT patients nationwide, though there is significant variation between services with high rates of diagnosis recorded in some and very low rates in others (HSCIC, 2015). The low availability of diagnosis is likely due to the increasing number of self-referrals to IAPT services, who would be unlikely to provide diagnostic information at assessment as this would usually be provided by a healthcare professional. Instead a DST for use in IAPT services would benefit from the use of routine patient characteristics that can be collected through patient self-reporting. This would likely increase data completion, but a DST using patient-report information only would reduce the need for clinician input, potentially freeing up more resource within the services.

IAPT services have adopted ROM, and so there is the potential is to incorporate this sessional symptom measure information in to a DST to inform more personalised treatment monitoring decisions. Treatment monitoring systems in psychological interventions such as the OQ-45 (Lambert, 2001) have not been adopted in IAPT at present. The expected response curves used as part of this system are not tailored to the patient other than being

based on their pre-treatment OQ-45 score, but the identification of patient characteristics associated with the likelihood of being 'not on track' could be used to tailor treatment. The identification of groups of patients with different responses to treatments could be supplemented with the identification of different expected response curves, to enable the development of a DST capable of being used for both treatment selection and treatment monitoring decisions. The elements of a successful DST were found to be provision of decision support as part of workflow, at the time the decision is made, providing a recommendation not assessment and being computer based (Kawamoto, Houlihan, Balas, & Lobach, 2005). Incorporating these elements where possible should therefore be considered in the development of a DST for CMHDs.

The stepped care model adopted by IAPT services provides a clear treatment selection decision (LI or HI) at initial treatment for which a DST could be developed, as well as the need for treatment monitoring decision support to as whether a patient is displaying symptomatic improvement or whether a change in treatment, for example stepping up from low to high intensity, is needed. In addition, IAPT services include a number of clinicians who have more limited experience of working in mental health settings, and could be considered paraprofessionals, and therefore these individuals might benefit most from decision support to improve both patient and service outcomes. The development of such a DST would also provide an opportunity to adopt a more personalised approach to treatment in IAPT services, by accounting for patient-specific characteristics that may be associated with treatment response.

The current allocation of patients to either LI or HI treatment in IAPT services cannot be considered random, as there are a number of clinical reasons why HI treatment may be considered more appropriate for certain patients. For example, for certain diagnoses such as PTSD and social anxiety disorder there is not an established evidence base for the use of LI interventions, and therefore HI treatments are recommended. Therefore, if these diagnoses are identified then it may be expected that clinicians allocate these patients to HI. In addition, it is generally accepted that HI treatments are more appropriate for individuals with more severe presentations of CMHD symptoms (NCCMH, 2011) which may inform some decisions to allocate to more intensive treatment. Therefore, it is possible that confounding by indication, which occurs when clinical information informs the selection a particular treatment also affects the outcome (Kyriacou & Lewis, 2016), may occur in relation to the choice to treatment for some IAPT patients, with an expectation that individuals who receive HI treatments will be more severe in nature than those routinely receiving LI. The measurement of symptom severity is available in IAPT datasets and can be used to explore whether there are differences in the patients receiving the different intensities of treatment, but it is also possible that there are unobservable factors that determine the allocation of treatments in IAPT. These may include patient specific characteristics such as childhood abuse which is not collected in IAPT minimum datasets but is linked to poorer response to

treatment (Williams, Debattista, Duchemin, Schatzberg, & Nemeroff, 2016) but also factors linked to service delivery such as the limited availability of HI therapists and wait-times. These will not be considered in this thesis.

The points raised above also inform the design of the studies presented in this thesis, and the question about whether patient characteristics can be deemed prescriptive or prognostic in nature. A characteristic would be considered prognostic if it were to indicate which patients are at an increased likelihood of an outcome relative to other patients, whereas prescriptive characteristics would indicate differential outcomes from one treatment compared to another (Fournier et al., 2009). These between-treatment comparisons can inform upon which treatment is the best for similar patients, but this prescriptive assumption assumes that all patients allocated to the compared treatments are the same. This assumption more likely holds true to randomised controlled trial (RCT) designs, where the random allocation of patients to treatments is expected to remove potential confounders of treatment choice. The use of data from routine treatment services, where randomisation is not used, instead lends itself to prognostic designs, where within treatments comparisons between patients would be explored. In order to develop a DST that can support decision making in the delivery of psychological interventions, an understanding of which patient characteristics are associated with outcomes is required. The next chapter presents a systematic review of patient characteristics associated with CMHD treatment outcomes, which could be used as part of a DST to support decision making in IAPT services.

Chapter 3. Patient characteristics associated with outcomes.

Abstract

The development of a DST to support a personalised treatment approach in CMHDs requires the identification of patient characteristics which have been associated with outcomes from treatment. A number of routinely available patient-specific characteristics (characteristics of the patient, not of the intervention, clinician or service) have been associated with treatment outcomes, for example, symptom severity and demographic information such as age or ethnicity (Arnow et al., 2007; Warden et al., 2009), and a systematic search of all potential predictors could provide more clinical utility to a DST for use in mental health services. The aim of this chapter was to conduct a meta-review; a systematic review of previously conducted reviews which had explored the associated between patient characteristics and treatment outcomes in people with CMHDs. A total of K=46 systematic reviews were identified in the literature, with K=22 focusing on predictors of outcome following pharmacological interventions, K=15 on predictors of psychological treatment outcome and K=9 considering both types of intervention. A range of patient characteristics were identified across reviews; severity, age, gender and comorbid mental health problems (including personality disorder) were the most frequently identified in reviews focusing on treatment response. Characteristics such as age and gender were inconsistently associated with treatment response across reviews, whereas other characteristics, for example higher pre-treatment symptom severity and the presence of comorbid conditions were more consistently associated with poor response to treatment. Some characteristics including duration of illness, personality characteristics and employment status, were associated with response to treatment but were only identified in a limited number of reviews yet may have potential value for a personalised treatment approaches. There was less available evidence on characteristics associated with treatment dropout, with only one review focusing on psychological interventions and K=9 providing information about associations with dropout from pharmacological interventions. Findings suggested that younger age and fear of stigma were frequently associated with increased dropout, and other patient characteristics were found to have inconsistent association with the likelihood of treatment dropout. The results of this meta-review provide information on the potential value of patient characteristics in predicting treatment outcomes, however the inconsistent associations for a number of characteristics suggests more research is required, and that associations could be specific to the interventions or environments in which treatment was delivered.

Introduction

The variation in outcomes for individuals with CMHDs receiving psychological interventions suggests that a more personalised approach to treatment selection may be of benefit (Cuijpers & Christensen, 2017). An understanding of which patient characteristics are associated with treatment outcomes is required to develop such an approach, as this information would likely inform which characteristics are of value to include in a DST to support treatment decisions. This includes the consideration of patient characteristics associated with dropout treatment, due to the cost associated with non-completion (Oldham et al., 2012). Allocating a patient to a treatment they will likely to not attend or complete would be considered a poor use of resource, and therefore incorporating this information into a potential DST could optimise healthcare delivery, as well as improving patient outcomes. The patient's attitude towards treatment and their experience of any harms (e.g. side effects) can influence whether treatment will be completed, and this is especially true of drug treatments where studies have found that nearly 50% of patients leave treatment within the first 12 months (Sansone & Sansone, 2012; Warden et al., 2009). Although concerns about side effects are a common reason for dropout of drug treatments, the fear of discussing uncomfortable issues in psychological interventions may also lead to the avoidance of treatment (Paige & Mansell, 2013).

A wide range of patient characteristics have been associated with positive outcomes from treatment (defined as 'treatment response' in this chapter), such as a clinically meaningful reduction in symptoms, as well as dropout from treatment. The level of pre-treatment symptom severity of the mental health disorder has regularly been proposed as a predictor of treatment response, but the direction of the association appears inconsistent, with some findings suggesting that higher symptom severity is more predictive of poorer outcomes, whereas others suggest higher severity increases the likelihood of response of treatment (Grammer et al., 2015; Mululo, de Menezes, Vigne, & Fontenelle, 2012; Solomon et al., 2008; Van, Schoevers, & Dekker, 2008). Other characteristics such as age, previous treatment, level of functioning and ethnic groups have also been associated with treatment dropout, as well as response to treatment (Arnow et al., 2007; Papakostas et al., 2003; Warden et al., 2009). Social characteristics, including martial distress or relationship quality have also been linked to outcomes following psychological treatment of depression (Snyder, Castellani, & Whisman, 2006).

Clinical characteristics including the presence of comorbid mental health conditions have also be proposed as predictive of treatment outcome, and researchers have suggested that both comorbid substance use disorders (Parker, Wilhelm, Mitchell, & Gladstone, 2000) and Axis-II disorders (personality disorders) (Newton-Howes, Tyrer, & Johnson, 2006) predict poorer outcomes, which has important implications for clinical practice. Population surveys have found that 20% of individuals meeting clinical criteria for depression, and 15% of those with an anxiety disorder also meet criteria for a substance use disorder (Grant et al., 2004),

and comorbid personality disorder, identified using structured clinical interviews, was found in 50% of outpatients with major depression (Sanderson, Wetzler, Beck, & Betz, 1992; Zimmerman, Rothschild, & Chelminski, 2005).

In addition to demographic and clinical characteristics of the patient, there has been growing interest in the role of biomarkers in predicting response to treatments in CMHD. One area is neuroimaging and the identification of areas of brain activation associated with treatment response. For example, research investigating cognitive biases in CMHDs suggests that abnormal cognitive processing may be involved in the development of symptoms, and therefore the potential focus of targeted treatments (Roiser, Elliott, & Sahakian, 2012). Functional magnetic resonance imaging (fMRI) has allowed researchers to assess levels of brain activation whilst undertaking cognitive processing tasks, such as facial recognition tests, and have found biases in emotional processing in individuals with mental health disorders (Surguladze et al., 2005). Others have suggested that levels of activation in specific brain regions to emotional stimuli can predict the response to both psychological interventions (Doehrmann et al., 2013; Siegle, Carter, & Thase, 2006) and pharmacological treatment (Fu et al., 2004) in depression and anxiety disorders. Further findings suggest that treatment may modify these processing biases in patients (Beard, 2011).

However, the costs and practical complications with administration involved with neuroimaging scanning may be too high to justify their use for all patients entering treatment, therefore limiting their current use in clinical practice (Evans et al., 2006). Further work is required, and there is potential for the development of more simple tests that can extrapolate information on specific brain areas, for example through measuring pupil dilation, which would allow a proxy measurement of brain activation in this area (Graur & Siegle, 2013).

As discussed in Chapter 1, genotyping has underpinned much of the recent advances in personalised medicine, but currently it has led to limited developments in the prediction of outcome in CMHDs (Oestergaard & Møldrup, 2009). A number of genetic markers have been investigated; the majority of these have focused on genes coding for proteins, such as the serotonin transporter (5-HTT) and the serotonin 5-HT-2 receptors (Papakostas & Fava, 2008). However, there has been little success with any translation of these findings into clinical practice, and the limited evidence at present does not support the use of routine genetic screening in routine treatment services for the purpose of treatment prediction in CMHDs (Licinio & Wong, 2011; Munafò et al., 2014).

There has also been research into the potential association between non-patient characteristics and treatment outcomes, for example characteristics of the therapists or treatment delivery that may predict treatment response. The analysis of therapist characteristics has become more increasingly popular, for example therapist adherence to intervention framework (Webb, DeRubeis, & Barber, 2010) and therapist alliance (Strunk,

Cooper, Ryan, DeRubeis, & Hollon, 2012), which would likely be influenced by patient characteristics such as personality disorder (Bender, 2005). Findings suggest that therapist effects may explain around 9% of the variance in outcomes between patients in IAPT services (Green, Barkham, Kellett, & Saxon, 2014), which is above the threshold for clinical importance proposed by researchers (Uher et al., 2012). However, the amount of variance explained seems low when compared to the variance that is explained by within treatment factors, for example the change in symptoms over the first three sessions, which is reported to explain up to 40% of the variance in eventual outcome (Lambert, 2013). Intervention delivery factors, such as more frequent sessions have also been found to improve outcomes in psychological treatment (Erekson, Lambert, & Eggett, 2015).

At present there is a lack of consistent evidence supporting the utility of costly and technically demanding patient data from genetic or neuroimaging procedures for predicting treatment outcomes in CMHDs. Instead, the use of routinely available patient characteristics to inform a personalised treatment approach in mental health services may have more clinical utility and applicability. IAPT services nationally collect the same standardised minimum dataset, and a DST which includes these characteristics would be applicable to all IAPT services in England.

However, it was expected that the review might also identify additional patient characteristics associated with treatment outcomes that were not collected by IAPT. If there was sufficient evidence for the ability of such characteristics to predict treatment outcomes, then this might support the inclusion of these characteristics in the IAPT MDS and a DST to support treatment decisions. The aim of this meta-review was to identify patient-specific characteristics which can be routinely collected in mental health services and have been associated with treatment outcomes in CMHDs. Findings from this review could then be used as the first stage in development of a DST that is able to predict the likelihood of treatment outcomes for patients entering IAPT services.

Method

From scoping searches conducted to identify primary randomised control trials (RCTs) and cohort studies of patient characteristics associated with outcome, it was found that there was a huge number of potential studies, typically post-hoc moderator analyses that were highly variable with regard to quality. It would not have been feasible to consider all this evidence within this thesis, so instead a more focused meta-review (a systematic review of previously conducted systematic reviews) was performed.

Meta-reviews are considered as an appropriate and efficient method of using existing literature where significant evidence exists (Harder, Remschmidt, Haller, Eckmanns, & Wichmann, 2016), and was seen as a method of identifying the best available evidence

within the resource limitations of this thesis. Previously conducted, high-quality systematic reviews arguably include the best available evidence, with a reduced risk of bias with regard to included studies and patient characteristics (Egger, Smith, & Altman, 2001). The review protocol is presented in Table 3.1 and described below. The search was first completed in July 2012, and then the same search string was repeated from July 2012 to February 2015 to identify additional reviews that had been published in the intervening period.

Table 3.1. Review protocol

Electronic databases	EMBASE, MEDLINE, PsycINFO
Date searched	Search 1 - Database inception to July 2012 Search 2 - July 2012 to February 2015
Study design	Systematic reviews
Population	Participants receiving pharmacological, psychological or a combination of the two treatments for common mental health disorders, which was the primary diagnosis.
Excluded populations	Diagnosis of severe mental illness (e.g. schizophrenia). Primary diagnosis of eating disorders or substance use disorders. Patients in remission before intervention (e.g. Relapse prevention).
Interventions	Any psychological or pharmacological intervention for common mental health disorders.
Included characteristics	Any patient specific characteristics, identifiable at pre-treatment assessment and/or interview (e.g. age, severity of illness).
Excluded characteristics	Non-patient characteristics (e.g. specific intervention type, therapist characteristics). Characteristics unidentifiable in routine assessment / interview (e.g. requiring neuroimaging data, genotyping).

Although the primary focus of this project is patient characteristics associated with treatment response to psychological interventions, scoping searches of the literature showed that many reviews combined predictors of both pharmacological and psychological interventions together. The presentation of these reviews made separating out psychological and pharmacological treatment predictors impractical as results were often combined, and therefore it was decided to include reviews of pharmacological treatment in addition to psychological. Including pharmacological interventions would also provide more information about potential patient characteristics associated with outcomes in CMHDs. This meta-review across types of intervention was conducted first, before a more detailed synthesis including only the reviews of patient characteristics associated with outcomes in psychological treatment was performed.

Study selection

The included clinical population was adults diagnosed with CMHDs, who were undergoing treatment for these disorders. The CMHD diagnoses considered were depressive disorders, panic disorder, generalised anxiety disorder (GAD), social anxiety disorder, obsessive-compulsive disorder, specific phobias and post-traumatic stress disorder.

Systematic reviews including populations with comorbid psychological illnesses (e.g. substance use disorders or personality disorders) were included, provided the CMHD was the primary diagnosis. Excluded populations were participants diagnosed with severe mental illness (such as psychosis or bipolar disorder) and reviews where interventions were conducted on participants already in remission from mental illness (such as relapse prevention interventions). Some reviews were conducted across all mental health conditions, and these were included as long as information on predictors of outcome in CMHDs could be extracted separately from the reporting.

The meta-review included only patient-specific characteristics that may be associated with treatment outcomes, including dropout. 'Patient characteristics' was defined as information related to the patient (not to the treatment or clinician) that could be routinely collected in mental health services, and made available to a clinician performing the pre-treatment assessment, for example level of symptom severity, comorbid conditions and employment status. Characteristics that need additional resource to collect, such as neuroimaging data or genotyping were therefore excluded. Characteristics related to the therapist, intervention or service (e.g. therapist alliance, type of intervention) were also excluded. Focusing on only patient-specific characteristics was important for the development of a personalised approach to treatment decisions in IAPT, as a DST developed for this thesis could only include characteristics that could be easily collected by both the patient and clinician.

Search strategy

The search strategy for this meta-review is presented in Appendix A, and was based on search strings originally developed for the NICE clinical guideline on CMHDs (NICE, 2011b), which included a review of recent systematic reviews (from 1993-2010) on treatment predictors. For the present review, the search strings were amended to include additional keywords and MeSH terms around decision making, algorithms and patient satisfaction, in addition to removing the lower limit on the date of publication (therefore including all titles since database inception).

Data extraction

Data was extracted from all included reviews by the author. The data extracted included: Information about the clinical diagnosis of the CMHD population that was under investigation, the type of intervention (psychological, pharmacological or both), whether response or dropout was the primary outcome explored by the review, the type of primary studies included by the review (controlled or uncontrolled studies) and the patient characteristics included in the review. Data was then extracted about the association between each included patient characteristic and treatment outcomes.

Review Quality

An assessment of quality and risk of bias was completed for each review using the methodology checklist: systematic reviews and meta-analyses adopted by NICE (NICE, 2009) (see Appendix B). The checklist includes five areas that are important to consider in a well-conducted review and asks the assessor to determine whether the included review satisfies each statement with “Yes”, “No” or “Unclear”. The five areas are: clarity of the research question, relevance of included studies, the coverage of the literature searches, whether study quality is reported and that the methods used are adequately reported.

Each review was given a quality score made up of the responses to the five checklist items, where a ‘Yes’ resulted in 1 point, ‘no’ resulted in 0 points and ‘unclear’ resulted in 0.5 points. Reviews were therefore scored out of a max of five points over the five checklist items. Scores of 4.5 and above were designated “very high” quality, scores of 4 were classed as “high quality”, scores of 3 and above were rated “moderate” quality and below 3 considered “low” quality. Any review that was considered as low quality following assessment was excluded from the meta-review.

Results.

Search 1.

Literature Search 1 identified 5206 citations. From initial screening of the titles and abstracts of these papers, 4892 citations were considered not relevant to the current meta-review. The remaining 314 papers were assessed for eligibility. From reviewing these studies, it was found that 39 were either books, conference reports or book chapters, 28 were focused on excluded populations, 113 were considered not relevant to the review as they included no relevant outcomes/predictive characteristics (for example genetics or biomarkers), and a further 92 papers were excluded as they were not systematic reviews.

The remaining 42 articles were judged relevant to the review question, and were assessed for quality and risk of bias using the methodology checklist. After assessing the systematic review quality, five of the systematic reviews were rated as poor quality, with a very high risk

of bias, and therefore excluded. This left 37 systematic reviews from search 1 to be included in the synthesis. A PRISMA (Moher, Liberati, Tetzlaff, Altman, et al., 2009) flow diagram for Search 1 is presented in figure 3.1.

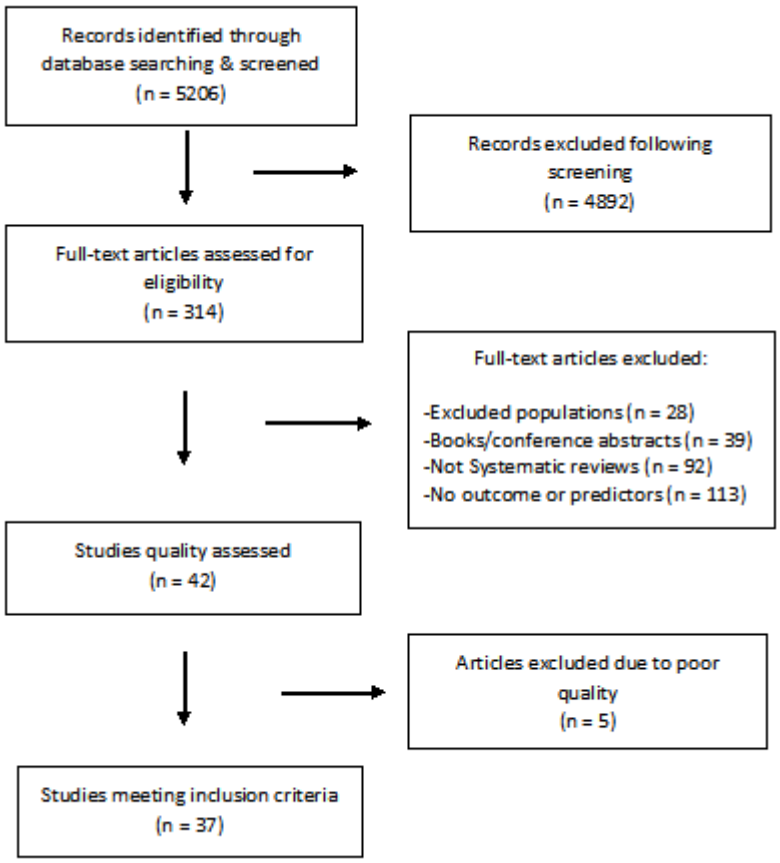

Figure 3.1. Flow diagram for search 1

Search 2.

Search 2 identified 2184 citations which were published between July 2012 and February 2015, of which 2062 were excluded following initial screening. Of the remaining 122 studies, 108 papers were excluded: 17 were not systematic reviews, 21 focused on excluded populations, 7 were from books or conference abstracts and 63 did not include any relevant patient characteristics.

The remaining 14 reviews were quality assessed using the NICE methodology checklist, and 4 studies were found to be of poor quality, with a very high risk of bias, and were excluded. The remaining 10 reviews were included in the synthesis. A PRISMA flow diagram for search 2 is presented in figure 3.2.

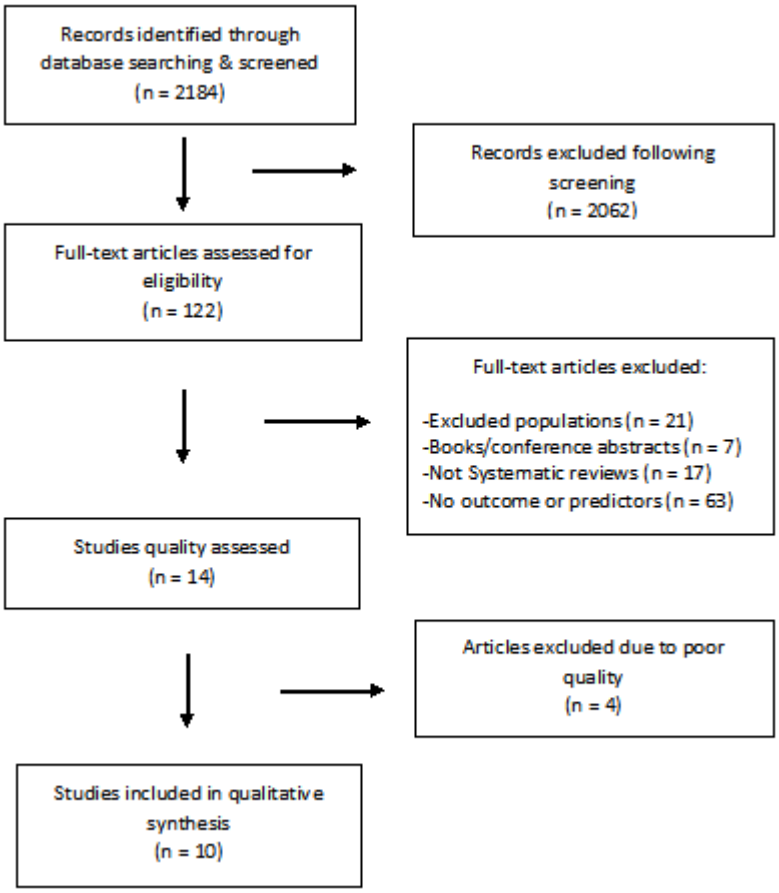

Figure 3.2. Flow diagram for search 2.

Characteristics of included studies

The systematic reviews identified from Search 1 and Search 2 were combined to produce the full list of included studies for this meta-review. It was found that one review in Search 1 (Pompili et al., 2009) had been updated in Search 2 (Pompili et al., 2013), and to avoid double counting information on patient characteristics, only the updated review was included. Thirty six (36) reviews were included from Search 1 and 10 reviews from Search 2, resulting in 46 included systematic reviews. Data was extracted from the reviews on any patient-specific characteristic where the association to either treatment response or dropout was reported. Additionally, information on the type of interventions, type of included primary studies and diagnoses of the populations included in the reviews was also extracted.

Review Quality

The NICE methodology checklist was completed for each included review. Of the K=46 systematic reviews that were included, only K=3 were rated as 'very high' quality using the methodology checklist. K=25 were rated as 'high' quality, and the remaining K=18 were viewed as 'moderate' quality. Reviews were typically downgraded to moderate or high quality, instead of very high quality, due to unclear literature searches or a lack of information about how primary studies were assessed for risk of bias.

The characteristics of included studies are presented in Table 3.2 below. Information is presented on:

- Type of intervention: Either psychological or pharmacological
- Quality of the systematic review with regard to risk of bias after completion of NICE checklist (Appendix B).
- The treatment outcome assessed: either treatment response or dropout
- The primary diagnosis of the population.
- Type of included primary studies: either controlled or uncontrolled.
- The patient related characteristics identified/included in the review.

Table 3.2. Key study characteristics of included reviews

Study ID	Intervention (Psychological or Pharmacological)	Review quality	Treatment Response or Dropout	Primary diagnosis of population	Types of Included primary studies	Patient characteristics identified
(1) Ackerman2002	Pharmacological	High	Response	Anxiety (OCD)	Controlled	Age of participant Age of onset/duration of illness Severity
(2) Al-Qasem2011	Pharmacological	Moderate	Dropout	Medical conditions (including CMHDs)	Controlled & uncontrolled	Gender Age of participant Patient perception of treatment Previous treatment Education Relationship status
(3) Christensen2009	Psychological (internet based interventions only)	Moderate	Dropout	Depression and Anxiety disorders	Controlled	Age of participant Severity Previous treatment Education
(4) Cuijpers2008	Psychological	High	Response	Depression	Controlled	Gender Age of participant Severity
(5) Cuijpers2009	Psychological	High	Response	Depression	Controlled	Age of participant
(6) Cuijpers2014	Psychological and Pharmacological	High	Response	Depression	Controlled	Gender
(7) Davis2006	Pharmacological	Moderate	Response	Anxiety (PTSD)	Controlled & uncontrolled	Comorbidity Severity Subtype (trauma type)

(8) Davis2014	Pharmacological	High	Response	Anxiety (social anxiety)	Controlled	Gender Age of participant
(9) Dodd2004	Pharmacological	High	Response	Depression and Anxiety disorders	Controlled & uncontrolled	Comorbidity – other MH Severity Personality factors (large range) Functioning
(10) Duncan2010	Psychological	Very high	Response	All mental health (including CMHDs)	Controlled & uncontrolled	Shared decision making
(11) Fekadu2009	Psychological and Pharmacological	Very high	Response	Depression (treatment resistant)	Controlled & uncontrolled	Gender Age of participant Previous treatment
(12) Fournier2010	Pharmacological	High	Response	Depression	Controlled	Severity
(13) Gava2007	Psychological	Very high	Response	Anxiety (OCD)	Controlled	Severity
(14) Haby2006	Psychological	Moderate	Response	Depression and Anxiety disorders	Controlled	Severity
(15) Haug2012	Psychological	High	Response	Anxiety disorders (any)	Controlled	Gender Age of participant
(16) Ipser2012	Pharmacological	High	Response	Anxiety (PTSD)	Controlled	Gender
(17) Johnston2013	Pharmacological	High	Dropout	Depression	Controlled & uncontrolled	Age of participant Previous treatment Shared decision making Patient perception/stigma
(18) Joosten2008	Pharmacological	High	Response and Dropout	Medical conditions (including CMHDs)	Controlled	Shared decision making

(19) Knopp2013	Psychological	High	Response	Anxiety (OCD)	Controlled	Age of participant Comorbidity – other MH Severity Education Employment Relationship status Medication use
(20) Kool2005	Pharmacological	High	Response	Depression	Controlled	Comorbidity – Personality disorder
(21) Levy2011	Psychological	Moderate	Response	All mental health (including CMHDs)	Controlled & uncontrolled	Attachment style
(22) Lingam2002	Pharmacological	Moderate	Dropout	Depression	Controlled & uncontrolled	Gender Age of participant Previous Treatment Patient perception/stigma
(23) Luszcyznska2009	Psychological	Moderate	Response	Anxiety (PTSD)	Controlled & uncontrolled	Personality – self efficacy
(24) Malpass2009	Pharmacological	High	Dropout	Depression	Controlled & uncontrolled	Patient perception/stigma
(25) Mancini2010	Pharmacological	High	Response	Anxiety disorders (any)	Controlled & uncontrolled	Comorbidity – Other MH
(26) McPherson2005	Psychological	Moderate	Response	Depression (treatment resistant)	Controlled & uncontrolled	Previous treatment
(27) Mulder2002	Psychological and Pharmacological	High	Response	Depression	Controlled & uncontrolled	Comorbidity – personality disorder Personality factors
(28) Mululo2012	Psychological and Pharmacological	High	Response	Anxiety (Social Anxiety Disorder)	Controlled & uncontrolled	Gender Age of participant Age of onset/duration of illness Comorbidity – personality disorder Comorbidity – Other MH Severity Duration of illness Previous treatment Subtype Education Relationship status

(29) Najt2011	Psychological and Pharmacological	Moderate	Response	All mental health (including CMHDs)	Controlled & uncontrolled	Gender Comorbidity- Other MH
(30) Nanni2012	Psychological and Pharmacological	Moderate	Response	Depression	Controlled & uncontrolled	Age of participant Childhood maltreatment
(31) Naudet2011	Pharmacological	Moderate	Response	Depression	Controlled & uncontrolled	Age of participant Severity
(32) Nelson2013	Pharmacological	Moderate	Response	Depression	Controlled	Gender Age of participant Severity Duration of illness
(33) Nunes2004	Psychological and Pharmacological	High	Response	Depression	Controlled	Age of participant Severity
(34) Ormerod2008	Pharmacological	Moderate	Dropout	All mental health (including CMHDs)	Controlled & uncontrolled	Ethnicity
(35) Pompili2013	Pharmacological	Moderate	Dropout	Depression	Controlled & uncontrolled	Age of participant Severity Comorbidity - personality disorder Patient perception/stigma
(36) Powers2010	Psychological	High	Response	Anxiety (PTSD)	Controlled	Age of onset/duration of illness Subtype (trauma type)
(37) Rivero-Santana2013	Pharmacological	Moderate	Dropout	Depression	Controlled & uncontrolled	Gender Age of participant Severity Ethnicity
(38) Sanchez-meca2010	Psychological	High	Response	Anxiety (Panic disorder)	Controlled	Gender Age of participant Age of onset/duration of illness Comorbidity – Other MH

(39) Serretti2009	Pharmacological	Moderate	Response	Depression and Anxiety	Controlled & uncontrolled	Gender Age of participant Age of onset/duration of illness Comorbidity – Other MH Comorbidity – Personality disorders Physical illness Severity Personality factors (range) Previous treatment Education Ethnicity Employment Relationship status Functioning Social Support
(40) Serretti2011	Pharmacological	High	Response	Depression	Controlled	Gender Age of participant Age of onset/duration of illness Ethnicity
(41) Silveira2013	Psychological	High	Response	Depression	Controlled & uncontrolled	Age of participant Severity
(42) Sullivan2005	Psychological and Pharmacological	High	Response	Depression	Controlled & uncontrolled	Comorbidity – Other MH
(43) Tedeschini2011	Pharmacological	High	Response	Depression	Controlled	Age of participant
(44) Thiel2013	Psychological and Pharmacological	Moderate	Response	Anxiety (OCD)	Controlled & uncontrolled	Comorbidity – Personality disorders
(45) Trivedi2011	Psychological	High	Response	Depression (treatment resistant)	Controlled	Previous treatment
(46) Yonkers2002	Pharmacological	Moderate	Response and Dropout	Depression	Controlled & uncontrolled	Gender

A total of K=15 reviews focused on only predictors of treatment outcome in psychological interventions, of which K=14 were concerned with treatment response and one review on predictors of dropout during psychological interventions. Pharmacological interventions were the sole focus of K=22 reviews (K=13 response only, K=7 dropout only and K=2 both response and dropout included), and the remaining K=9 included both psychological and pharmacological interventions, all of which reported on treatment response only.

The included diagnoses also varied considerably between the reviews. K=11 reviews included specific anxiety disorders only (OCD K=4, PTSD K=4, social anxiety disorder K=2, Panic disorder K=1), K=2 focused on any anxiety disorder, K=23 considered only patients with depression and K=4 reviewed both depression and anxiety disorder. The final K=6 reviews reported on characteristics associated with outcomes for any physical and/or mental health condition, although depression and anxiety disorders were separated out in the reviews for data extraction. Of the included reviews, K=21 included only controlled trials, whereas the remaining K=25 included both controlled and uncontrolled studies in their included primary studies. Uncontrolled studies included cohort and case series studies of populations with CMHDs.

The number of patient characteristics reported within each review ranged significantly, with a number of reviews focusing their research question to just one particular characteristic (e.g. Yonker2002), whereas other reviews ranged from two to 15 characteristics (e.g. Serretti2009).

Data synthesis (psychological and pharmacological interventions)

Only a third of included reviews were focused exclusively on psychological treatments, and therefore the first synthesis of the data included reviews on both pharmacological and psychological interventions. This initial stage allowed the synthesis of all available information on patient characteristics associated with outcomes, and is supplemented by a more focused synthesis of data from reviews of psychological interventions, which is presented further below.

Information about included patient characteristics (e.g. severity, previous treatment) varied considerably between reviews with regard to methods of study selection, reporting, aggregating and synthesising. Reviews that included only controlled intervention studies were more likely to have used quantitative methods (e.g. meta-regression) to statistically evaluate the association between patient characteristics and treatment outcomes (e.g. Ackerman2002; Cuijpers2009). However, as less than 50% of included reviews restricted their inclusion criteria to only controlled trials, quantitative analysis was not conducted in the majority of reviews. In most reviews the information on predictors of treatment outcome was

presented narratively (e.g. Davis2006; Mancini2010). Data on predictors could not be extracted from statistical reporting in these reviews, and instead the association of specific characteristics with outcome was determined by exploring the narrative review provided within the study.

Due to the lack of poolable quantitative data across the included systematic reviews, a narrative approach was adopted for this meta-review, to make best use of available information from the identified systematic reviews on patient characteristics associated with treatment outcomes (Popay et al., 2006). For each patient characteristic identified across the reviews, information was extracted first on the frequency of it's reporting in the literature, regardless of association with outcome.

Next, data was extracted for all patient characteristics explored within the included reviews about the reported association with treatment outcome. Evidence considered within the reviews could suggest that each patient characteristic: was associated with outcomes (as well as the direction of the association), had an inconsistent association (varied amongst included primary studies) or that there was limited evidence of any association (unlikely to predict outcome). For studies using meta-regression or other more quantitative methods, the results of statistical analyses were used to inform the decision as to whether there was an association or not, whereas qualitative analysis was used to inform associations in reviews which were more narrative.

Characteristics associated with treatment response

Table 3.3 presents the patient characteristics which were explored in relation to treatment response. Age of participant and gender were amongst the most frequently reported patient characteristics in the literature, but both were found to have little association overall with treatment response. The severity of illness at baseline and the presence of comorbid mental health issues (including both personality disorders and other mental health conditions) were also frequently explored and were often associated with treatment response. It was generally suggested in the reviews that the presence of a co-morbidity and higher initial severity of illness were both linked to poorer outcomes, although a few reviews indicated that higher severity had a positive effect on treatment outcome. Both these characteristics have been promoted as potential predictors by researchers (Shea et al., 1990; Van et al., 2008), and findings of this review suggest that these characteristics could inform likely treatment response.

Patients reporting previous treatment for a mental health illness were associated with poorer response to current treatment in four of the five reviews which considered this characteristic, and could be a useful predictor for clinical purposes considering this level of agreement across reviews. Duration of illness, however, had a much more inconsistent association with treatment response, with two of six reviews finding no association, one review suggesting a

positive effect of longer duration on outcome and three reviews suggesting that a longer duration has a negative association with response.

The remaining patient characteristics were mentioned less frequently in the literature, with three or less systematic reviews providing information about them. Despite the lack of previous research, these characteristics may have clinical utility in predicting treatment response. Both older age of onset and being in a relationship where associated with better response to treatment in two of three reviews. Higher self-efficacy/directedness was also found to be consistently associated with positive response to treatment in all three reviews reporting this factor. All three of these characteristics can be assessed using simple direct questions or self-reported measures (such as the Self-efficacy scale (Sherer & Maddux, 1982)) which may be of value to consider in clinical practice, especially as the reviews noted these associations across different diagnostic groups.

Other characteristics associated with positive treatment response reported in the literature were higher level of interpersonal functioning, lower neuroticism, higher co-operation and secure attachment in childhood. Patient characteristics such as ethnicity, physical illness, medication use, social support level and employment were mentioned rarely in the literature, and reviews including these characteristics suggested that the association with response was limited or inconsistent.

Shared decision making was associated with better response in one systematic review, and no association in a second review. Although this characteristic could be viewed as a description of the assessment process rather than a true patient-specific characteristic, the emphasis of shared decision making in IAPT would suggest that role of this characteristic would be important to consider when developing a DST to promote personalised medicine approaches in these services.

Characteristics associated with treatment dropout

Table 3.4 displays the patient related characteristics identified in reviews of treatment dropout. The role of patient characteristics associated with treatment dropout have been considerably less studied than those involved in treatment response, and only nine of the K=46 included reviews reported any information on patient specific predictors of treatment dropout. It should be noted that all but one of these reviews included only pharmacological interventions, and the one review of psychological interventions was focused on internet-delivered therapies only. Therefore their relevance to the broader range psychological treatments delivered in IAPT services is more questionable.

Table 3.3. Characteristics associated with treatment response.

Patient characteristic	Number of reviews reporting	Reviews concluding associations between factor and treatment outcome	Inconsistent association	No association found
Age of participant	K=17	K = 4 - K=2 older age positive - K=2 younger age positive	K=1	K = 12
Severity	K=14	K=11 - K=4 higher severity positive - K=7 lower severity positive		K = 3
Gender	K=13	K = 3 - K=2 Males negative - K=1 Female negative	K=2	K=8
Comorbidity – Other MH	K=9	K = 9 - K=8 comorbidity negative		K=1
Duration of illness	K=6	K=4 - K=3 longer duration negative - K=1 longer duration positive		K=2
Comorbidity _ Personality disorders	K=5	K = 3 Comorbidity negative	K=1	K=1
Previous treatment	K=5	K=4 Previous treatment negative		K=1
Education	K=3	K=1 Less education negative		K=2
Age of onset	K=3	K=2 Older age positive	K=1	
Relationship status	K=3	K=2 Married/in relationship positive		K=1
Self-efficacy / Self-directedness	K=3	K=3 Higher efficacy positive		
Illness / trauma subtype	K=3	K=2 -K=1 combat trauma positive -K=1 social anxiety negative		K=1
Ethnicity	K=2	K=1 Caucasian positive	K=1	
Functioning	K=2	K=2 Lower functioning negative		
Shared decision making	K=2	K=1 Shared decision making positive		K=1
Neuroticism	K=2	K=2 Higher neuroticism negative		
Cooperation	K=2	K=2 More cooperation positive		
Employment	K=2	K=1 Unemployed negative		K=1
Physical illness	K=1	K=1 Negative		
Medication use	K=1			K=1
Social Support	K=1			K=1
Childhood maltreatment	K=1	K=1 Maltreatment negative		
Attachment style	K=1	K=1 Secure attachment positive		

Of the K=9 systematic reviews reporting characteristics associated with treatment dropout, K=5 explored the patient's initial perception of treatment and potential fear of stigma as a potential characteristic associated with treatment dropout. Four of these five reviews found that negative perceptions of treatment (e.g. concern about side effects), or fears around the potential stigma of receiving treatment for mental health conditions were associated with increased dropout from treatments. This finding would suggest that improving patient understanding of CMHDs to reduce their fears of stigma and potential side effects could increase adherence to treatment. Age of patient was considered in the majority of included reviews, and results suggested that younger patients were more likely to drop out of treatment. It is of interest that age was found to have an inconsistent association with response, but a more clear association with treatment dropout. The association between previous treatment and dropout also appears to differ from the association between this characteristic and treatment response, as patients reporting previous treatment were associated with more dropout (a poorer outcome) than patients who had not received previous treatment.

Table 3.4. Characteristics associated with treatment dropout.

Patient characteristic	Number of reviews reporting	Reviews concluding associations between factor and treatment dropout	Inconsistent association	No association found
Age of participant	K=6	K = 5 - K=4 older age positive - K=1 younger positive	K=1	
Patient perception of treatment / stigma	K=5	K=4 Negative view has negative impact on dropout	K=1	
Gender	K=4		K=4	
Previous treatment	K=4	K=4 -K=2 previous worse, -K=2 Previous positive		
Severity	K=3	K=2 -K=1 lower severity positive -K=1 lower severity negative	K=1	
Education	K=2	K=1 Lower education level negative		K=1
Shared decision making	K=2	K=2 Shared decision making positive		
Ethnicity	K=2	K=1 Caucasian positive		K=1
Relationships	K=1	K=1 No relationship negative		
Comorbidity – personality disorders	K=1	K=1 Comorbidity negative		

A number of demographic characteristics were also considered in reviews of treatment dropout. Gender was found to have no association with dropout in the literature, which was similar to findings in the reviews of characteristics associated with treatment response. Not being in a relationship and having a comorbid condition were also associated with an

increased risk treatment dropout. Whereas high pre-treatment symptom severity was more frequently associated with poorer response to treatment, severity appeared to have a more inconsistent association with dropout. The three reviews that explored severity in relation to treatment dropout each reported different associations.

Although patient characteristics associated with dropout have received limited attention in previous systematic reviews, some factors such as age and patient perception of treatment were found to be consistently associated with dropout. Younger patients and patients with more concerns about treatment, or stigma associated with mental health treatment were associated with more dropout from mental health treatment. Other patient characteristics were considered in a limited number of previous systematic reviews, and as a result further research would be of value to understand the association between these characteristics and the risk of treatment dropout.

Data synthesis (psychological interventions only)

In the next stage of analysis, only reviews of characteristics associated with outcomes following psychological interventions were included. A more focused review of psychological interventions was expected to provide more relevant information for predicting outcomes in IAPT services, but it could also explain some of the inconsistencies in findings for patient characteristics indicated in the previous analysis. Data on patient characteristics was extracted for each review, and the assessment of review quality was also considered to grade the quality of the evidence for each predictor.

Characteristics associated with treatment response

All included reviews of patient characteristics associated with treatment response in psychological interventions are presented in Table 3.5. Each review is listed, as well as each patient characteristic that was associated with treatment response in the review. The notation (+) is used when the patient characteristic was is found to be positively associated with treatment response in the review, (-) is used when the association was negative and (o) is used to indicate that no association was found, or that the association was inconsistent. The number of notations is used to indicate the strength of the evidence of this finding, derived from the quality of the review: one notation (+) indicates that the evidence is extracted for a moderate quality review, two notations (++) for high quality evidence and three notations (+++) indicates evidence from a very high quality review.

There were K=2 systematic reviews focusing on psychological interventions that were rated as very high quality using the NICE methodology checklist (Duncan2002, Gava2007). These reviews were focused on only one patient characteristic. Duncan2002 found that although

shared decision making had a positive effect on patient satisfaction with care, it had no association with improved outcomes across depression or anxiety disorder treatment. Gava2007 focused exclusively on the association of severity with treatment response in OCD and found that lower severity levels at baseline were associated with better response to psychological treatment. These reviews were conducted according to Cochrane standards and the evidence provided by these reviews was considered very strong.

The high quality reviews (K=8) provided the most information on patient characteristics associated with treatment response. Age was included as a factor in K=6 reviews, although four of these reviews concluded a limited or inconsistent association between age and outcome. The two reviews (Cuijpers2008; Silveira2013) which indicated an association between age and response both suggested that younger patients were more likely to have poorer outcomes. These two reviews were focused on depressed populations only, whereas three of the reviews finding no association included only anxiety populations, and it is possible that the differences here may be linked diagnosis/clinical presentations of patients. Three of the included reviews considered gender as a potential predictor: two of the reviews (Cuijper2009; Sanchez-Meca2010) found no association between gender and outcome, whereas one suggested that females had poorer response to anxiety disorder treatment (Haug2012).

Baseline severity of illness was explored as a predictor in three of the high quality systematic reviews of psychological interventions. Although one review (Cuijper2008) found no association between severity and response to depression treatment, two reviews (Knopp2013; Silveira2013) found that lower severity was associated with better response to treatment in OCD and depression populations. Lower pre-treatment severity was also associated with better response to treatment in the very high quality review (Gava2007), and this may suggest that this characteristic may be important when predicting response to treatment.

The presence of comorbid mental health conditions was associated with less response to treatment in two high quality studies (Knopp2013; Sanchez-meca2010) and type of trauma subtype was not associated with response to PTSD treatment in the one high quality review exploring this factor (Powers2010).

The duration of illness was included in two high quality reviews, with inconsistent findings. One review found that there was no association between time since traumatic experience and treatment response in PTSD patients (Powers2010) whereas a review of panic disorder suggested a longer duration of illness was associated with poorer outcome (Sanchez-meca2010). This impact of this characteristic on treatment response by differ between diagnoses. One high quality review included previous treatment, and the findings suggested that previous treatment was linked to poorer response (Trivedi2011).

Table 3.5. Included reviews: response to psychological treatment.

Study ID	Primary diagnosis of participants	Younger age	Female	Lower severity	Presence of Comorbidity	Duration of illness	Previous Treatment	Trauma subtype	Shared decision making	In a relationship	Higher education	Using medication	Employed	High self-efficacy	Secure attachment
Very high quality reviews															
Duncan2010	All Mental health								ooo						
Gava2007	Anxiety (OCD)			+ + +											
High quality reviews															
Cuijpers2008	Depression	- -	oo	oo											
Cuijpers2009	Depression	oo													
Haug2012	Anxiety disorders	oo	- -												
Knopp2013	Anxiety (OCD)	oo		+ +	- -					+ +	oo	oo	+ +		
Powers2010	Anxiety (PTSD)					oo		oo							
Sanchez-Meca2010	Anxiety (Panic disorder)	oo	oo		- -	- -									
Silveira2013	Depression	- -		+ +											
Trivedi2011	Depression						- -								
Moderate quality reviews															
Haby2006	Depression and Anxiety			+											
Levy2011	All Mental health														+
Luszcyznska 2009	Anxiety (PTSD)													+	
McPherson 2005	Depression						-								

"+" = Positive association; "-" = negative association; "o" = No association

One high quality review considered a range of demographic characteristics associated with treatment response to psychological interventions in addition to age, and was limited to OCD populations (Knopp2013). The findings of this review suggested that being in a relationship and in employment were associated with better response to treatment, whereas level of education and receiving medication in additional to psychological interventions had no impact on outcomes.

There were only K=4 moderate quality reviews of psychological interventions, and each review was focused on only one patient characteristic. The number of previous treatments was associated with poorer outcome to depression (McPherson2005), supporting the finding from the higher quality review (Trivedi2011). Severity of baseline symptoms was reported in the Haby2006 review, where findings supported both high and very high quality reviews that lower severity was associated with better response to psychological treatments. The Levy2011 and Luszcyznska2009 reviews focused on self-efficacy and attachment style respectively, and findings from these reviews suggested that higher self-efficacy and a more secure attachment were associated with better response to treatment.

Characteristics associated with treatment dropout

From the K=15 systematic reviews that focused on predictors of psychological treatment outcomes, only one review looked at patient characteristics associated with treatment dropout (Christensen2009). As this review included only internet-based psychological interventions, no systematic reviews of patient characteristics associated with treatment dropout in face-to-face psychological interventions were identified in the literature search.

The findings from the Christensen2009 review indicated that younger participants were more likely to stay in depression treatment, although no evidence was found in relation to anxiety disorders. As the interventions were all delivered by the internet, it is possible that younger participants would feel more comfortable with this method of treatment delivery, which would explain these findings. Lower pre-treatment severity of illness was associated with less dropout in both depression and anxiety studies. Education was also explored in this review, although no association with dropout was found.

Discussion

The findings from this meta-review have identified a number of patient characteristics that may be associated with treatment response and dropout, in addition to characteristics that appear to show little association or a more inconsistent association across studies and populations. There was less available evidence on characteristics associated with response to psychological interventions compared to pharmacological interventions, yet a number of

characteristics were frequently associated with response in these treatments, which could inform a personalised medicine approach in IAPT services. Only one review explored characteristics associated with dropout from psychological treatments, and therefore it is possible that there are important differences between characteristics associated with dropout in pharmacological and psychological treatments that could not be identified in this meta-review. Furthermore, the range of populations, included studies and methods of analysis of primary studies included in the reviews synthesised in this chapter may explain some of the inconsistencies in findings.

Review of findings: Treatment response

Higher pre-treatment severity of symptoms and the presence of comorbid mental illness were associated with a poorer response to treatments in the majority of reviews which considered these patient characteristics. Both high and very high quality reviews that focused on psychological interventions identified these relationships, which suggests that more intensive treatment might be recommended for more severe/complex patients. Four reviews suggested that higher pre-treatment severity increased the likelihood of positive outcomes (rather than reduced them), but as these reviews all included pharmacological treatment, this may suggest a more inconsistent relationship in drug treatments for CMHDs.

Gender appeared to have little association with treatment response, and the age of patient appears to have an inconsistent association with this outcome, which was evidenced across a large number of reviews which included this characteristic. Most reviews of psychological interventions found no relationship between age and outcome, although two reviews (Cuijpers2008; Silveira2013) did find that older age was associated with better outcomes in the treatment of depression. Two reviews of pharmacological interventions found that higher age was associated with poorer outcomes (Tedeschini2001; Naudet2011) which may suggest a differential relationship between age and type of intervention. However, as K=12 further reviews of pharmacological treatment suggested no association between age and outcome, it may be that age only influences treatment outcomes under certain circumstances, for example in the presence of other patient characteristics, although this was not explored in the current review.

Participants reporting previous treatment for a CMHD were more likely to have a poorer response to treatment across diagnoses. Two reviews (McPherson2005; Trivedi2011) specifically focused on treatment resistant depression, which may explain why previous treatment was frequently associated with worse outcome in these reviews. It would be of more value if characteristics that were associated with treatment resistance could be identified in routinely collected data, and there is potential that previous treatment may be linked to other characteristics that could help estimate the likelihood of successful treatment. Duration of illness was explored in relation to treatment response in six reviews, with

inconsistent findings. Three reviews found that a longer duration of illness was associated with poorer outcomes to treatment for depression (Serretti2011), panic disorder (Sanchez-meca2010) and across CMHDs (Serretti2009). However, no association was found in reviews of social anxiety (Mululo2012) and for PTSD (Power2010), whereas longer duration was associated with positive response to depression treatment in one review (Nelson2013). It is unclear whether the inconsistency is due to the different CMHDs under review, variation in the quality of the interventions delivered in primary studies or whether other patient characteristics (for example severity) may have further influenced findings.

Some patient characteristics appear to have potential value in predicting treatment outcomes but were only considered in a limited number of reviews. For example, being in a relationship was associated with positive treatment outcomes suggesting potential to inform clinical decision making, but it was only considered in three of the included reviews. Other characteristics such as ethnicity, self-efficacy and level of functioning also had limited evidence, but the results presented in this Chapter suggest that they may have potential utility in a DST to inform treatment decisions in CMHDs.

Although it could be argued that shared decision making is not a patient-specific characteristic, it is an important part of IAPT delivered care and therefore should be considered as part of a personalised medicine approach developed in this thesis and was expected to improve outcomes following treatment. However, results were mixed with one review of pharmacological interventions for depression suggesting better outcomes following shared decision making, whereas a review of response in psychological interventions suggested there was no association between sharing decisions and outcomes (Duncan2010). It is possible that shared decision making is more important when considering pharmacological interventions rather than psychological, and this may be linked to concerns about side effects and perceptions that are associated with drug treatments.

Review of findings: Treatment dropout

The number of reviews exploring patient characteristics associated with treatment dropout was significantly lower than those which focused on treatment response. The main characteristics associated with treatment dropout in the literature were the patient's perception of treatment and the possible stigma attached to receiving the intervention, as well as the age of the participant. Concerns around stigma and potential side effects of mental health treatment were all identified in systematic reviews of pharmacological interventions for CMHDs. Although the sole study of psychological interventions and dropout did not mention this characteristic, it is possible that the preference for psychological over pharmacological treatment (Eiring et al., 2015; Williams et al., 2016) reduces the impact of patient perception on dropout in psychological treatment. Negative perception and younger age were associated with worse adherence and more drop out in reviews of

pharmacological treatment. Conversely, the one review of psychological interventions found that younger age was associated with better adherence to treatment, but as this review focused on internet-based interventions only, it may be that age is linked to the method of intervention delivery.

The use of shared decision making was found to be positively associated with treatment dropout, as the two reviews exploring this characteristic both reported that dropout was less likely when a shared decision-making process is used. The potential importance of shared decision making for reducing dropout appears important to consider for services such as IAPT, as although shared decision making wasn't associated with treatment response in psychological interventions, it could improve rates of treatment completion.

Reporting previous treatment for mental health conditions was also linked to reduced rates of dropout in two of the three reviews that included this characteristic. The opposite was identified in reviews that focused on treatment response, where previous treatment was linked to poorer outcomes, and the findings of this meta-review could suggest that individuals who report previous treatment are more likely to stay in treatment, but are less likely to benefit from it. Researchers have suggested that patients may not always successfully recall their previous response to treatment for CMHDs such as depression, and therefore consideration on how this information is collected at assessment is required for it have potential clinical utility (Simon, Rutter, Stewart, Pabiniak, & Wehnes, 2012).

Limitations

One limitation of this meta-review is the level of variation across the included reviews with regard to the populations, the types of primary studies included (controlled or uncontrolled studies) as well as different interventions delivered. Although some reviews had used quantitative methods, such as meta-regression, to explore the predictive nature of patient characteristics, many more used narrative review methods. It is possible that these differences in data aggregation could explain some inconsistencies found between reviews.

The inconsistency regarding certain patient characteristics, could also be linked to the definition of 'response' used by the included reviews, as well as by the primary studies which they included. As discussed in Chapter 1, there are differences between 'recovery', the reduction in symptoms below a clinical cut-off, and 'reliable change', a reduction in symptoms by a certain proportion/absolute value (Boessen, Groenwold, Knol, Grobbee, & Roes, 2012). A large number of included reviews did not explicitly record how 'response' was defined in their primary studies, and therefore this could explain some inconsistencies in the current synthesis if different approaches to defining outcomes were grouped together.

Another limitation of this review is that it has not considered the potential interaction effects between patient characteristics that may impact their ability to predict treatment outcomes.

Although each patient characteristic may inform a decision on possible treatment, incorporating information across multiple characteristics would allow for a more nuanced picture of the patient, and could better inform a DST for use in treatment services. This could not be explored within this meta-review but using a dataset of patients receiving IAPT treatments would provide an opportunity to explore the association of multiple patient characteristics and treatment outcomes, potentially informing personalised treatment decisions when many patient characteristics are available to the decision maker.

Although this thesis is focused on predicting outcome in psychological interventions, this review also included reviews of pharmacological interventions, due to the limited number of studies exploring predictors of psychological interventions alone. Only K=15 of the K=46 included studies were focused on psychological interventions, and only one of these looked at treatment dropout in computerised treatments (Christensen2009). This highlights the lack of high quality reviews into the area, especially patient characteristics associated with treatment dropout. Using a patient dataset from psychological treatment services could provide more information on the role of patient characteristics and risks of dropout to inform a DST for use in IAPT services.

One further limitation of this meta-review is that the identification of reviews, rating of review quality and the data extraction was conducted by the author only, and there was no independent validation of the reliability. This may raise questions about the reliability of the categorisation of reviews and patient characteristics. An independent rater was not used in the current study due to time restrictions, but future work in this area should consider the use of an independent rater, even if just on a subsample of identified reviews (e.g. 10% of papers identified in the literature search), in order to assess the reliability of the categorisations used.

Summary

The meta-review presented in this chapter has identified a number of routinely collected patient characteristics that may be associated with outcomes in the treatment of CMHDs, as well as characteristics which appear to have a more inconsistent association with outcomes in the literature. This information may be of value when determining which patient characteristics should be included as part of routine patient datasets due to their potential to inform treatment outcomes, and some characteristics identified could be used to predict prognosis. For example, higher severity of illness pre-treatment was frequently associated with poorer response to psychological treatment in the included reviews, and this information could be used as part of treatment planning to determine most appropriate treatment for more severe patients. However, the review was limited by the variation in the included systematic reviews with regard to diagnoses, methods of reporting and types of primary studies that were considered, which may have resulted in some inconsistencies with regard

to the association of specific patient characteristics and outcomes. In addition, the review did not consider the potential interactions of co-occurring patient characteristics, which may influence the potential predictive ability of individual patient characteristics. To address these limitations, the next chapter presents analysis performed on a patient dataset collected in two IAPT services, which explored the association between multiple routinely collected patient characteristics and psychological treatment outcomes.

Chapter 4. Patient characteristics and IAPT treatment outcome.

Abstract

The results of the meta-review presented in the previous chapter suggested a number of patient characteristics that may be associated with treatment outcomes, although there were inconsistencies with some findings which may be due to differing populations and methods of reporting treatment outcomes. Supplementing these findings with analyses of patient characteristics and treatment outcomes from IAPT datasets was expected to yield the most clinically useful and relevant information to help personalise treatment in these services. The aim of this chapter was to explore the use of standardised IAPT data from two services (n=10693 patients) in North London to understand the association between routinely available patient characteristics and outcomes from psychological treatments delivered by IAPT. Stepwise multiple regression was used to explore the association between nine patient-specific characteristics available in the IAPT dataset and five outcomes; including recovery, reliable change and improvement, deterioration and dropout from treatment. Analysis was performed on the full sample of patients receiving treatment from the services, but also split by patients who received only low intensity (LI) interventions, high intensity (HI) interventions and patients who were stepped up during treatment. Results indicated that each of the patient characteristics was statistically associated with at least one outcome across sub-analyses, but depression and anxiety severity, functional impairment and welfare status were most predictive of response to treatment, and depression severity, age and welfare status were most predictive of treatment dropout. The patient characteristics considered in this analysis could be used to inform the prediction of likely response to treatment as well as risk of deterioration and dropout. However, the variance explained by some of the final regression models was low, especially for treatment dropout, which may suggest that additional characteristics not available in the dataset may better predict some treatment outcomes.

Introduction

In order to personalise treatment decisions in IAPT services, an understanding of which patient characteristics are associated with treatment outcomes is needed. The systematic review presented in the previous chapter identified a number of patient-specific characteristics that have been consistently associated with treatment outcomes in CMHD, for example comorbid mental health issues (Serretti et al, 2009; Mululo et al, 2012). However, for a number of characteristics the pattern was more inconsistent which suggests that more research is needed in order to understand their association with outcomes, especially in relevant clinical environments. There was also significantly less evidence available for predictors of outcomes following psychological interventions when compared to pharmacological interventions for CMHD. Therefore further analysis was needed to understand the association between patient characteristics and psychological treatment outcomes, and using data from IAPT services was expected to increase the practical and clinical utility of a DST developed as part of this thesis.

One reason for the inconsistencies found with some patient characteristics across reviews may in part be due to the different types of study design (e.g. controlled vs uncontrolled trials) that were included across the reviews, which might have impacted how comparative analyses on predictors. Another possible reason is the different populations, especially with regard to diagnoses, that were included across the reviews of CMHDs. Although IAPT services see patients across the range of CMHDs, formal diagnosis has been generally low, and 48.7% of patients referred to IAPT between 2014 and 2015 were reported to have an invalid or unknown diagnostic code (HCSIC, 2015). As a result, a DST for IAPT that used diagnosis may have little clinical utility in IAPT services, as this information may not be routinely available for half of patients. Instead, focusing on standardised patient characteristics that are routinely collected and able to be self-completed by patients, could result in a DST with more clinical utility in routine services. An analysis of patient characteristics that are routinely collected at all IAPT services would supplement the results of the meta-review and potentially provide more relevant information on predicting outcomes in IAPT services.

As discussed in Chapter 1, IAPT services are the main provider of psychological interventions for CMHDs in England. IAPT services have adopted a stepped care model (see figure 1.1, Chapter 1), and provide low intensity (LI) interventions at Step 2 or high intensity (HI) interventions at Step 3. The main key performance indicator (KPI) for IAPT services nationally is ‘Moving to Recovery’ (NHS Digital, 2016). IAPT defines this as scoring clinical ‘caseness’ on either the Patient Health Questionnaire 9 (PHQ-9; Kroenke, et al., 2001) or the Generalised Anxiety Disorder scale 7 (GAD-7; Spitzer, et al., 2006) at initial contact with the service and then being below caseness at the end of treatment on both symptom measures. Clinical caseness is defined as scoring 10 or higher on the PHQ-9 and/or 8 or above on the GAD-7 at initial assessment, and scoring below caseness on both

measures at the end of treatment is defined as ‘recovery’. Considering the importance of this KPI for evaluating IAPT services, determining patient characteristics associated with recovery will have practical benefits for services nationally in England.

An alternative method of evaluating patient response to treatment is to consider whether patients have reported a clinically meaningful reduction in symptom scores. One of the most commonly used methods of estimating clinically significant reductions in symptoms used in both research and clinical practice is the Reliable Change Index (RCI) (Jacobson & Truax, 1991), a statistic which identifies whether the change in scores is more than could attributed to measurement error. This method can also be applied to increases in patient symptoms, where clinically significant increases in symptom severity may indicate that the patient’s mental health has deteriorated during treatment. This is important as an increase in symptoms during interventions could indicate a harmful treatment effect, and previous research suggests around 5-10% of patients could be classed as deteriorating following psychological interventions (Boisvert & Faust, 2003; Rozental et al., 2017).

The meta-review presented in the previous chapter identified only nine previously conducted reviews that had explored the association between patient characteristics and dropout from treatment, none of which were for face-to-face psychological interventions. Of the n=953522 patients entering IAPT treatment between April 2015 and March 2016, n=416391 did not complete treatment (44%) (NHS Digital, 2016), which is considerably higher than the 30% estimated in routine psychological care (Wells et al., 2013). Identifying patient characteristics associated with increased dropout might inform a more personalised approach to treatment, which could increase engagement and reduce early termination of treatment in IAPT services.

The aim of this chapter was to explore the association between patient characteristics routinely collected as part of the standardised IAPT minimum dataset (MDS) and outcomes following treatment from IAPT services. It was theorised that certain characteristics would have more influence on outcomes for specific intensities of IAPT treatment, and therefore sub-analyses were also conducted for patients receiving different intensities of psychological interventions. The results of this analysis could be used to predict the likelihood of treatment outcomes in IAPT services based on routine assessment data, and could inform the first stages of a personalised treatment approach in IAPT settings.

Methods

Patient sample & Dataset

The sample used for this analysis came from two IAPT services in North London with data on individual patients collected between September 2008 and March 2012. This dataset included every patient referred to the service between this period; a total of n=34741 patients. Initial assessment information (referred to as time 1; 'T1') including data on patient characteristics was available for n=19817 patients. Following the T1 assessment, n=3181 patients either declined treatment from IAPT, were still in treatment or referral according to the patient management system, or were deemed not suitable for treatment from the services. This left n=16636 patients who were assessed and taken into treatment within the dataset.

To calculate the key IAPT outcome, 'moving to recovery', data is needed on both the first and final assessment scores on the PHQ-9 and GAD-7 for patients. Therefore, only patients who had both T1 and T2 (Time 2; endpoint) data available on either symptom measure were included in the analysis. A number of patients accepted into treatment at IAPT services will have only a single treatment session for advice and consultation with a clinician, and these patients would therefore be excluded from analysis as they would provide data for only one time point. In addition, the moving to recovery IAPT outcome requires patients to be scoring above caseness on either symptom measure at T1, and therefore patients who scored below the cut-offs on both measures at T1 were not included in the current analysis (as they did not meet caseness). From the total of n=16636 patient taken into treatment, n=2691 scored below clinical caseness on both symptom measures pre-treatment, and n=3252 patients only received a single treatment session and therefore had no T2 data collected. This resulted in n=10693 patients included in the analysis. A full flow diagram is presented in figure 4.1.

Patient characteristics (Indicator variables)

A number of patient characteristics identified in the systematic review of the literature (Chapter 3) were also present in the IAPT dataset. Although some patient characteristics (e.g. gender) were not consistently associated with outcomes in the meta-review, they were included in this exploratory analysis to confirm whether a pattern of limited association was observed in outcomes following IAPT treatments. The IAPT MDS is a standardised mandatory set of information that includes a range of clinical, demographic and service based variables to be collected at each assessment by all services (HSCIC, 2015). Nine patient characteristics collected within the IAPT MDS were selected for this analysis, and are presented in Table 4.1. Severity scores were taken from the T1 assessment measures on the PHQ-9 and GAD-7. Caseness for phobia, defined as scoring 4 or above on any of the

three IAPT phobia items included in the IAPT dataset (IAPT, 2011) was included as a dichotomous patient characteristic. The level of social and occupational functioning was assessed using the Work and Social Adjustment scale (W&SAS) (Mundt, Marks, Shear, & Greist, 2002) score at T1. Although a range of categories were collected for ethnicity within the dataset, a very large percentage were from white ethnic groups (78%) and it was therefore decided to combine all other ethnicities into a 'non-white' group to reduce the risk of the regression analysis being under powered. The dataset also contains information on whether medication had been prescribed at T1, although there is no data regarding medication adherence. Lastly, information on whether patients were receiving welfare benefits was used as a dichotomous variable, as information on employment status was not available within the dataset provided by the services.

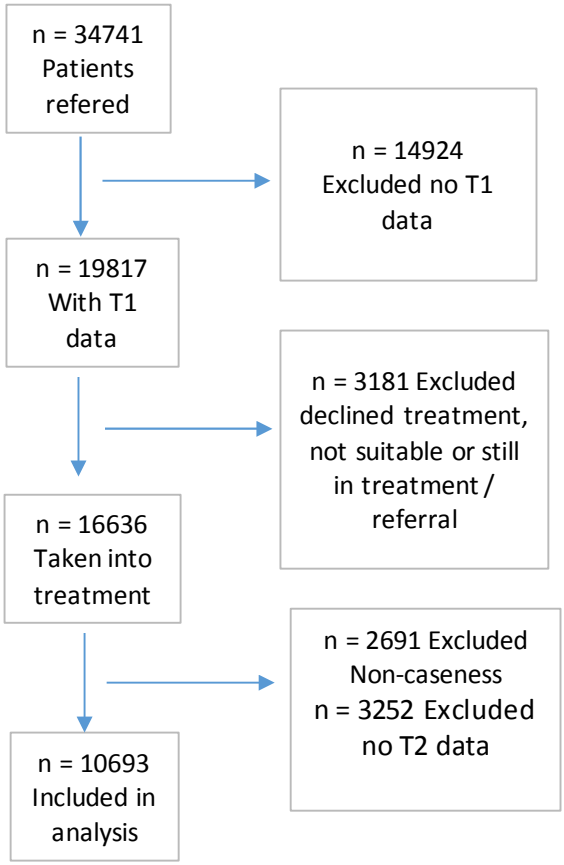

Figure 4.1. Patient flow - analysis of treatment outcomes.

Table 4.1. Included patient characteristics.

Variable	Type of variable	Description
Age at referral	Continuous	Age of patient
Gender	Dichotomous	'Male' or 'female'
Self-rating of depressive symptoms	Continuous	Score on Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001)
Self-rating of anxiety symptoms	Continuous	Score on Generalised Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006)
Level of personal and social functioning	Continuous	Score on Work and Social adjustment Scale (W&SAS; Mundt et al., 2002).
Medication prescription status	Dichotomous	'Prescribed' or 'not prescribed' psychotropic medication at referral.
Welfare status	Dichotomous	'Receiving benefits' or 'not receiving benefits' from UK welfare support.
Ethnic group	Dichotomous	'White' or 'non-white' ethnic group
Phobia self-rating	Dichotomous	'Phobia' or 'non-phobia', classified by a score of 4 or more any one of the three phobia items (IAPT, 2011).

As discussed previously, the IAPT dataset includes a provisional diagnosis variable, or 'problem description' but completion of this data is poor, and missing data is recorded for nearly half of patients (HSCIC, 2015). A DST which included this factor would potentially be unusable for half of patients entering treatment, and therefore diagnosis was not explored in the analysis. A number of patient characteristics identified in the meta-review presented in Chapter 3 which were associated with outcomes were not included in the IAPT dataset (e.g. self-efficacy, and previous treatment), and therefore could not be explored in the analysis. Despite being encouraged by the services, information about shared decision making is not included in the IAPT MDS and therefore was not considered in the analysis presented below.

Outcomes

There were five outcomes of interest in this analysis, all of which were dichotomous in nature, and these are detailed in Table 4.2.

The first outcome explored, 'Moving to Recovery' (Recovery), is the main IAPT national KPI measuring service level performance and was defined as moving from caseness to non-caseness following treatment. Patients scoring above the cut-off point on either measure at T1 needed to score below caseness on both assessment measures (PHQ-9 and GAD-7) at T2 to achieve this outcome. The cut-offs used by IAPT services are a score of 10 on the PHQ-9 and of 8 on the GAD-7 (HSCIS, 2015).

Table 4.2. Treatment outcomes: definitions and criteria.

Name	Definition	Criteria
Recovery	Percentage of patients recovered	Scoring above clinical caseness cut-off at initial assessment on either symptom measure and scoring below the cut-off on both at final assessment following treatment.
Reliable change	Percentage of patient showing reliable change	Reliable decrease in symptoms between first and final assessment on both the PHQ-9 and GAD-7. Calculated using the reliable change index (Jacobsen & Traux, 1991).
Recovery or improvement	Percentage of patients recovered or showing improvement	Either in recovery or showing reliable change on either measure, as defined above, or reliable change in one measure whilst being non-caseness on the second measure.
Deterioration	Percentage of patients in showing clinical deterioration	Deterioration was defined as reliable clinical change in a negative direction (increase in score) between initial and final assessment on either symptom scale. Calculated using the reliable change index (Jacobsen & Traux, 1991)
Dropout	Percentage of patients not completing treatment	Defined as either dropping out of treatment as recorded in the IAPT dataset, or as declining treatment after two or more treatment sessions by the service.

The second outcome of interest was whether patients showed a reliable change on both symptom severity measures following treatment. Reliable change was calculated on both measures for each patient using the reliable change index (RCI) formula (Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991). This method of estimating clinically important change in symptoms is widely used and recommended in outcome research (Bauer, Lambert, & Nielsen, 2004). To calculate this outcome, the standard error of measurement and the error of difference were calculated on both symptom measures using the standard deviation of the current IAPT dataset used for this analysis and the alpha coefficient of reliability presented in the original publications of the assessment tools (Kroenke et al., 2001; Spitzer et al., 2006). Using these values, the reliable change index (RCI) for each patient on both measures was generated, and RCI values over 1.96 indicated reliable change on that measure had been achieved. Calculations of the RCI showed that a change score of 5 or above on the PHQ and a score of 4 or above on the GAD-7 indicated reliable change. Patients were only considered to have met reliable change if they showed reliable change on both measures, rather than just one.

A third outcome was then generated by combining the first and second outcomes to create a 'recovery or improvement' outcome, to incorporate these two different types of treatment effect calculations into one positive outcome, which may be a more valuable indicator of overall performance from the perspective of the service. To achieve this outcome, patients had to either meet the recovery criteria, meet the reliable change criteria (change on both

measures) or display reliable change on one measure whilst being below the clinical cut-off for caseness on the second scale.

In addition to positive outcomes from treatment, the analysis was also interested in potential negative outcomes, for example increases in symptoms following IAPT treatment, and therefore clinical deterioration was also included as an outcome. To meet criteria for clinical deterioration, a patient must show reliable change in a negative direction (5+ on the PHQ-9 and 4+ on the GAD) on either scale. Patients were classed as 'deteriorated' if they met these criteria on just one scale, regardless of whether there was a reduction in symptom scores on the second measure.

The final outcome of interest was dropout from treatment. Data on dropout was obtained from the 'reason for end of contact' record in the service records system. Patients were also coded as dropping out from treatment if they either; 1) were recorded as 'declining care' but had had two or more sessions with a clinician, 2) were recorded as 'not suitable' but had received three or more contacts with the service or 3) were recorded as having dropped out of treatment regardless of number of sessions. It was decided to code patients who were recorded as 'declining' or 'not suitable' following multiple contacts as 'dropout', as these were viewed as coding errors. Patients were expected to have completed an extended assessment by their second contact with the service, and therefore any patient considered as "not suitable", or declining further sessions after the second contact was viewed as a treatment dropout in this analysis.

Analysis was conducted on both the full sample of included patients ('all patient' sample), but also by sub-groups of patients receiving specific intensities of psychological intervention. The three subgroups were patients receiving LI interventions only, patients receiving HI interventions only and patients who were stepped up during treatment. Stepped up patients were defined as individuals who started treatment at Step 2 (LI) but were 'stepped up' to Step 3 (HI) during contact with the services, usually due to lack of response to treatment at Step 2.

Analysis

To explore the association between the nine patient characteristics and the five dichotomous treatment outcomes, multiple logistic regression analysis was performed. Regression analysis is routinely used for this type of research question, and have been used to explore the role of patient characteristics in CMHDs previously (Blom et al., 2007; Gyani, Shafran, Layard, & Clark, 2011).

Binary logistic regression analysis was performed on the dataset to determine which characteristics were significantly associated with each of these five dichotomous outcomes listed Table 4.2 above. All outcomes were coded as 1 for outcome achieved and 0 for not

achieved. As multiple independent variables were to be added to the regression models, a stepwise approach was adopted. There are three typical approaches to consider in stepwise regression (Wang, Wright, Buswell, & Brownlee, 2013):

*Forward selection – starts from an empty model and adds variables one at a time until no more variables can significantly improve the regression model based on criteria defined by the researcher (see below). Variables are added based on pre-selection criteria.

*Backward selection – Model starts with all variables included and variables are then deleted by the least improving variable first and so on until no more deleted variables can significantly improve the model.

*Bidirectional elimination – Uses the forward selection procedure but at each step all variables in the model at that step are investigated to see if any need to be deleted based on pre-selected criteria.

For this analysis, a Bidirectional elimination method was selected with variables added one by one, and all variables in the model at that step checked to make sure they still had a significant association with the outcome, when controlling for all other characteristics in the model. The criteria for including a variable in the model were a significant p-value for the odds ratio ($p<0.05$) and an increase in the R^2 value following the introduction on the variable, indicating an increase in variance explained by the regression model. Analysis was performed using STATA 12 (StataCorp LP, 2011).

The analysis for each outcome is presented below, with the Odds Ratio (OR), 95% confidence of each patient characteristic and the R^2 for the model at that step displayed. The order that the variables were added to the regression models was informed by the results of the meta-review presented in Chapter 3. Patient characteristics with more evidence for association from the full review were added first, and those with more inconsistent or limited findings were added afterwards. The order of patient characteristics added to the analyses on recovery, reliable change, recovery or improvement and deterioration outcomes are displayed in Table 4.3 below. As the patient characteristics associated with treatment dropout identified in the meta-review differed from those associated with treatment response, a different order of variable addition was employed for these analyses, and this is displayed in Table 4.4.

Table 4.3. Order of patient variables added to 'response' analyses.

Variable name	How measured
Severity	PHQ-9 T1 score
Severity	GAD-7 T1 score
Comorbidity	Phobia caseness at T1
Functioning	W&SAS T1 score
Ethnicity	White or non-white ethnic group
Welfare status	On benefits at T1
Age	Age at referral
Gender	Gender recorded
Medication use	Prescribed medication at T1

Table 4.4. Order of patient variables added to 'dropout' analyses.

Variable name	How measured
Age	Age at referral
Severity	PHQ-9 T1 score
Severity	GAD-7 T1 score
Comorbidity	Phobia caseness at T1
Gender	Gender recorded
Ethnicity	White or non-white ethnic group
Functioning	W&SAS T1 score
Welfare status	On benefits at T1
Medication use	Prescribed medication at T1

Results

All patients entering treatment

Table 4.5 displays the percentage of patients achieving each of the five outcomes for 'all patients' entering treatment at the services (n=10693). Just over 40% of the sample achieved the recovery and reliable change outcomes following treatment, and nearly 48% of patients achieved the recovery or improvement outcome. The recovery rate for the services was similar to the median rate reported by the one-year IAPT report (41%; Gyani et al., 2011) but lower than more recent national evaluations of IAPT recovery figures (46%; NHS Digital, 2016).

Deterioration on at least one of the symptom severity measures was recorded for just under 9% of the full sample, which falls within the 5-10% suggested by previous researchers (Crawford et al., 2016; Rozental et al., 2016. Dropout was indicated in over 33% of patients, which is similar to levels reported to other evaluations of routine psychological treatment (Wells et al., 2013) and lower than the most recent national IAPT report (NHS Digital, 2016). Although patients who dropped out did not complete treatment, the patient's last score on the symptom scales was available and used as T2 data for the analysis.

Table 4.5. All patients: Outcomes following treatment.

Outcome	No	Yes	%
Recovery	6,410	4,283	40.05%
Reliable change	6,352	4,341	40.60%
Recovery or improvement	5,568	5,125	47.93%
Deterioration	9,762	931	8.71%
Dropout	7,114	3,579	33.47%

Recovery – 'All Patients'

The results of the stepwise logistic regression for recovery in the 'all patients' sample is displayed in Table 4.6. The patient characteristics were added to the regression model in the order displayed in Table 4.3 above. At the first step of the regression model, the PHQ-9 score at T1 was found to have a significant odds ratio of 0.879 (p<0.001), suggesting that for each increase in score on the PHQ-9 score at T1 (initial assessment), the odds of moving to recovery following treatment drop by 0.879. This indicates that higher levels of depression severity at assessment were significantly associated with a lower likelihood of recovery following treatment. The variance explained by PHQ-9 score at T1 alone was 8.38%, as indicated by the R^2 value.

Table 4.6. Stepwise regression model: recovery (all patients)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.879	0.873 to 0.886	0.000	8.38%
Model Step 2				
PHQ-9 T1	0.904	0.896 to 0.912	0.000	9.53%
GAD-7 T1	0.931	0.921 to 0.941	0.000	
Model Step 3				
PHQ-9 T1	0.907	0.899 to 0.915	0.000	9.79%
GAD-7 T1	0.937	0.926 to 0.947	0.000	
Phobia caseness	0.773	0.709 to 0.844	0.000	
Model Step 4				
PHQ-9 T1	0.924	0.915 to 0.933	0.000	10.27%
GAD-7 T1	0.938	0.927 to 0.948	0.000	
Phobia caseness	0.832	0.761 to 0.91	0.000	
Functioning	0.976	0.971 to 0.982	0.000	
Model Step 5				
PHQ-9 T1	0.93	0.92 to 0.94	0.000	11.08%
GAD-7 T1	0.936	0.925 to 0.948	0.000	
Phobia caseness	0.914	0.83 to 1.007	0.07	
Functioning	0.981	0.975 to 0.988	0.000	
Welfare status	0.523	0.469 to 0.584	0.000	
Model Step 6				
PHQ-9 T1	0.929	0.919 to 0.939	0.000	11.12%
GAD-7 T1	0.935	0.924 to 0.946	0.000	
Functioning	0.981	0.975 to 0.987	0.000	
Welfare status	0.515	0.461 to 0.574	0.000	

At the next step, the GAD-7 score at T1 was added to the model, which increased the R^2 to 9.53%, and had a significant odds ratio of 0.931 (p<0.001), suggesting higher scores on the GAD-7 at T1 reduced the odds of recovery. By adding GAD-7 score to the regression model, the odds ratio for PHQ-9 score was adjusted slightly to 0.904, although the PHQ-9 score remained significant (p<0.001). Phobia caseness and the level of functioning were then added to the model in the next two steps, and both showed significant odds ratios as well as increasing the R^2 to 10.27%. Scoring caseness for phobia at T1 and having lower levels of social and occupational functioning (indicated by a higher scores at T1 on the W&SAS) significantly reduced the odds of recovery following treatment. Ethnicity was added to the model, but this characteristic was not significantly associated with the recovery (p>0.05). At

Step 5 of the regression model, welfare status was added and this factor had a significant association with recovery, as receiving benefits was found to reduce the odds of recovery as well as increasing the R^2 to 11.08%. However, by adding welfare status to the model, the odds ratio for phobia caseness became non-significant (p=0.07) and this factor was therefore dropped at Step 6. Age, gender and medication prescription status were also included in the final stages, but none of these patient characteristics were found to significantly improve the model and therefore the final regression model included PHQ-9, GAD-7 and W&SAS scores at T1 as well as welfare status, and this model explained 11.12% of the variation in recovery.

Reliable change – 'All Patients'

The stepwise regression model for reliable change in 'all patients' is presented in Table 4.7. In the first step, the PHQ-9 score at T1 was found to be significantly associated with achieving reliable change, indicated by an odds ratio of 1.057 (p<0.001). An odds ratio above 1 indicates the odds of achieving the outcome increase with higher PHQ-9 T1 scores, suggesting that higher depression scores at assessment increased the likelihood of reliable change following treatment. In Step 2 of the model, GAD-7 at T1 was added and again had a similar significant association to the PHQ-9 score, as higher T1 GAD-7 score was found to increase odds of achieving this outcome. These findings differ to those for recovery presented in Table 4.6 where higher scores decreased the odds of achieving the recovery outcome following treatment.

In the next steps, caseness for phobia at T1 and level of functioning were added to the model and both had significant associations with the outcome. Lower levels of functioning and scoring caseness for phobia reduced the odds of reliable change following treatment. The next step involved the addition of ethnicity group to the model, and there was a significant association with being from a non-white ethnic group and lower odds of achieving reliable change following treatment in the services, although just 3.21% of variance was explained by this model. The addition of welfare status increased the R^2 to 4.96% and suggested that being on benefits reduced the odds of reliable change. However, the addition of welfare status to the model resulted in phobia caseness becoming non-significant and therefore phobia was dropped at Step 7. In the final step, medication prescription status at T1 was found to be significantly associated with achieving reliable change, as individuals prescribed psychotropic medication were less likely to reliably change on both symptom scales following treatment. The final multivariate regression model accounted for 4.93% of the variance.

Table 4.7. Stepwise regression model: reliable change (all patients)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	1.057	1.049 to 1.064	0.000	1.70%
Model Step 2				
PHQ-9 T1	1.036	1.027 to 1.044	0.000	2.26%
GAD-7 T1	1.049	1.038 to 1.06	0.000	
Model Step 3				
PHQ-9 T1	1.041	1.032 to 1.05	0.000	2.61%
GAD-7 T1	1.056	1.044 to 1.068	0.000	
Phobia caseness	0.763	0.7 to 0.831	0.000	
Model Step 4				
PHQ-9 T1	1.06	1.05 to 1.07	0.000	3.12%
GAD-7 T1	1.059	1.046 to 1.07	0.000	
Phobia caseness	0.818	0.749 to 0.892	0.000	
Functioning	0.978	0.973 to 0.983	0.000	
Model Step 5				
PHQ-9 T1	1.061	1.05 to 1.072	0.000	3.21%
GAD-7 T1	1.058	1.046 to 1.071	0.000	
Phobia caseness	0.828	0.754 to 0.91	0.000	
Functioning	0.979	0.973 to 0.984	0.000	
Ethnicity Group	0.843	0.759 to 0.936	0.001	
Model Step 6				
PHQ-9 T1	1.074	1.062 to 1.086	0	4.96%
GAD-7 T1	1.06	1.046 to 1.074	0	
Phobia caseness	0.926	0.836 to 1.024	0.135	
Functioning	0.984	0.978 to 0.99	0	
Ethnicity Group	0.889	0.793 to 0.996	0.042	
Welfare status	0.463	0.414 to 0.518	0	
Model Step 7				
PHQ-9 T1	1.073	1.061 to 1.085	0	4.92%
GAD-7 T1	1.059	1.045 to 1.072	0	
Functioning	0.983	0.977 to 0.989	0	
Ethnicity Group	0.879	0.785 to 0.984	0.026	
Welfare status	0.46	0.412 to 0.514	0	

Model Step 8				
PHQ-9 T1	1.076	1.064 to 1.089	0	4.93%
GAD-7 T1	1.057	1.043 to 1.071	0	
Functioning	0.985	0.979 to 0.991	0	
Ethnicity Group	0.878	0.779 to 0.989	0.032	
Welfare status	0.473	0.421 to 0.531	0	
Medication prescribed	0.871	0.786 to 0.966	0.009	

Comparing the final regression model for reliable change (Table 4.7) to the model for recovery (Table 4.6) shows an important difference in the role of severity with these two outcomes. Higher severity was found to decrease the odds of recovery following treatment, but conversely increased the odds of a patient achieving reliable change. This may be explained by the way in which the outcomes are defined and achieved. For recovery, individuals needed to move from above to below a clinical cut off, and arguably being closer to the cut off at T1 should make it easier to move below (as less change is needed). In practice, patients with lower severity scores have a head start on patients scoring higher at T1 when both are trying to reach the same goal (symptom scores under the threshold). When considering reliable change, the goal of this outcome is to reduce symptom scores by a certain absolute amount of scale points (5 points or more on the PHQ-9, 4 or more on the GAD-7) and therefore patients scoring higher have more available points in which they can decrease. This phenomenon is referred to as regression to the mean, where scoring closer to the extreme increases the likelihood of the scoring closer to the mean on a second measurement (Stigler, 1997). Although severity had a different association with the outcomes, the level of functioning, phobia caseness and welfare status all had the same association with both outcomes, suggesting that being phobic, on benefits and lower levels of functioning decrease a patient's odds of a positive outcome (both recovery and reliable change) following treatment at the IAPT services.

Recovery or improvement – 'All Patients'

Following the individual outcomes of recovery and reliable change, the results of 'recovery or improvement' are presented in Table 4.8. The results of this analysis were very similar to those for recovery (Table 4.6). Higher PHQ-9 and GAD-7 scores at T1 were significantly associated with decreased odds of achieving a positive outcome following treatment. Phobia caseness and lower levels of social and occupational functioning were then added to the model and both characteristics reduced the odds of a positive outcome, with 7.53% of the variance explained at this step. Welfare status was then added as the last step, and being on benefits was found to significantly reduce the odds of positive outcome when controlling for the other variables in the model. No other patient characteristics improved the model and

the final variance explained was 8.48%.The variance explained was therefore lower than for the regression model of recovery (Table 4.6), and suggests that the combined outcome is more difficult to predict from the included patient characteristics, which is most likely due to the opposing influence of initial severity on the recovery and reliable change outcomes.

Table 4.8. Stepwise regression model: recovery or improvement (all patients)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.896	0.889 to 0.903	0	6.36%
Model Step 2				
PHQ-9 T1	0.908	0.9 to 0.915	0	6.57%
GAD-7 T1	0.97	0.96 to 0.98	0	
Model Step 3				
PHQ-9 T1	0.911	0.904 to 0.92	0	6.86%
GAD-7 T1	0.975	0.965 to 0.986	0	
Phobia caseness	0.757	0.696 to 0.824	0	
Model Step 4				
PHQ-9 T1	0.927	0.918 to 0.936	0	7.35%
GAD-7 T1	0.977	0.966 to 0.988	0	
Phobia caseness	0.813	0.745 to 0.887	0	
Functioning	0.978	0.972 to 0.983	0	
Model Step 5				
PHQ-9 T1	0.935	0.925 to 0.945	0	8.48%
GAD-7 T1	0.976	0.964 to 0.987	0	
Phobia caseness	0.892	0.811 to 0.98	0.018	
Functioning	0.982	0.976 to 0.988	0	
Welfare status	0.505	0.456 to 0.56	0	

Deterioration – 'All Patients'

The regression model for clinical deterioration, an increase in one or both symptom scale score(s) over the reliable change threshold following treatment, is presented in Table 4.9. In the first two steps of the model, higher PHQ-9 and GAD-7 scores were found to be significantly associated with lower odds of deteriorating following treatment. This suggests that lower severity at T1 increased the likelihood of deteriorating by T2. The addition of phobia caseness and then the level of functioning increased the variance explained to 4.62%, and the model suggested that being phobic and having lower levels of functioning pre-treatment increased the odds of clinical deterioration.

Table 4.9. Stepwise regression model: deterioration (all patients)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.94	0.929 to 0.951	0	1.66%
Model Step 2				
PHQ-9 T1	0.976	0.963 to 0.99	0.001	3.73%
GAD-7 T1	0.903	0.888 to 0.919	0	
Model Step 3				
PHQ-9 T1	0.972	0.958 to 0.986	0	4.21%
GAD-7 T1	0.895	0.879 to 0.911	0	
Phobia caseness	1.54	1.329 to 1.784	0	
Model Step 4				
PHQ-9 T1	0.951	0.935 to 0.967	0	4.62%
GAD-7 T1	0.894	0.878 to 0.91	0	
Phobia caseness	1.39	1.194 to 1.617	0	
Functioning	1.028	1.018 to 1.038	0	
Model Step 5				
PHQ-9 T1	0.953	0.936 to 0.97	0	5.36%
GAD-7 T1	0.883	0.867 to 0.901	0	
Phobia caseness	1.42	1.207 to 1.67	0	
Functioning	1.026	1.016 to 1.037	0	
Ethnicity group	1.306	1.095 to 1.556	0.003	
Model Step 6				
PHQ-9 T1	0.941	0.923 to 0.959	0	6.47%
GAD-7 T1	0.889	0.871 to 0.908	0	
Phobia caseness	1.249	1.047 to 1.491	0.013	
Functioning	1.021	1.009 to 1.032	0	
Ethnicity group	1.196	0.987 to 1.45	0.068	
Welfare status	2.149	1.799 to 2.567	0	
Model Step 7				
PHQ-9 T1	0.938	0.921 to 0.955	0	5.66%
GAD-7 T1	0.902	0.885 to 0.912	0	
Phobia caseness	1.227	1.04 to 1.447	0.015	
Functioning	1.022	1.011 to 1.032	0	
Welfare status	2.126	1.8 to 2.511	0	

In the next step of the model, being in a non-white ethnic group was found to significantly increase the odds of deterioration. Being on benefits was then associated with increased odds of clinical deterioration following treatment, but the addition of this patient characteristic to the regression model adjusted the odds ratio of ethnicity group so that it was no longer statistically significant (*p*=0.068) and this variable was dropped at Step 7 of the model. No other variables improved the model fit and the final R^2 was 5.66%.

Treatment dropout – 'All Patients'

The order of patients characteristics added to the stepwise regression model for treatment dropout is presented in Table 4.4, and differed from the order used in the analyses of recovery, reliable change and deteriorated presented above. The results of the multivariate logistic regression for treatment dropout in the 'all patient' sample is presented in Table 4.10.

The first patient characteristic added to the model was age of patient, which was found to be significantly associated with dropout from treatment, as older age was associated with decreased odds of treatment dropout. In the next model step, a higher PHQ-9 score at T1 was found to increase the odds of treatment dropout. Welfare status was then added to regression model, and the results indicated that patients who were receiving benefits had a significantly increased risk of dropout. No other items were found to significantly improve the model. The final R^2 value was 0.0192, suggesting less than 2% of the variance in dropout was explained by the available patient characteristics.

Table 4.10. Stepwise regression model: dropout (all patients)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
Age	0.984	0.98 to 0.987	0	0.76%
Model Step 2				
Age	0.982	0.979 to 0.985	0	1.75%
PHQ-9 T1	1.043	1.036 to 1.051	0	
Model Step 3				
Age	0.981	0.978 to 0.985	0	1.92%
PHQ-9 T1	1.036	1.027 to 1.044	0	
Welfare status	1.368	1.239 to 1.512	0	

Summary – 'All Patients'

The regression analyses of 'all patients' entering treatment at the IAPT services identified a number of patient characteristics that were significantly associated with treatment outcomes in IAPT. However, the variance explained by these models was limited, with only the final model for the recovery outcome explaining over 10% of the variance. The regression model for treatment dropout found that only three characteristics were significantly associated with dropping out of treatment, and the variance explained by this model was just 1.92%.

The analysis of characteristics associated with recovery supported findings from the systematic review presented in Chapter 3 as lower severity, being in employment and higher levels of functioning were all associated with better outcomes. Although ethnicity group and medication status were not associated with the recovery outcome, they were both found to be significantly associated with the reliable change outcome, as being from a non-white ethnic group and prescribed medication reduced the likelihood of achieving reliable change following treatment. Potentially these two characteristics may indicate greater complexity and/or social adversity which may result in reduced odds of a positive outcome from treatment. When the recovery or improvement outcome was explored, the associated patient characteristics matched those for the recovery alone outcome, but with a reduced variation explained by the final model. It is likely that the inclusion of 'improvement' to the model, which would include reliable change on one measure for some individuals, is contributing more variability in the predictive ability of the patient characteristics, hence the reduced variation explained.

The results of the regression analysis of clinical deterioration suggested that lower severity at T1 is associated with a higher likelihood of symptom scores increasing by the end of contact with the services. Although this could appear concerning, it may also be explained by regression to the mean, as individuals who are already scoring higher on the measures at T1 have less available points on which to increase compared to individuals scoring at the lower range of the symptoms measures at assessment. Alternatively, lower severity may have been associated with an increased likelihood of deterioration as some patients may have attended the services before the symptoms had reached a peak. For example, some patients with phobia may display an increase in anxiety symptoms as they confront their phobia during initial treatment. The other patient characteristics associated with deterioration were the same as those identified in the model predicting recovery, but with an opposite direction of effort (as expected): the presence of phobia, higher levels of work and social functioning impairment and receiving welfare benefits all increased the likelihood of increased symptoms following treatment. These characteristics could all be linked to more complex presentations in the services, which if combined with stressful life events and limited social support could increase the risks of a worsening of clinical symptoms (Steger & Kashdan, 2009).

Increased odds of treatment dropout were associated with lower age, higher depression scores at T1 and being on welfare benefits at assessment, although this model explained very little variance in the outcome (1.92%). Younger age was also associated with less treatment dropout in the meta-review, suggesting the findings of this analysis support findings identified in the previous literature. Severity of depression was linked to treatment dropout, but anxiety scores were not, which may explain the inconsistent role of severity across the included systematic reviews identified in Chapter 3. Employment status was not included in any of the included systematic reviews of treatment dropout, but the findings of this regression analysis suggest they may be important in predicting whether a patient will complete treatment.

The findings of the regression analyses performed identified a number of patient characteristics that were associated with treatment outcomes in patients receiving IAPT treatment. However, the variance explained by the models was limited for predicting reliable change, deterioration and dropout. The variance explained by the regression model for recovery was higher (over 11%) and was above the level that would be considered of clinical importance to researchers (Uher et al., 2012). It was possible that the inclusion of patients attending different intensities of psychological treatment reduced the variance explain by some of the regression models, as characteristics may have a different association with outcomes depending on the treatments received. Therefore the analysis was replicated using sub-samples of patients within the dataset who received LI or HI interventions, as well as patients who were stepped up during contact with the services.

Sub-analyses by intensity of intervention

The next stage of analysis was to explore the role of patient characteristics associated with outcome following different IAPT-delivered interventions. However, the nature of the IAPT services meant that one patient may have received a number of different interventions within an episode of care. This may be because they were ‘stepped up’ during their episode of care and therefore received both a self-help low intensity intervention and a therapist-delivered treatment high intensity intervention. In addition, some patients may have received a change in treatment modality during treatment, for example receiving both CBT and counselling during an episode of care. Appendix C presents the count of interventions received by the n=10693 included patients. Self-help (other) (51.45%) and CBT (39.67%) were the interventions most commonly received by any patient. Specific modalities of treatment such as IPT and Couples Therapy were received relatively infrequently in the dataset and received by just n=63 (0.59%) and n=37 (0.35%) of the n=10693 included patients. As a result, it was decided to group patients into the intensity of intervention they received rather than focus on individual treatment modalities, and the sample was split by intensity of intervention received, and regression analysis was performed by these sub-groups.

Individuals who were stepped up were considered as a separate group as they would have received both intensities of treatment.

Table 4.11 shows the number of patients who received interventions at each step of the stepped care model in the dataset. Of the n=10693 patients included in the full sample analysis, only n=38 patients were recorded as receiving treatment at either Step 1 or Step 4 and these patients were not considered for analysis, as these interventions are not typically delivered in IAPT services. Some patients were stepped down from Step 3 to Step 2 during the course of treatment but as this group consisted of less than 2% of the included sample, they were not explored in this analysis. N=568 patients did not have their intervention type recorded and therefore could not be included, but there was information on n=9894 patients who received treatment at Step 2 (LI), Step 3 (HI) or were 'stepped up' during contact with the services. This made up 92.52% of the initial included sample ('all patients') and regression analyses for the five outcomes were performed on these three intervention groups separately.

Table 4.11. Number of patients by intensity of intervention.

Step of Intervention	Freq.	%
Step 1 or Step 4	38	0.36%
Stepped Down	193	1.80%
Low Intensity	5,271	49.29%
High Intensity	3,063	28.64%
Stepped Up	1,560	14.59%
Not known/recorded	568	5.31%
Total	10,693	

A comparison of the means and proportions of the nine patient characteristics between patients receiving either LI, HI or 'stepped up' patients is displayed in Table 4.12. One-way ANOVA and chi^2 statistics indicate that the means and proportions were significantly different between intervention groups. Age was slightly higher for HI patients than LI or stepped up patients, which may suggest more chronic or re-occurring issues requiring higher intensity treatment. Pre-treatment PHQ and GAD scores were higher in the stepped up group when compared to HI and LI patients, whereas functional impairment as scored on the W&SAS was highest in patients who received only HI treatment.

Table 4.12. Descriptive statistics by intensity of intervention.

	LI	**HI**	**Stepped Up**		
	Mean (SD)	Mean (SD)	Mean (SD)	*F*	*p-value*
Age at referral	37.4 (12.8)	39.1 (13.4)	37.0 (12.8)	*22.64*	*<0.000*
Pre-treatment PHQ-9	15.0 (5.5)	16.1 (6.0)	16.5 (5.6)	*60.04*	*<0.000*
Pre-treatment GAD-7	13.6 (4.3)	14.2 (4.7)	14.7 (4.3)	*41.68*	*<0.000*
Pre-treatment W&SAS	18.1 (8.8)	21.3 (9.5)	20.7 (9.0)	*127.26*	*<0.000*
	N (%)	N (%)	N (%)	*Chi2*	*p-value*
Gender - Female	3508 (68%)	1905 (63%)	1012 (65%)	*16.25*	*<0.000*
Phobia caseness - Yes	2524 (50%)	1894 (63%)	924 (60%)	*131.76*	*<0.000*
Receiving welfare benefits - Yes	1104 (25%)	1024 (38%)	394 (28%)	*150.96*	*<0.000*
Ethnic group - Non-white	1032 (23%)	629 (23%)	286 (20%)	*5.23*	*0.073*
Medication prescribed - Yes	1717 (34%)	1337 (55%)	572 (38%)	*302.83*	*<0.000*

Although the proportion of people from non-white ethnic background was not significantly different between intervention groups, the other four categorical variables showed significant differences between groups. A higher proportion of patients receiving HI treatments were male, receiving benefits, met caseness for phobia and were prescribed medication compared to patients receiving LI only. This would be expected if the more complex nature of the presentation contributed to the decision to select HI intervention instead of LI. Stepped up patients had a very similar proportions of female patients, patients prescribed medication and patients receiving benefits compared to LI, but stepped up patients had a higher proportion of individuals scoring caseness for phobia.

Patients receiving Low Intensity treatments

Patients receiving LI interventions were seen for shorter periods of time than those receiving HI interventions, and with an expectation that less intensive interventions will be satisfactory to provide a positive outcome (IAPT, 2011). Table 4.13 displays the percentage of patients achieving each of the patient outcomes in this subgroup. Overall, the percentages are not dissimilar from 'all patients' entering treatment (Table 4.5), although more patients were moving to recovery following LI interventions (41.4% vs 40.05%) and the percentage of patients deteriorating is slightly lower (7.74% vs 8.71%).

Table 4.13. Low intensity: Outcomes following treatment.

Outcome	No	Yes	%
Recovery	3,089	2,182	41.40%
Reliable change	3,126	2,145	40.69%
Recovery or improvement	2,675	2,596	49.25%
Deterioration	4,863	408	7.74%
Dropout	3,510	1,761	33.41%

Stepwise logistic regression was performed using the same nine patient variables, for the five treatment outcomes explored in the analysis in the 'all patient' sample. For the stepwise analysis, the patient variables were entered in the order presented in Table 4.3 for the recovery, reliable change, recovery or improvement and deterioration analyses, and the order in Table 4.4 was used in the analyses of treatment dropout.

Recovery - Low Intensity

The stepwise regression model for moving to recovery is presented in Table 4.14. The significant associations and model steps were identical to those in the stepwise regression model for recovery in 'all patients'. The similarities in findings between these comparisons may be influenced by the large number of patients in the LI group, who were therefore also present in the 'all patients' sample (5271 of 10693; 49%). The ORs in the final step of the model were very similar between the models for 'all patients' and LI only. The variance explained by the LI recovery model was 10.6%, which was lower than for the 'all patient' sample, using the same patient characteristics.

Table 4.14. Stepwise regression model: recovery (Low intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.88	0.87 to 0.89	0	7.69%
Model Step 2				
PHQ-9 T1	0.905	0.894 to 0.917	0	8.79%
GAD-7 T1	0.931	0.916 to 0.946	0	
Model Step 3				
PHQ-9 T1	0.908	0.896 to 0.92	0	8.99%
GAD-7 T1	0.934	0.919 to 0.95	0	
Phobia caseness	0.805	0.711 to 0.91	0.001	
Model Step 4				
PHQ-9 T1	0.927	0.914 to 0.941	0	9.50%
GAD-7 T1	0.934	0.919 to 0.95	0	
Phobia caseness	0.862	0.76 to 0.978	0.021	
Functioning	0.973	0.965 to 0.981	0	
Model Step 5				
PHQ-9 T1	0.935	0.92 to 0.95	0	10.56%
GAD-7 T1	0.934	0.917 to 0.95	0	
Phobia caseness	0.919	0.802 to 1.055	0.23	
Functioning	0.976	0.968 to 0.985	0	
Welfare status	0.47	0.397 to 0.556	0	
Model Step 6				
PHQ-9 T1	0.934	0.92 to 0.949	0	10.60%
GAD-7 T1	0.932	0.916 to 0.949	0	
Functioning	0.976	0.967 to 0.984	0	
Welfare status	0.463	0.392 to 0.548	0	

Reliable change - Low Intensity

The stepwise regression model for reliable change in patients attending LI interventions is presented in Table 4.15. The results of this analysis were again very similar to those for reliable change in the ‘all patients’ patients, with the same patient characteristics significantly associated with the outcome.

Table 4.15. Stepwise regression model: reliable change (Low intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	1.067	1.056 to 1.078	0	2.19%
Model Step 2				
PHQ-9 T1	1.046	1.034 to 1.059	0	2.73%
GAD-7 T1	1.049	1.033 to 1.066	0	
Model Step 3				
PHQ-9 T1	1.051	1.038 to 1.065	0	2.99%
GAD-7 T1	1.054	1.037 to 1.071	0	
Phobia caseness	0.803	0.712 to 0.906	0	
Model Step 4				
PHQ-9 T1	1.073	1.058 to 1.088	0	3.62%
GAD-7 T1	1.055	1.038 to 1.072	0	
Phobia caseness	0.86	0.76 to 0.973	0.017	
Functioning	0.975	0.967 to 0.983	0	
Model Step 5				
PHQ-9 T1	1.075	1.059 to 1.092	0	3.86%
GAD-7 T1	1.055	1.037 to 1.074	0	
Phobia caseness	0.866	0.758 to 0.989	0.034	
Functioning	0.975	0.967 to 0.984	0	
Ethnicity Group	0.797	0.633 to 0.859	0	
Model Step 6				
PHQ-9 T1	1.089	1.071 to 1.108	0	5.37%
GAD-7 T1	1.057	1.037 to 1.077	0	
Phobia caseness	0.926	0.8 to 1.07	0.297	
Functioning	0.98	0.971 to 0.989	0	
Ethnicity Group	0.764	0.646 to 0.902	0.002	
Welfare status	0.499	0.421 to 0.592	0	
Model Step 7				
PHQ-9 T1	1.089	1.071 to 1.108	0	5.39%
GAD-7 T1	1.056	1.037 to 1.076	0	
Functioning	0.979	0.97 to 0.988	0	
Ethnicity Group	0.755	0.639 to 0.891	0.001	
Welfare status	0.495	0.418 to 0.586	0	

Model Step 8				
PHQ-9 T1	1.093	1.074 to 1.112	0	5.55%
GAD-7 T1	1.055	1.035 to 1.075	0	
Functioning	0.98	0.97 to 0.989	0	
Ethnicity Group	0.746	0.629 to 0.884	0.001	
Welfare status	0.504	0.424 to 0.6	0	
Medication prescribed	0.793	0.681 to 0.923	0.003	

The initial PHQ-9, then GAD-7 score were significantly associated with achieving reliable change, as higher scores on both measures increased the odds of reliable change. Phobia caseness and lower social functioning (as measured on the W&SAS) were significantly associated with reduced odds of reliable change. The next steps of the model indicated that being from a non-white ethnic group significantly reduced the odds of reliable change, and that receiving welfare benefits was found to further reduce the likelihood of achieving reliable change. Adding these variables to the model resulted in the phobia caseness variable becoming non-significant, and therefore it was removed at Step 7. Finally, at Step 8, having medication prescribed at the time of the initial assessment was significantly associated with a reduced likelihood of achieving reliable change following treatment. The overall variance explained by the included patient characteristics was 5.55% which is slightly higher than the variance explained by the same model in the 'all patients' sample ($R^2 = 0.493$).

Recovery or Improvement - Low Intensity

The stepwise regression analysis exploring characteristics associated with recovery or improvement is presented in Table 4.16. PHQ-9, GAD-7, Phobia caseness and the level of functioning were all found to be associated with the positive treatment outcome, with lower severity, not meeting caseness for phobia and higher levels of functioning at assessment associated with increased odds of achieving the outcome. Welfare status was then added to the model and was significantly associated with the outcome (being on benefits reduced the odds of achieving the outcome), but this resulted in phobia caseness becoming non-significant and therefore this characteristic was removed at the following step (Step 6). No other patient characteristics were significantly associated with the outcome. The final model included PHQ-9, GAD-7, functioning, and welfare status, and accounted for 7.65% of the variance.

The final regression model for recovery or improvement in LI cases was similar to the regression model for the same outcome in the 'all patients' sample, with the exception of phobia caseness which remained significantly associated with the outcome in the final model for the 'all patients' sample. This suggests that phobia caseness may have less impact on this outcome in LI patients, compared to patients attending other interventions.

Table 4.16. Stepwise regression model: recovery or improvement (Low intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.898	0.888 to 0.907	0	5.74%
Model Step 2				
PHQ-9 T1	0.908	0.896 to 0.919	0	5.86%
GAD-7 T1	0.974	0.96 to 0.99	0.001	
Model Step 3				
PHQ-9 T1	0.911	0.899 to 0.923	0	6.11%
GAD-7 T1	0.979	0.963 to 0.995	0.009	
Phobia caseness	0.772	0.685 to 0.87	0	
Model Step 4				
PHQ-9 T1	0.929	0.915 to 0.942	0	6.61%
GAD-7 T1	0.979	0.963 to 0.995	0.01	
Phobia caseness	0.823	0.728 to 0.931	0.002	
Functioning	0.976	0.968 to 0.983	0	
Model Step 5				
PHQ-9 T1	0.938	0.924 to 0.953	0	7.73%
GAD-7 T1	0.978	0.961 to 0.995	0.012	
Phobia caseness	0.883	0.772 to 1.009	0.068	
Functioning	0.978	0.97 to 0.986	0	
Welfare status	0.49	0.419 to 0.572	0	
Model Step 6				
PHQ-9 T1	0.937	0.923 to 0.952	0	7.73%
GAD-7 T1	0.976	0.959 to 0.993	0.006	
Functioning	0.977	0.968 to 0.985	0	
Welfare status	0.478	0.409 to 0.558	0	

Deterioration - Low Intensity

The analysis of deterioration in LI patients is presented in Table 4.17. The PHQ-9 score at T1 was significantly associated in the first step but the addition of GAD-7 adjusted the OR for PHQ-9 so that it was no longer statistically significant (p=0.074) and was removed by Step 3 of the regression model. Phobia caseness, then lower levels of functioning were found to be significantly associated with increased deterioration and added to the regression model. Finally, being from of a non-white ethnic group was found increase the odds of deterioration following treatment. No other patient characteristics were significantly

associated with the outcome when controlling for the characteristics already included in the model, and the variance explained was 5.55%.

The results of this analysis were different from those presented in the 'all patients' analysis (Table 4.9). Initial severity of depression scored on the PHQ-9 was not significantly associated with deterioration in the final LI model, whereas it was significantly associated with the outcome in the 'all patients' analysis. Welfare status was also significantly associated with deterioration for all patients, but it was not associated with deterioration in patients who received only LI treatments, as it did not increase the variance explained by the model. Ethnicity group was found to be significantly associated with deterioration in the final LI regression model, but it was not in the 'all patient' model as it was no longer significant in the final step. These results suggest there may be a slightly different role of patient characteristics that are associated deterioration in symptoms following treatment between the LI interventions and all patients receiving any IAPT interventions.

Table 4.17. Stepwise regression model: deterioration (Low intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.941	0.923 to 0.958	0	1.45%
Model Step 2				
PHQ-9 T1	0.98	0.959 to 1.002	0.074	3.76%
GAD-7 T1	0.894	0.87 to 0.718	0	
Model Step 3				
GAD-7 T1	0.884	0.863 to 0.906	0	3.65%
Model Step 4				
GAD-7 T1	0.875	0.853 to 0.898	0	4.04%
Phobia caseness	1.404	1.13 to 1.745	0.002	
Model Step 5				
GAD-7 T1	0.871	0.848 to 0.894	0	4.15%
Phobia caseness	1.301	1.039 to 1.631	0.022	
Functioning	1.017	1.003 to 1.03	0.016	
Model Step 6				
GAD-7 T1	0.854	0.83 to 0.878	0	5.55%
Phobia caseness	1.419	1.11 to 1.813	0.005	
Functioning	1.015	1.001 to 1.03	0.038	
Ethnicity Group	1.412	1.081 to 1.845	0.011	

Treatment dropout - Low Intensity

The stepwise regression model for treatment dropout during LI interventions is presented in Table 4.18. The order of patient variables added to the regression model were the same as those entered in the dropout analysis for 'all patients', and is presented in Table 4.4. Higher age was found to be reduce the odds of dropping out of treatment in the first step of the model, and then higher PHQ-9 scores were found to significantly increase the odds of dropout. Next, gender was found to be significantly associated with dropping out of treatment, as female patients less likely to dropout of treatment compared to male patients in receipt of LI treatments. Receiving welfare benefits was significantly associated with higher odds of dropping out of treatment at Step 4 of the regression model, and having been prescribed medication was then significantly associated with a higher risk of dropping out of treatment. No other patient characteristics were significantly associated with dropout, and the variance explained by this final regression model was just 1.58%.

Table 4.18. Stepwise regression model: dropout (Low intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
Age	0.984	0.979 to 0.988	0	0.74%
Model Step 2				
Age	0.983	0.978 to 0.987	0	1.32%
PHQ-9 T1	1.034	1.023 to 1.045	0	
Model Step 3				
Age	0.982	0.978 to 0.987	0	1.44%
PHQ-9 T1	1.034	1.023 to 1.046	0	
Gender	0.84	0.742 to 0.951	0.006	
Model Step 4				
Age	0.982	0.976 to 0.987	0	1.50%
PHQ-9 T1	1.027	1.015 to 1.04	0	
Gender	0.859	0.751 to 0.983	0.027	
Welfare status	1.296	1.115 to 1.506	0.001	
Model Step 5				
Age	0.981	0.976 to 0.987	0	1.58%
PHQ-9 T1	1.024	1.011 to 1.036	0	
Gender	0.87	0.759 to 0.997	0.045	
Welfare status	1.305	1.118 to 1.522	0.001	
Medication prescribed	1.21	1.052 to 1.391	0.008	

The results of the regression model for dropout of treatment was quite different from the regression model for 'all patients' presented in Table 4.6. Gender and medication prescription status were found to be associated with dropout in LI only, and not for 'all patients'. This suggests that these patient variables may need further consideration by clinicians working in LI, as they may affect treatment adherence.

Summary – Low Intensity

The findings presented in these analyses of patients attending LI interventions appear similar for the three recovery/reliable change outcomes, but differ in relation to deterioration and dropout when compared to the 'all patients' analyses. As nearly 50% of the full sample was made up of individuals receiving LI interventions only, it may not be surprising that the final regression models for positive outcomes in this subgroup were similar to those for the full sample for some of the outcomes. For both the recovery and reliable change outcomes, the final regression models included the same patient characteristics in both the LI and 'all patients' sample, with both the direction and magnitude of effects being very similar.

A slight difference was identified between samples for the 'recovery or improvement' outcome, as phobia caseness was not associated with the outcome in the LI sample, but was included in the final regression model for the 'all patients' analysis. One explanation for this is that individuals with phobias may have been viewed as a more complex group of patients and there were more likely to be seen at HI initially, therefore had less representation at Step 2. The final model for deterioration following LI interventions was also different to that for the 'all patients' sample and suggests that although welfare status does not increase the risk of deterioration in LI treatments, patients from non-white ethnic groups are at greater risks of negative outcomes than white patients. This may be due to the nature of interventions provided, as LI interventions will typically use more printed or computer-based literature, for example cCBT or self-help books, and it is possible that these materials may not be culturally relevant for people of non-white ethnic background, increasing risk of poorer outcomes.

Two additional patient characteristics were associated with dropout in LI that were not significantly associated in the analysis of 'all patients' entering treatment. Gender was found to be associated with dropout, as men were more likely to drop out of treatment compared to women, and being prescribed medication was also associated with a lower likelihood of completing LI treatment. It may be that patients who were prescribed medication were more severe or complex in nature than individuals who were not prescribed psychotropic medications, which may have been linked to increased dropout. An alternative theory could be that individuals receiving medication as well as psychological treatment preferred the pharmacological treatment, and therefore dropped out of treatment as they no longer wished to receive IAPT treatments. As higher severity of depression was also associated with more dropout of treatment, it is possible that LI interventions were not sufficient to treat more

severe or complex patients, and potentially HI interventions may have been a better initial treatment which could have reduced the risk of dropout.

Patients receiving High Intensity treatments

The outcomes for patients attending HI interventions are displayed in Table 4.19. Compared to the results for 'all patients' entering treatment (Table 4.5), there was a slightly higher percentage of patients in recovery, achieving reliable change and recovery or improvement. There was also much less dropout of treatment (24.84% HI vs 33.47% for 'all patients') although a very slightly higher percentage deteriorating following treatment (9.08% vs 8.71%).

In comparison to outcomes for patients attending LI interventions only (Table 4.13), it was found that the percentages of recovery and for the recovery or improvement outcome were very similar, however a higher percentage of patients achieved reliable change following HI interventions than LI, which may be explained by the higher pre-treatment severity scores for patients receiving HI only (Table 4.12). There was less dropout in HI interventions than LI (24.84% vs 33.41%) but more deterioration (9.08% vs 7.74%).

Table 4.19. High intensity: Outcomes following treatment.

Outcome	No	Yes	%
Recovery	1,789	1,274	41.59%
Reliable change	1,774	1,289	42.08%
Recovery or improvement	1,556	1,507	49.20%
Deterioration	2,785	278	9.08%
Dropout	2,302	761	24.84%

Recovery - High Intensity

The results of the stepwise regression of recovery is presented in Table 4.20. The T1 PHQ-9 score had a significant association with recovery following treatment, and as with 'all patients' and LI only patients analyses, higher depression scores at T1 reduced the odds of recovery. Higher GAD-7 at T1 and Phobia caseness were both significantly associated with reduced odds of recovery in the next steps of the model. The level of work and social functioning in addition to welfare status were then found to be significantly associated with the outcome, as higher functioning impairment and being on benefits reduced the odds of recovery. No other patient characteristics were significantly associated with the recovery and the variance explained by the final model was 12.03%.

The results of the logistic regression analysis for HI only patients was similar to those for 'all patients' and LI only patients, with the exception that phobia caseness remained significantly associated with recovery in HI cases. This suggests that co-occurring phobia may have

more impact on a patient's likelihood of recovery, when controlling for other patient characteristics, in HI interventions compared to LI treatments.

Table 4.20. Stepwise regression model: recovery (High intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.88	0.868 to 0.892	0	9.09%
Model Step 2				
PHQ-9 T1	0.904	0.89 to 0.919	0	10.19%
GAD-7 T1	0.935	0.917 to 0.954	0	
Model Step 3				
PHQ-9 T1	0.911	0.896 to 0.926	0	10.84%
GAD-7 T1	0.942	0.924 to 0.961	0	
Phobia caseness	0.651	0.552 to 0.769	0	
Model Step 4				
PHQ-9 T1	0.924	0.908 to 0.941	0	11.25%
GAD-7 T1	0.944	0.925 to 0.963	0	
Phobia caseness	0.698	0.588 to 0.828	0	
Functioning	0.98	0.97 to 0.99	0	
Model Step 5				
PHQ-9 T1	0.933	0.915 to 0.951	0	12.03%
GAD-7 T1	0.945	0.925 to 0.965	0	
Phobia caseness	0.75	0.624 to 0.9	0.002	
Functioning	0.986	0.975 to 0.998	0.019	
Welfare status	0.524	0.435 to 0.631	0	

Reliable change - High Intensity

The results of the stepwise logistic regression analysis of patient characteristics associated with achieving reliable change when in receipt of HI interventions are presented in Table 4.21. Higher initial depression and anxiety severity were both found to be significantly associated with achieving reliable change following treatment. In the next steps of the analysis, being phobic and poorer levels of functioning reduced the odds of achieving reliable change in patients receiving HI interventions. In the final step, welfare status was added to the regression model and it was found that being on benefits significantly reduced the odds of reliable change. No other patient characteristics were significantly associated

with achieving reliable change and the final model explained 5.36% of the variance in outcomes.

The final regression model for reliable change in HI interventions differed from the models for 'all patients' and LI only as neither ethnicity nor medication prescription status were found to be significantly associated with the outcome in HI only patients. These characteristics may have less impact on reliable change in HI interventions.

Table 4.21. Stepwise regression model: reliable change (High intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	1.05	1.037 to 1.063	0	1.49%
Model Step 2				
PHQ-9 T1	1.024	1.009 to 1.039	0.002	2.37%
GAD-7 T1	1.06	1.04 to 1.081	0	
Model Step 3				
PHQ-9 T1	1.031	1.016 to 1.047	0	2.81%
GAD-7 T1	1.067	1.046 to 1.088	0	
Phobia caseness	0.707	0.601 to 0.833	0	
Model Step 4				
PHQ-9 T1	1.046	1.028 to 1.064	0	3.13%
GAD-7 T1	1.069	1.049 to 1.091	0	
Phobia caseness	0.755	0.638 to 0.893	0.001	
Functioning	0.982	0.973 to 0.992	0	
Model Step 5				
PHQ-9 T1	1.064	1.044 to 1.084	0	5.36%
GAD-7 T1	1.072	1.05 to 1.095	0	
Phobia caseness	0.807	0.673 to 0.968	0.021	
Functioning	0.987	0.976 to 0.998	0.018	
Welfare status	0.446	0.372 to 0.536	0	

Recovery or Improvement - High Intensity

The results of the analysis for recovery or improvement following HI treatment are presented in Table 4.22. The ORs for PHQ-9 and GAD-7 scores suggest that lower levels of severity at T1 were associated with increased odds of achieving the positive outcome, whereas phobia caseness was associated with lower odds of recovery or improvement being achieved. The

level of functioning was then added to the model and found to be positivity associated with the outcomes, as lower functioning at T1 was associated with reduced odds of positive outcome. However, the addition of phobia caseness to the model resulted in the GAD-7 score becoming non-significant and therefore anxiety severity was dropped from the regression model in the following step. In the final model step, welfare status was added, and it was found that receiving benefits was associated with a lower likelihood of achieving the positive outcome from treatment. No other patient variables improved the model fit, and the variance explained was 10.12%.

Table 4.22. Stepwise regression model: recovery or improvement (High intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.891	0.88 to 0.903	0	7.59%
Model Step 2				
PHQ-9 T1	0.901	0.887 to 0.915	0	7.75%
GAD-7 T1	0.975	0.956 to 0.994	0.009	
Model Step 3				
PHQ-9 T1	0.908	0.893 to 0.922	0	8.29%
GAD-7 T1	0.98	0.961 to 0.999	0.044	
Phobia caseness	0.673	0.571 to 0.792	0	
Model Step 4				
PHQ-9 T1	0.919	0.903 to 0.936	0	8.67%
GAD-7 T1	0.983	0.964 to 1.003	0.089	
Phobia caseness	0.72	0.608 to 0.852	0	
Functioning	0.982	0.972 to 0.992	0	
Model Step 5				
PHQ-9 T1	0.913	0.899 to 0.928	0	8.60%
Phobia caseness	0.706	0.597 to 0.835	0	
Functioning	0.981	0.971 to 0.991	0	
Model Step 6				
PHQ-9 T1	0.923	0.907 to 0.94	0	10.12%
Phobia caseness	0.758	0.632 to 0.908	0.003	
Functioning	0.988	0.977 to 0.999	0.036	
Welfare status	0.475	0.398 to 0.568	0	

The findings presented for characteristic associated with recovery or improvement following HI treatments were similar to those for ‘all patients’ receiving treatment from the services, except that GAD-7 was not significantly associated with the positive outcome in the final

multivariate model. These findings are surprising considering that GAD-7 at T1 was associated with both recovery and the reliable change alone outcomes in HI treatment. Compared to LI interventions, phobia caseness was significantly associated with the outcome in HI interventions, but not LI, suggesting phobia may have more influence on the outcomes at Step 3 of IAPT treatment.

Deterioration - High Intensity

Table 4.23. Stepwise regression model: deterioration (High intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.942	0.923 to 0.962	0	1.71%
Model Step 2				
PHQ-9 T1	0.984	0.96 to 1.009	0.212	4.42%
GAD-7 T1	0.896	0.87 to 0.923	0	
Model Step 3				
GAD-7 T1	0.887	0.864 to 0.911	0	4.34%
Model Step 4				
GAD-7 T1	0.878	0.854 to 0.902	0	4.80%
Phobia caseness	1.636	1.24 to 2.157	0	
Model Step 5				
GAD-7 T1	0.873	0.848 to 0.899	0	5.54%
Phobia caseness	1.659	1.235 to 2.229	0.001	
Ethnicity Group	1.52	1.128 to 2.048	0.006	
Model Step 6				
GAD-7 T1	0.857	0.83 to 0.884	0	7.95%
Phobia caseness	1.431	1.031 to 1.986	0.032	
Ethnicity Group	1.324	0.952 to 1.843	0.095	
Welfare status	2.333	1.717 to 3.17	0	
Model Step 7				
GAD-7 T1	0.864	0.839 to 0.89	0	7.11%
Phobia caseness	1.416	1.042 to 1.924	0.026	
Welfare status	2.311	1.734 to 3.082	0	

The results of the analysis on deterioration from treatment are presented in Table 4.23. Although a higher score on the PHQ-9 at T1 was associated with lower odds of deteriorating following treatment initially, adding GAD-7 scores at T1 to the model resulted in depression

severity becoming non-significant. This would suggest that baseline anxiety has a stronger association with deterioration than baseline depression in patients receiving HI interventions. Being phobic and from a non-white ethnic group were also found to be significantly associated with increased odds of deterioration, but by adding welfare status in the next step of the regression model, ethnicity became non-significant. Welfare status was included in the final model and it was found that receiving benefits was significantly associated with increased odds of deterioration in symptoms following treatment. No other patient characteristics improved the model, so the final model included just GAD-7, phobia caseness and welfare status, with 7.11% of the variation explained by these characteristics.

The results of this regression model for HI interventions was similar to that for LI interventions, except that higher levels of impairment in work and social functioning were significantly associated with the negative outcome in LI interventions. The regression model for 'all patients' (Table 4.9) found that PHQ-9 was associated with deterioration but this was not found to be significantly associated with deterioration in patients attending HI only.

Treatment dropout – High Intensity

The results of the logistic regression analysis for treatment dropout in HI patients is presented in Table 4.24. The final regression model included patient age, T1 PHQ-9 score and welfare status, with higher age decreasing the odds of dropping out of treatment, whereas higher severity of depression and receiving benefits increased the odds of treatment dropout. No other variables were significantly associated with treatment dropout in the stepwise regression model.

The variance explained by the final regression model was 3.43%, which was higher than the percentage in the regression model for dropout in the 'all patients' sample (1.92%) even though it included the same patient characteristics.

Table 4.24. Stepwise regression model: dropout (High intensity)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
Age	0.983	0.976 to 0.989	0	0.83%
Model Step 2				
Age	0.98	0.973 to 0.986	0	2.95%
PHQ-9 T1	1.064	1.048 to 1.079	0	
Model Step 3				
Age	0.978	0.97 to 0.986	0	3.43%
PHQ-9 T1	1.048	1.032 to 1.066	0	
Welfare status	1.627	1.344 to 1.969	0	

Summary – High Intensity

A number of patient characteristics identified in the low intensity analysis were also significantly associated with outcomes in the high intensity intervention sample, suggesting these characteristics may predict outcomes across interventions. For example, severity, level of work and social functioning, as well as welfare status were all associated with recovery, reliable change and the recovery or improvement outcomes in the same direction across intervention samples. It is of interest that the variance explained by the regression models of the HI intervention sample were typically higher than the R^2 for the LI regression models, or those for the 'all patients' sample.

One difference between the regression models for HI interventions when compared to LI was that phobia caseness was significantly associated with both the recovery and reliable change outcomes in HI, but not LI. One suggestion for this difference is that individuals meeting caseness for phobia may have been more frequently allocated to HI interventions as an initial step, and therefore the representation of phobia at LI was not high enough to find an association in the LI analysis. Phobia was significantly associated with all four of the treatment response/deterioration outcomes in the HI sample, with the presence of phobia predicting poorer outcomes, and would suggest that it may be an important characteristic to consider when developing HI treatment plans.

Whereas medication prescription status and ethnicity group were associated with reliable change in LI, neither factor was significantly associated with this outcome, or any other outcome, in HI patients. It is possible that ethnicity is not associated with outcomes in HI interventions as the more intensive and person-centred nature of the interventions decreases the likelihood of any cultural biases that may be present in more text-focused interventions (e.g. self-help books at LI). Medication prescription status may be linked to more chronic, severe or complex presentation hence the higher proportion of patients prescribed medication referred to HI compared to LI (Table 4.12), resulting in a lack of predictive ability of this characteristic in patients receiving HI treatments.

Stepped up patients

Patients who entered treatment at Step 2/LI treatment but were later moved to Step 3/HI are defined as 'stepped up' patients. Their outcomes following treatment are presented in Table 4.25. Compared to the outcomes for 'all patients', the recovery rates were around 3% lower, although reliable change was slightly higher. The percentage of stepped up patients who dropped out was much higher than the proportion of the 'all patient' sample who dropped out (38.14% versus 33.47%), and percentage of patients deteriorating were higher, approaching 10%.

Table 4.25. Stepped up: Outcomes following treatment.

Outcome	No	Yes	%
Recovery	985	575	36.86%
Reliable change	919	641	41.09%
Recovery or improvement	853	707	45.32%
Deterioration	1,407	153	9.81%
Dropout	965	595	38.14%

Recovery – Stepped Up

The results of the stepwise multivariate logistic regression analysis for recovery in patients who were stepped up is presented in Table 4.26. The final model includes the same variables as the 'all patients' and LI analyses, and the only difference with the final HI regression model is that phobia caseness was not significantly associated with recovery for stepped up patients. The variance explained was 11.23%, which is similar to that found in the 'all patients' analysis.

Table 4.26. Stepwise regression model: recovery (Stepped up)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.878	0.86 to 0.896	0	8.48%
Model Step 2				
PHQ-9 T1	0.902	0.881 to 0.923	0	9.62%
GAD-7 T1	0.931	0.904 to 0.958	0	
Model Step 3				
PHQ-9 T1	0.918	0.895 to 0.942	0	10.07%
GAD-7 T1	0.936	0.909 to 0.964	0	
Functioning	0.975	0.961 to 0.989	0.001	
Model Step 4				
PHQ-9 T1	0.915	0.891 to 0.94	0	11.23%
GAD-7 T1	0.935	0.906 to 0.964	0	
Functioning	0.98	0.965 to 0.996	0.012	
Welfare status	0.621	0.468 to 0.826	0.001	

It is of interest that phobia caseness was not significantly associated with recovery in stepped up patients even in the first steps of the regression model. In LI and the 'all patients' analysis, the ORs for phobia caseness are statistically significant in the earlier stages of the

stepwise regression model, but the lack of association with recovery in stepped up patients suggests the presence of phobia may not have an important role when deciding whether stepping up would benefit the patient or not. It may be that patients with phobia were more likely to be referred straight to HI interventions for initial treatment.

Reliable change – Stepped Up

The results of the stepwise model for reliable change is presented in Table 4.27. The final regression model included PHQ-9, GAD-7, the level of social and occupational functioning and welfare status.

Table 4.27. Stepwise regression model: reliable change (Stepped up)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	1.04	1.021 to 1.059	0	0.86%
Model Step 2				
PHQ-9 T1	1.022	1.001 to 1.044	0.044	1.31%
GAD-7 T1	1.045	1.016 to 1.076	0.002	
Model Step 3				
PHQ-9 T1	1.026	1.004 to 1.048	0.023	1.69%
GAD-7 T1	1.055	1.024 to 1.086	0	
Phobia caseness	0.798	0.642 to 0.993	0.043	
Model Step 4				
PHQ-9 T1	1.045	1.019 to 1.07	0	2.22%
GAD-7 T1	1.056	1.025 to 1.087	0	
Phobia caseness	0.859	0.687 to 1.072	0.179	
Functioning	0.977	0.964 to 0.991	0.001	
Model Step 5				
PHQ-9 T1	1.044	1.019 to 1.07	0.001	2.14%
GAD-7 T1	1.053	1.023 to 1.084	0	
Functioning	0.975	0.962 to 0.989	0	
Model Step 6				
PHQ-9 T1	1.05	1.024 to 1.078	0	3.72%
GAD-7 T1	1.059	1.027 to 1.092	0	
Functioning	0.979	0.965 to 0.993	0.004	
Welfare status	0.476	0.366 to 0.618	0	

The results of this analysis were very similar to those for HI interventions, except that phobia caseness did not remain significantly associated with reliable change for stepped up cases. Compared to the analysis LI only patients, the stepped up results did not find ethnicity group or medication prescription status to be significantly associated with reliable change whereas they did seem important for reliable change in LI. The variance explained by the final model for stepped up cases was very low, at just 3.72%.

Recovery or Improvement – Stepped Up

The results for the stepwise regression model for recovery or improvement is presented in Table 4.28. Higher PHQ-9 and GAD-7 scores, lower scores on the work and social functioning scale, receiving benefits and being female were all found to reduce the odds of achieving a positive outcome.

The final regression looks quite different from those for the other samples. One major difference is that gender was found to be significantly associated with the outcome, as females were less likely to achieve a positive outcome if they were stepped up, although the p-value was very close to non-significance (p=0.047). At 8.65%, the amount of variance explained by the final model was higher compared the R^2 of the LI model, but lower than that for HI.

Table 4.28. Stepwise regression model: recovery or improvement (Stepped up)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.898	0.881 to 0.915	0	6.05%
Model Step 2				
PHQ-9 T1	0.913	0.893 to 0.933	0	6.45%
GAD-7 T1	0.959	0.932 to 0.986	0.004	
Model Step 3				
PHQ-9 T1	0.933	0.91 to 0.956	0	6.99%
GAD-7 T1	0.964	0.936 to 0.992	0.011	
Functioning	0.972	0.959 to 0.986	0	
Model Step 4				
PHQ-9 T1	0.935	0.912 to 0.96	0	7.06%
GAD-7 T1	0.965	0.936 to 0.994	0.019	
Functioning	0.93	0.959 to 0.988	0	
Ethnicity group	0.731	0.55 to 0.972	0.031	
Model Step 5				
PHQ-9 T1	0.936	0.911 to 0.962	0	8.56%
GAD-7 T1	0.962	0.931 to 0.993	0.016	
Functioning	0.979	0.964 to 0.995	0.008	
Ethnicity group	0.806	0.596 to 1.091	0.163	
Welfare status	0.525	0.398 to 0.692	0	
Model Step 6				
PHQ-9 T1	0.934	0.91 to 0.959	0	8.45%
GAD-7 T1	0.961	0.932 to 0.99	0.009	
Functioning	0.978	0.963 to 0.992	0.003	
Welfare status	0.538	0.413 to 0.7	0	
Model Step 7				
PHQ-9 T1	0.933	0.909 to 0.958	0	8.65%
GAD-7 T1	0.964	0.935 to 0.993	0.017	
Functioning	0.977	0.962 to 0.992	0.002	
Welfare status	0.533	0.409 to 0.695	0	
Gender	0.787	0.621 to 0.997	0.047	

Deterioration – Stepped Up

The results of the stepwise regression analysis for deterioration in stepped up patients is presented in Table 4.29. The final regression included just GAD-7 at T1 and welfare status, with higher severity of anxiety reducing the odds of deterioration, whereas being on benefits increased the likelihood. Although PHQ-9 at T1 and phobia caseness had a significant association with deterioration at certain steps of the model, their adjusted ORs were not statistically significant when controlling for the other included characteristics. The variance explained was limited to just 4.3%.

The results of this analysis are different to the other deterioration analyses, as just two characteristics remained in the final model. Patient characteristics such as depression severity, phobia caseness and level of functioning were all included in the final model for deterioration in the 'all patients' analysis, and ethnicity was also included in the LI final regression model, which suggests the role of patient characteristics with deterioration in symptoms may vary between types of intervention.

Table 4.29. Stepwise regression model: deterioration (Stepped up)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
PHQ-9 T1	0.936	0.908 to 0.964	0	1.96%
Model Step 2				
PHQ-9 T1	0.969	0.937 to 1.003	0.075	3.92%
GAD-7 T1	0.907	0.869 to 0.946	0	
Model Step 3				
GAD-7 T1	0.891	0.858 to 0.925	0	3.61%
Model Step 4				
GAD-7 T1	0.883	0.849 to 0.919	0	3.90%
Phobia caseness	1.448	1.007 to 2.082	0.046	
Model Step 5				
GAD-7 T1	0.886	0.85 to 0.923	0	4.16%
Phobia caseness	1.216	0.827 to 1.789	0.32	
Welfare status	1.597	1.08 to 2.362	0.019	
Model Step 6				
GAD-7 T1	0.888	0.854 to 0.924	0	4.30%
Welfare status	1.733	1.186 to 2.532	0.004	

Treatment dropout – Stepped Up

The stepwise multivariate logistic regression for treatment dropout in stepped up patients is presented in Table 4.30. The final regression model includes age, PHQ-9 baseline severity and welfare status, and therefore was similar in structure to the analysis of HI and 'all patients' entering treatment. The variance explained by this final model was low at 1.69%.

Table 4.30. Stepwise regression model: dropout (Stepped up)

Step and characteristics included	Odds Ratio	95% confidence intervals	P-value	R^2
Model Step 1				
Age	0.989	0.981 to 0.997	0.007	0.35%
Model Step 2				
Age	0.987	0.979 to 0.995	0.002	1.06%
PHQ-9 T1	1.037	1.018 to 1.056	0	
Model Step 3				
Age	0.986	0.978 to 0.995	0.002	1.69%
PHQ-9 T1	1.037	1.016 to 1.058	0	
Welfare status	1.442	1.126 to 1.847	0.004	

Summary – Stepped Up

Patients who were stepped up during treatment were the smallest subgroup included in this analysis (less than 15% of the full sample of patients). However, the results of this subgroup of patients indicated that some of the characteristics that are significantly associated with outcomes in the other intervention subgroups were similarly associated with outcomes in this sub-population, especially with regard to severity, social functioning and welfare status.

As with the results of the LI intervention analysis, phobia caseness was not associated with the recovery, reliable change or the recovery or improvement outcomes whereas this characteristic significantly contributed to the regression models for the HI intervention analysis. As mentioned previously, it is possible that phobia was associated with the criteria to allocate to HI as initial treatment as it may increase the potential complexity of the presenting symptoms, and therefore phobia was less prevalent in the sample of individuals who were stepped up/started at Step 2. Gender was also associated with the recovery or improvement outcome, and the results suggested that female patients had a reduced likelihood of achieving this outcome compared to men. As the p-value for this odds ratio was very close to 0.05, it is possible that this finding may have limited clinical significant, and therefore replication in another dataset would be useful to compare results.

The final regression model for deterioration in stepped up individuals only included baseline GAD-7 score and welfare status as significant predictors of negative outcomes following treatment. Other patient characteristics such as social functioning and phobia caseness were associated with this outcome in the other intervention subgroups, and the finding that these patient characteristics did not predict deterioration in stepped up patients might be due to patients receiving both types of intervention. This cumulative treatment effect may not be predicted by either functioning or phobia caseness, and instead other patient characteristics not included in the current IAPT MDS may have more association with this outcome.

Discussion.

The analyses presented in this chapter have identified a number of associations between patient characteristics collected as part of the IAPT MDS and treatment outcomes within the services. Many of the characteristics were shown to have consistent associations with treatment outcomes, including across patients receiving different intensities of interventions. Additionally, there were a number of patient characteristics that had different associations between intensity of intervention and with different outcomes, and therefore could provide useful information on treatment selection decisions on the choice between LI and HI treatment. However, the amount of variance explained by any of the regression models was quite low, with 12% the highest explained by any model.

The results from the systematic review in Chapter 3 suggested that severity is among the most frequently reported patient characteristics associated with response to treatment, although the direction of this association was inconsistent with some studies finding higher severity predicted better response, others the opposite. The results of the analyses presented in this chapter suggest that the association between severity and outcome depends on whether absolute change in symptom scores or a decrease below a threshold is the criteria for achieving a positive outcome. For the outcome ‘moving to recovery’, higher severity was consistently associated with lower odds of recovery, across different interventions. When ‘reliable change’ was the outcome of interest, higher severity was associated with increased odds of achieving the outcome. This may be related to the criteria for achieving the outcomes themselves, as moving to recovery requires the patient to move from above a cut-off/threshold to below following treatment, and therefore the closer to the threshold a patient scores at T1 (lower severity) the less symptoms needed to decrease in order to be under the cut-off. For the reliable change outcome, a higher pre-treatment score provides more opportunity for a decrease in symptoms to reach a clinically significant amount, but regression to the mean (the increased likelihood of more extreme scores moving towards to mean over multiple time points), may also contribute some of the findings when severity is considered. As residual symptoms are found to be a factor in predicting later relapse to depression following successful treatment (Lin et al, 1998), reducing the

severity of symptoms may reduce the likelihood of recurrence and the need for further episodes of treatment.

One patient characteristic that was also frequently associated with response to treatment in the analyses was welfare status, as patients receiving benefits were regularly associated with worse outcomes (less recovery/change, more deterioration and dropout) across the different intervention intensities. It is possible that welfare status predicts poorer outcomes from treatment due to the social adversity that is potentially associated with not having a job, for example financial pressures or chronic physical illness, and previous studies have found these to be related to poorer outcomes in CMHD regardless of treatment type (Brown et al., 2010). Collecting more information on social adversity and employment status may help further refine the predictive ability of this patient characteristic. For now, it would appear that receiving welfare benefits is associated with poorer outcomes, and therefore it may be of value for clinicians to consider this when aiming to personalise treatment.

Although severity and welfare status showed a consistent association across treatment intensity, other patient characteristics were found to have different associations across the different treatment intensities. Ethnicity group for example had a significant association with treatment response outcomes in LI interventions, as non-white groups were associated with poorer outcomes. However, ethnicity was not significantly associated with outcomes in the HI or stepped up samples. This may be explained by the mode of delivery of LI interventions compared to HI interventions, as LI involves more self-help manual and texts, and these may not be seen as culturally relevant for non-white groups. Gender was also found to be significantly associated with the recovery or improvement outcome in stepped up cases only, with males more likely to achieve a positive outcome. However, the confidence intervals were very close to non-significance and therefore more research on this finding would be recommended.

The level of social functioning was frequently associated with poorer outcome following treatment in this analysis, regardless of whether it was recovery or reliable change that was being explored. However, it was only included in the final model for deterioration for LI cases when controlling for the other included variables and this suggests it may have less of a role in predicting deterioration in HI or stepped up cases. Poorer social functioning may be associated with more severe symptoms and greater complexity, as well as additional comorbid conditions other than depression or anxiety, which may have resulted in LI being less appropriate treatment than HI for these cases. Clinicians may benefit from considering the level of work and social functioning when considering LI interventions, and this may involve focusing on social or employment support to achieve a better outcome.

It is also of interest that PHQ-9 severity was significantly associated with deterioration in the ‘all patient’ sample but was not associated with this outcome in either of the three sub-analyses by intervention. Instead, it was found that the level of anxiety was consistently

associated with deterioration across samples, as higher pre-treatment anxiety reduced the odds of deterioration. One possible explanation for this finding may be that some patients who suffer from anxiety related issues and are avoidant may present with limited symptoms as measured on the GAD-7 but aspects of the interventions, for example behavioural experiments, may increase anxiety as they begin to through distressing symptoms.

The amount of variance explained by the regression models for reliable change and deterioration presented in this chapter appear low, especially for predicting treatment dropout. However, the R^2 values reported in the models for recovery ranged between 10.5% and 12%, which was above cut-off proposed for clinical importance suggested by researchers (Uher et al, 2012). Other analyses using IAPT data have reported higher R^2 values, but only when including a number of additional patient characteristics not included in the MDS. For example, Gyani and colleagues (2011) analysed national IAPT data using a range of characteristics including patient specific (e.g. demographics and clinical factors) as well as characteristics associated with treatment (e.g. type of therapy, number of sessions). The variance explained by their final logistic regression model for recovery was around 17%, suggesting only 6-7% more variance was explained when analysis included a large number of additional characteristics associated with interventions and therapists. The R^2 values for the analyses of treatment dropout were comparatively low, the highest being 3.43% which was found for HI interventions.

It is possible that patient characteristics identified in the meta-review, but not included or available in the IAPT dataset (e.g. previous treatment, comorbid personality disorders, and self-efficacy) could further predict IAPT treatment outcomes. One limitation of the present IAPT dataset is that it includes only a select number of patient characteristics, and some of these characteristics were limited as to the information they contained. For example, welfare status was used instead of employment because information about employment status was not available in the dataset. In addition, the only mental health comorbidity available in the dataset was the presence of phobia caseness or not, which was scored using the three IAPT phobia items in the dataset (IAPT, 2011). However, the psychometric properties of these items has not be investigated thoroughly, and future research might benefit from using additional measures of specific anxiety disorders, as well as personality disorders. This may provide more information about comorbidities, with a view to refining the regression models presented in this chapter. For the current project, additional characteristics could not be collected, and therefore alternative methods of using the nine available patient characteristics was explored in the next stages of this thesis.

Although multiple regression analysis allows the inclusion of a number of patient characteristics within a regression model, there may be more direct interactions between characteristics that may be associated with differential outcomes (Kraemer, 2013). For example, it may be that the baseline level of severity moderates the effect of gender on outcome. However, the inclusion of interaction terms in regression should always be

facilitated by some theory of the potential association, as a lack of prior theory could be viewed as merely 'fishing' (Johansson & Høglend, 2007). Therefore, further work on theoretical models of patient characteristics associated with treatment response may provide the ground work for further analysis of interaction terms using the patient characteristics included in this analysis.

As there were nine included patient variables in the current analysis, exploring the interactions between all variables within a standard regression model framework would become too complex, but also challenging to translate into a DST which may have clinical utility. It would also be likely that different interaction models would be built for the different outcomes considered in these analysis, resulting in reduced practical utility. As research into decision making has suggested that clinicians will often compare new patients to prototypes of patients with similar characteristics (Garb, 2005), a statistical method of identifying subgroups of patients from the available patient characteristics may prove clinically useful if these groupings were associated with different treatment outcomes. New patients could be compare to these subgroups, and the expected utility of treatments for each subgroup could be calculated to aid clinicians with decision making. Exploring methods of identifying sub-groups of patients entering treatment at the services, and comparing the likelihood of treatment outcomes between these sub-groups was the focus of the next two chapters.

Chapter 5. Identifying latent profiles of patients using routine IAPT data.

Abstract

The patient characteristics included in the IAPT minimum dataset could be used to predict the likelihood of treatment outcomes following psychological interventions. However, the results of the regression analysis were limited by the low amount of variance explained by the models, and therefore alternative statistical methods may provide a more clinically useful way of using the available patient characteristics. Statistical methods of grouping patients based on shared characteristics has been used to identify homogeneous groups of patients in clinical populations, and could be used to identify profiles of patients with differential response to IAPT treatment. Latent class methods, such as latent profile analysis (LPA) (Lazarsfield & Henry, 1968) have previously been used in mental health populations to provide information on types or profiles of patients presenting in clinical populations, but not to inform response to treatments (Rosellini & Brown, 2014; Unick, Snowden, & Hastings, 2009). This chapter presents the results from latent profile analysis (LPA) performed on the dataset of patients entering treatment at the IAPT services described in Chapter 4. The LPA identified eight distinct groups of patients with statistically different constellations of patent characteristics. For example, one profile displayed very severe symptoms, as well as a high likelihood of receiving welfare benefits and being prescribed psychotropic medication, whereas another profile was characterised by low symptom severity, younger aged, a low likelihood of phobic symptoms and less functional impairment. Further details on each of the eight profiles is described, and analysis of treatment outcomes for each of these profiles is presented in Chapter 6.

Introduction

The results of the multiple regression analysis presented in Chapter 4 suggest that a number of patient-specific characteristics are associated with recovery, reliable change, clinical deterioration and the early termination of treatment in the IAPT services. Most of these characteristics were also identified in the meta-review presented in Chapter 3, suggesting these characteristics may have value in predicting treatment outcomes for CMHD patients. However, the variance explained by the regression models was generally low, with a range of 1.58 to 12.03% across all outcomes and intensities of intervention, and therefore alternative methods of exploring the predictive ability of these patient characteristics was considered.

One possible reason for this relatively small amount of variance explained is that multiple regression analysis explores the impact of each characteristic on the outcome, whilst controlling for the other included characteristics, but it does not explore potential interactions between variables (Cox, 1984). Whilst interactions between multiple variables can be explored in regression analyses, prior theory is required to justify the use of interaction terms within the analysis (Johansson & Høglend, 2007). As there were no prior theories of interactions between all nine patient characteristics in the current analysis, and because a regression model containing the interactions between all nine characteristics would be difficult to interpret, an alternative method of exploring the patient characteristics together was considered.

Previous research into clinical decision making theorises that clinicians compare new patients to 'prototypes', subgroups of patients created in the clinician's mind that they may have treated previously and/or who may typify a specific diagnosis or illness (Garb, 2005). The clinician may make a judgement as to which 'prototype' a new patient most resembles, and therefore will likely recommend treatments they feel are most appropriate for this group of patients, incorporating previous experience with these prototypes. Less experienced clinicians may be at a disadvantage having seen fewer previous patients (and therefore prototypes). Whereas prototypes are subjective in nature, an objective method of using patient characteristics to identify common 'prototypes' or profiles of patients with similar characteristics could create subgroups of patients attending the services that could be used by all clinicians, and may be especially valuable to less experienced clinicians. There may be further clinical utility if differential response to treatments was found between these profiles of patients. The 'expected utility' of treatment outcomes for differing profiles may have the potential to inform clinical decision making concerning treatment selection in mental health services.

A grouping of patients in this way may also provide a better understanding of the nature of mental illness beyond diagnosis (Insel, 2006), especially if patient demographics in addition to clinical symptoms were included in the method of profiling patients. Clinical diagnoses

typically concern the symptoms of the illness, however the results of the regression analysis in Chapter 4 suggested that other characteristics such as level of social functioning and receiving welfare benefits are also associated with treatment outcomes, across intensities of interventions. Therefore, a multi-factorial approach using all available characteristics to generate profiles of patients entering treatment may provide a more valuable method of informing clinical decision making than focusing exclusively on diagnosis or individual characteristics alone.

Identifying clusters of patients may also have the benefit of allowing multiple treatment outcomes to be explored across stratified groupings of patients with similar characteristics, which may increase their potential clinical utility. The regression modelling presented in Chapter 4 resulted in a series of models, identifying different patient characteristics identified with different treatment outcomes. The results of this modelling suggested that characteristics may be associated with some outcomes but not others, for example, gender was not associated with recovery or reliable change in patients receiving HI treatments, whereas it was associated with dropout for the group. A DST that is able to consider multiple treatment outcomes could be of more value to treatment services such as IAPT, as identifying patients at risk of negative outcomes such as deterioration and treatment dropout may help inform treatment selection, as well as treatment monitoring decisions. Using the regression models presented in Chapter 4 to develop a DST therefore runs the risk that certain patient characteristics will be associated with a higher positive likelihood for some outcomes and lower likelihood of other outcomes, which would be difficult for clinicians to interpret in relation to supporting a clinical decision. The benefit of a clustering approach is that the identified groupings could be explored over a number of outcomes and potentially a DST using a clustering method could provide the probability of each outcome to the clinician. The probability of different treatment outcomes could then be used to estimate the likely expected utility of treatment and could then be used to inform decisions about the most appropriate treatment.

The two main statistical methods of identifying subgroups of patients are clustering methods, such as K-means clustering (Cox, 1957), and latent class methods (Goodman, 1974), for example latent class and latent profile analysis. Both methods are concerned with the recognition of patterns within the dataset of responses or patient characteristics that can be grouped into specific clusters or classes of responders, containing individuals with similar patterns of responses or characteristics.

K-means clustering (Cox, 1957; MacQueen, 1967) is a partitional clustering method popular across disciplines including the social sciences, although more frequently used as a marketing segmentation methods in consumer research (Schreiber & Pekarik, 2014). This method attempts to reduce all included variables into smaller groupings. The original method of K-means clustering uses interval or ratio data only, as equal distances are needed between values, although more recent extensions for categorical data exist (San, Huynh, &

Nakamori, 2004). It is a simpler method of clustering to implement which has increased its popularity amongst researchers (Jain, Murty, & Flynn, 1999). However, a major criticism of k-means clustering is that there are no objective criteria for judging the suitability of the clustering solution and therefore there is an inherent danger that a solution can be shaped to fit with the researcher's expectations. There are also concerns that K-means methods lack rigour and consistency, as adding new data may result in a quite different cluster solution (Krantz, Korn, & Menninger, 2009).

An alternative method of identifying distinct groups in a dataset is the use of Latent Class (LC) methods, which are used to identify homogenous sub-groups, referred to as "classes" or "profiles" depending on the specific analysis used, of individuals based on responses to variables (Goodman, 1974). On one level it is similar to k-means clustering, however model 'fit' is established using statistics available within data output, which results in a more objective method of determining the number of distinct profiles that exist within the dataset. Latent class analysis (LCA) was originally developed to include categorical variables only, and latent profile analysis (LPA) provides an extension that can incorporate categorical, continuous and ordinal variables (Hagenaars & McCutcheon, 2002; Lazarsfield & Henry, 1968). The type of data to be included within the analysis determines whether 'classes' or 'profiles' is used to name the subgroups identified, although both LCA and LPA fall within 'Latent Class methods'.

In LC methods, the probability of the specific pattern of responses **y**, P(Y = **y**), is a weighted average of the C class-specific probabilities P(Y = y|X = x); expressed as:

$$P(Y = y) = \sum_{x=1}^{C} P(X = x)P(Y = y|X = x)$$

Where P(X = x) is the proportion of individuals in Latent Class x. Groups of individuals with a similar pattern of responses or characteristics are identified to determine statistically differing groups within the dataset. Each individual case can be classified into a subgroup (referred to as the individual's latent class or latent profile) from their observed pattern of responses (e.g. to questionnaire items, presence of particular symptoms, dichotomous variables such as 'Male' or 'Female'). In LC methods, the inter-item relationship is explained by the presence of an unknown subgroup (the latent class/profile). Individual differences in the observed item response patterns are explained by differences in latent class/profile membership which are probabilistic.

Simulation and comparison studies suggest results of K-means and LC method analyses can vary, with a different number or clusters/classes identified using the same dataset (Eshghi, et al., 2011). When directly comparing methods, the majority of researchers suggest a better performance of latent class methods over K-means clustering (Schreiber & Pekarik, 2014). The preference for LC methods is also linked to the availability of statistical tests to confirm the number of classes/profiles, or model fit that do not exist in K-means

clustering methods (Magidson & Vermunt, 2002). A further benefit is that LC methods use a probabilistic approach, which reduces the vulnerability of these analyses to extreme scores or outliers when compared to k-means clustering (Schreiber & Pekarik, 2014).

LC methods have been used within mental health populations previously to identify sub-groups of patients within diagnostic groups based on a combinations of assessment scores and demographic information such as eating disorders (Duncan et al., 2005; Wade, Crosby, & Martin, 2006) and personality disorders (Bucholz, Hesselbrock, Heath, Kramer, & Schuckit, 2000; Fossati et al., 2001). These analyses have typically included the presence of certain symptoms or personality scores as indicator variables to identify clinical subgroups, such as 'binge-eating disorder' or specific personality disorders. Researchers suggest that these profiles may be useful in tailoring treatment, although these analyses did not go on to compare treatment outcomes between the profiles identified.

Within CMHDs, researchers have used LC methods to identify subgroups based on clinical symptoms. Rosellini and Brown (2014) performed LPA on a dataset of more than 1200 patients using six symptom severity subscale scores from questionnaires such as the Beck Depression Inventory (Beck, Steer, & Brown, 1996), the Anxiety Sensitivity Index (Peterson & Reiss, 1992) and the Social Interaction Anxiety Scale (Mattick, Peters, & Clarke, 1989) as the patient characteristics, and identified 6 latent profiles of patients (for example, "Obsessed-worried" and "Mildly neurotic"). Unick, Snowden and Hastings (2009) performed LCA on the presence or absence of DSM depression and anxiety symptoms using the University of Michigan Composite International Diagnostic Interview (UM-CIDI) (Kessler et al., 1994) to create 23 dichotomous items as the included patient characteristics. An analysis of n=1009 patients identified seven classes, for example "Mild Somatic Anxiety" and "Mild Psychological Depression". Both sets of authors focused only on symptom scores as patient characteristics and did not include demographic information. Patient demographic information may provide important information for the development of classes or profiles that may be predictive of treatment outcome. Although Unick et al looked at differences in service utilisation between the classes, neither sets of authors looked at differences in outcomes following treatment between the classes/profiles they identified.

One additional benefit of LC methods as an approach to identifying groups of individuals entering treatment for CMHDs is that any individual can be assigned to previously identified latent classes/profiles by calculating the probability of class/profile membership, referred to as posterior membership probability. This is calculated using Bayes rule (Bayes, 1763). For categorical variables, this is expressed as:

$$P(X = x|Y = y) = \frac{P(X = x)P(Y = y|X = x)}{P(Y = y)}$$

For continuous variables, the probability density distribution is used, and expressed as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}}\ exp\left(\frac{-(x_i - \mu_{ij})^2}{\sigma_{ij}^2}\right)$$

Where σ_{ij}^2 is the variance of item *i* within *j* class, x_i is the value of item *i* and $\mu_{ij}$is the mean of item *i* for class *j*. The potential clinical value of these posterior membership probabilities for this thesis was that they could be included within a DST to calculate class/profile membership for any new patient entering the service. An algorithm including the posterior probabilities from a latent class/profiles analysis would be straight-forward to develop and could be hosted by the local treatment services' EPMS to provide class/profile information to clinicians, potentially in real time.

The aim of this chapter was to identify statistically distinct profiles of patients entering treatment at the IAPT services by performing LPA on patient characteristics included in data MDS. Once the profile structure in the dataset has been identified, an algorithm using the posterior probabilities could be produced to generate the likelihood of profile membership for any new patient referred to the services from their patient characteristics. The differential treatment outcomes across the different intensities of psychological treatment was then explored in the next chapter.

Method

Sample

The dataset used for this analysis was the same data as described in Chapter 4. Whereas the multiple regression analyses excluded patients who received only a single treatment session and individuals who scored below clinical caseness, it was decided to include these individuals in the LPA. Therefore, the LPA included all patients who entered treatment at the services and had T1 assessment information, regardless of their symptom severity or number of treatment sessions. This decision was made so that the LPA could provide information and probability of profile membership for any new patient entering the service and completing an assessment, rather than just a subset who were clinical caseness or attended more than one session. This was deemed especially important if the LPA would have potential utility as a DST, as it would likely be more acceptable to clinicians if all patients could be entered into it, rather than a sub-group of patients referred to services.

The intensity of intervention was also not considered as inclusion criteria for the LPA analysis. It was expected that individuals belonging to the same profile could receive different intensities of intervention, and it was considered of interest to explore the distribution of profiles to interventions in the following chapter. The potential role of a DST would be to provide information about the most appropriate intensity of intervention for an individual in a given profile, and therefore outcomes between interventions within an identified profile would be explored in subsequent analyses.

From the sample of n=34741 patients who were referred to the two IAPT services in North London between September 2008 and March 2012, n=19817 included T1 assessment data. Of these individuals, n=3181 patients either declined treatment from IAPT, were still in treatment or referral, or were deemed not suitable for treatment from the service. A total of n=16636 patients were therefore included in the analysis, and a full flow diagram is presented in figure 5.1.

Measures

The patient characteristics included in the LPA were the same variables explored in the regression analysis presented in Chapter 4 (Table 4.1). Although some of these patient characteristics did not appear to be significantly associated with treatment outcomes across both intensities of psychological treatment, it was decided that they may still be of value to the LPA, especially as they may inform specific patterns of patient characteristics and/or contribute to specific patient profiles.

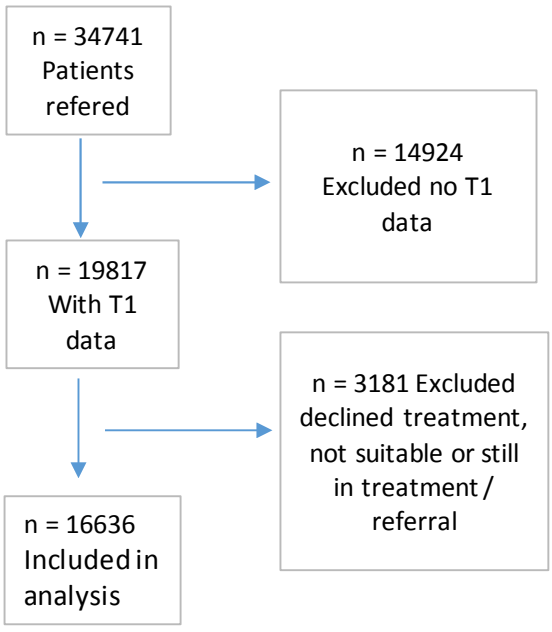

Figure 5.1. Patient flow - latent profile analysis

Data analysis

Data analysis was conducted in Mplus version 7 (Muthén & Muthén, 2012), on the initial Sept 2008 to March 2012 dataset. This primary dataset of n=16636 patients was split into two independent samples and LPA was performed separately on these samples to allow comparison and confirmation of the profile structure in two samples. This was viewed important to confirm whether the profile structure would be replicated in independent samples.

To identify the best fitting model for the datasets, the Vuong-Lo-Medell-Rubin Likelihood Ratio test (VLMR-LRT) (Lo, Mendell, & Rubin, 2001) and the Bootstrap Likelihood Ratio Difference test (B-LRT) were compared alongside the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and entropy values. Both the B-LRT and the VLMR-LRT compare the K model (current model with K number of profiles) to K-1 model (model with one less profile), with a significant p-value indicating the K model fits the data better than the model with one less profile. A non-significant finding (p-value >=0.05) suggests that the model with one less profile provides a better fit for the data, and the more parsimonious model would be preferred. Lower AIC and BIC value indicate better model fit, whereas higher entropy values indicate higher accuracy in classification for the model.

As there was no prior hypothesis on the exact number of patient profile groups from the data, the analysis was conducted starting with a two profile model and increasing the number of profiles until the VLMR-LRT became non-significant. The B-LRT was then used to confirm the K-1 model using a parametric bootstrap procedure (Asparouhov & Muthén, 2012; Geiser, 2013). Information from the AIC, BIC and entropy values were also used to inform model fit. This method was applied to both split samples of the primary dataset to confirm whether the same profile structures were identified across samples.

Results

A random split of the primary dataset resulted in two samples of n=8321 (Sample 1a) and n=8315 patients (Sample 1b). LPA was conducted on both samples independently using the same procedure, and model comparison statistics are presented in Table 5.1 for both split-samples. The LPA for Sample 1a yielded significant p-values on the VLMR-LRT comparing successive models, from a two-profile solution to an eight-profile solution (p=0.0057 at the eight-profile model), as well as decreasing AIC and BIC values. Although the BIC and AIC values were slightly lower for the nine-profile solution, the VLMR-LRT produced a non-significant p-value (p=0.3551) suggesting that increasing the number of profiles was not a better fit for the data (Asparouhov & Muthén, 2012; Geiser, 2013). The entropy value was also higher for the eight-profile solution suggesting higher classification accuracy, and

therefore the eight-profile solution was preferred. B-LRT was performed on the eight-profile solution, with a significant finding between the eight and seven profile models (p<0.0001).

Table 5.1. Model comparison statistics for both split samples.

Sample 1a					
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**
k = 2	265442	265604	265531	<0.05	0.798
k = 3	262848	263080	262975	<0.05	0.765
k = 4	262079	262382	262245	<0.05	0.725
k = 5	261548	261920	261752	<0.05	0.743
k = 6	260872	261315	261114	<0.05	0.751
k = 7	260533	261046	260814	<0.05	0.73
k = 8	260282	260865	260602	<0.05	0.722
k = 9	260044	260697	260402	0.3512	0.71
Sample 1b					
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**
k = 2	265744	265906	265833	<0.05	0.788
k = 3	263059	263290	263186	<0.05	0.771
k = 4	262377	262680	262543	<0.05	0.718
k = 5	261769	262141	261973	<0.05	0.734
k = 6	261079	261522	261321	<0.05	0.745
k = 7	260689	261202	260970	<0.05	0.728
k = 8	260341	260925	260661	<0.05	0.739
k = 9	260147	260800	260505	0.094	0.714

The LPA for Sample 1b also yielded significant increases in model fit according to the VLMR-LRT up to the eight-profile model (p<0.0001 for the eight-profile model compared to the seven-profile model) with decreasing and AIC and BIC values, and again the nine-profile model produced a non-significant VLMR-LRT p-value (p=0.940). The B-LRT confirmed a significant p-value for the eight-profile model compared to the seven-profile model (p<0.0001).

Following confirmation of an eight-profile model structure from the two independent split samples, the dataset was re-combined, and the same method of LPA was applied to the full sample of n=16636 to generate probabilities of profile membership for each patient included in the analysis. As before, VLMR-LRT showed a significant p-value up to the nine-profile model (p=0.699) with the eight-profile model selected and probability of profile membership assigned to each patient in the sample. Each patient was allocated to a latent profile (LP) by using the highest probability of profile membership as the patient's LP, with probabilities provided by Mplus.

Table 5.2. Latent profiles and associated patient characteristics

	Full sample	LP1 (n=3002)	LP2 (n=3771)	LP3 (n=513)	LP4 (n=674)	LP5 (n=1414)	LP6 (n=1514)	LP7 (n=2070)	LP8 (n=3678)
Age - Mean(SD)	37.9 (13.36)	33.47 (8.46)	30.74 (7.48)	66.83 (9.71)	65.16 (8.88)	54.25 (7.85)	40.72 (9.10)	42.74 (9.44)	29.68 (6.90)
PHQ-9 - Mean(SD)	13.85 (6.67)	5.37 (3.03)	11.28 (3.16)	4.59 (3.09)	10.76 (3.59)	17.88 (3.43)	13.36 (3.38)	22.86 (2.78)	18.85 (3.14)
GAD-7 - Mean(SD)	12.35 (5.51)	5.26 (2.71)	12.56 (2.98)	3.74 (2.58)	10.85 (3.26)	15.94 (2.88)	7.99 (2.78)	18.38 (2.50)	16.43 (2.79)
W&SAS - Mean(SD)	17.85 (9.69)	8.69 (5.79)	14.53 (6.23)	6.99 (6.32)	11.79 (7.14)	18.15 (7.38)	20.87 (6.69)	31.72 (5.74)	22.28 (7.11)
Gender - n(%) female	10793 (66.15%)	1906 (65.21%)	2570 (69.46%)	335 (67.81%)	451 (68.23%)	905 (65.01%)	850 (57.32%)	1166 (57.13%)	2610 (72.06%)
Ethnic Group – n (%) Non-White	3151 (22.24%)	452 (17.66%)	547 (17.33%)	46 (10.90%)	70 (12.22%)	219 (18.28%)	346 (26.59%)	498 (27.51%)	973 (30.89%)
Medication prescribed – n (%) Prescribed	5802 (38.71%)	721 (26.50%)	691 (20.02%)	114 (27.60%)	195 (33.22%)	659 (52.97%)	789 (59.01%)	1357 (72.61%)	1276 (37.90%)
Welfare status - n (%) on benefits	3834 (28.10%)	297 (12.00%)	262 (8.33%)	32 (8.67%)	48 (9.36%)	472 (40.45%)	679 (53.93%)	1258 (73.70%)	786 (26.12%)
Phobia Self-rating - n(%) phobia	7592 (50.66%)	585 (21.30%)	1261 (36.76%)	99 (22.86%)	205 (35.28%)	755 (60.35%)	750 (54.78%)	1738 (93.09%)	2199 (66.48%)

The distribution of patient characteristics for each LP alongside the full sample descriptives are displayed in Table 5.2 and a graphical representation of each profile is presented in Appendix D.

Comparing the means and proportions of each of the nine patient characteristics by LP with both the full sample and other LPs in Table 5.2 provides an understanding of the make-up of each profile of patients. For example, members of latent profile 1 (LP1) are younger, reporting lower symptom severity (on both PHQ-9 and GAD-7), and tend to score low the phobia scale as well as being less likely to be in receipt of welfare benefits and psychotropic medication when compared to the overall sample. The LP2 has a similar age, gender and ethnic group distribution to LP1, but with higher symptom severity and lower social functioning, as well as having a higher probability of phobic caseness. Comparing LP2 to the full sample means and proportions suggests a similar level of anxiety symptoms, slightly lower depression severity and nearly 15% less phobia caseness than the full population entering treatment at the services.

The highest symptom severity scores are for patients in LP7 (means of 23 and 19 on the PHQ and GAD), and this group also have a much higher proportion of patients receiving welfare benefits and prescribed psychotropic medication compared to the full sample and the other profiles. These characteristics may be inter-dependent, so higher severity of

depression or anxiety is associated with a patient receiving welfare benefits and prescribed psychotropic medication. However, LP8 have a considerably lower likelihood of receiving welfare benefits compared to LP7 (26% vs 73%) yet the severity scores are similar to those of LP7. This would suggest that the profiles are identifying clusters of patients with distinct constellations of patient characteristics, which may result in differential treatment outcomes.

There was a large variation in the size of the patient profiles, which suggests that certain patient groups are more typical of the service population that others. LP3 and LP4 have the lowest share of the population with 3.1% and 4.1% respectively. LP3 are a group of patients presenting with mild levels of severity, who are very likely to be white and very unlikely to be receiving welfare benefits, whereas members of LP4 report higher severity than LP3 (but less than the average for the full population). However, both LP3 and LP4 have much higher mean age than the other groups, and it may be that these profiles occur much less frequently in the IAPT services as these types of patients may be more likely to attend older adult services. The profiles most frequently taken into treatment by the IAPT services were LP2 and LP8, with 22.7% and 22.1% of the population respectively.

Description of the latent profiles

A more detailed description of each LP is provided below:

LP1.

In comparison to the full sample of patients attending the services, members of LP1 are younger and have lower mean scores on both the PHQ-9 and GAD-7, suggesting a less severe population of patients. The mean W&SAS score is also much lower for members of LP1 when compared to the full sample mean, suggesting social and occupational functioning is less impaired in this profile. Patients in LP1 were less likely to meet caseness for phobia, be in receipt of welfare benefits or prescribed medication. The percentage of the population in LP1 who are female is very similar to the full sample percentage, as is the proportion of individuals from non-white ethnic groups. In sum LP1 are a younger and less severe population, with limited functional impairment issues, low scores on the phobia scales, and are less likely to be prescribed psychotropic medication or to be on welfare benefits. LP1 make up 18% of the population taken into treatment.

LP2.

Members of LP2 have a lower mean age than the full sample of patients attending the services and are slightly younger than members of LP1. The mean level of anxiety severity (GAD-7) for patients in LP2 is similar to the full sample mean, whereas mean depression severity (PHQ-9) is slightly lower. The mean social and occupational impairment score for this profile is lower than for the full sample suggesting less impairment from symptoms. Individuals in LP2 have a lower probability of phobia, are very unlikely to be receiving

welfare or prescribed medication when compared to the average patient entering treatment at the services. The percentage of females in this profile is slightly higher than for the full population (69% vs 66%), and the percentage of individuals from non-white ethnic groups is slightly lower than the full sample (17% vs 22%). The description of this population suggests a group of patients whose symptoms are less chronic or potentially the first presentation to the services with consideration to the younger age, low probability of psychotropic medication and welfare benefit receipt. LP2 are most frequent LP taken into treatment (22.7% of the full sample).

LP3.

This profile of patients has the lowest mean symptom severity scores and the lowest functional impairment issues compared to all the LPs, suggesting a less disabled group of patients. The likelihood of receiving welfare benefits, prescribed medication and caseness for phobia is also low for this profile. The proportion of individuals from non-white ethnic groups is very low at 11% compared to the full sample (22%) and is the lowest of all the profiles. LP3 is also has the highest mean age (67 years old). This sub group make up only 3.1% of the full sample taken into treatment. This is a profile of older patients, who may be attending services for less severe common mental health disorders.

LP4.

Patients in LP4 have a mean age of 65 years old, making them the second oldest group of patients after LP3. Other similarities to LP3 are the proportion of female patients, medication prescription, non-white ethnic group proportion and the likelihood of receiving benefits. However, patients in LP4 typically have higher mean symptom scores (mild to moderate symptom severity), and more functioning issues as scored on the W&SAS compared to LP3. The levels of severity reported by members of this profile are slightly lower than the mean for the full sample. The percentage of the population in LP4 is just 4.1% suggesting this group of patients are not frequent attenders of the services.

LP5.

In comparison to the full sample of patients, LP5 have a higher mean age, as well as higher mean PHQ-9 and GAD-7 scores at presentation to the services. The level of functioning appears to be similar to the average impairment displayed by the full sample of patients entering treatment. The incidence of phobia, proportion prescribed medication and the likelihood of receiving welfare benefits is higher than for the full population, which suggests that there may be more disability as a result of symptoms in this profile of patients. LP5 have a slightly lower percentage of individuals from ethnic groups compared to the full sample. The incidence of LP5 in the full population is 8.5%. Compared to the previous four profiles, LP5 appear to have more severe levels of disability, indicated by higher symptoms, functional impairment, higher likelihood of psychotropic medication prescription and higher incidence of phobia.

LP6.

Whereas the other LPs and the full sample show similar levels of symptom severity on both the PHQ-9 and GAD-7 (i.e. within a point or two on each measure), LP6 are the only profile where there is a reasonable difference in mean depression and anxiety severity. The mean anxiety severity is 8 compared to a mean depression score of 13 for this profile, suggesting they may be attending the services for more depression focused symptoms and treatment. The mean age is similar to the full sample mean, and percentage of female patients is also close. The probability of being prescribed medication (59%) and receiving welfare benefits (54%) is higher in this profile of patients compared to the full sample, which may suggest a more chronic group of patients. The proportion of patients taken into treatment who are from LP6 was 9.1%.

LP7.

Patients in LP7 present to services with the highest baseline symptom severity compared to all other profiles, with means scores 9 and 6 points higher than the full sample means on the PHQ-9 and GAD-7 respectively. Mean age is slightly higher than the full sample mean and that of LP6. The level of functional impairment is also high, which may be consistent with the higher symptom severity scores. Over 90% of the group are self-rated as phobic, over 70% have been prescribed psychotropic medication and nearly 75% report receiving welfare benefits suggesting a much more disabled group. The percentage of non-white ethnic groups is higher compared to the full sample (28% vs 22%). LP7 make up over 9% of the sample of patients entering treatment.

LP8.

LP8 are the profile of patients with the highest proportion of non-white ethnic group patients attending the services (31%). They are also the youngest group of patients attending the services with a mean age of almost 30 years. Individuals in this profile are more likely to be female compared to the other profiles (72%). The mean PHQ-9 and GAD-7 scores are higher than the full sample mean scores, and patients in this profile have the second highest symptom scores of all profiles after LP7. The probability of receiving welfare benefits or being prescribed medication is very similar to the overall percentage of the population and therefore this group of patients may represent a younger and more acute profile of patients attending the services, potentially in their first episode of care.

It is interesting to note the outcome of a meeting where the results of the LPA and profile descriptions set out above were presented to a group of senior IAPT clinicians and PWPs as part of a dissemination of service outcomes. The clinicians reported that they could see value in the LPs and their descriptions, and importantly that they could recognise the different ‘types’ of patients amongst their own current caseloads. This provides some

encouragement for the role of LPAs in the development of a DST, as the ‘face’ validity that this represents may make it possible to engage clinicians in the further development and use of DSTs based on these LPs.

Discussion.

The LPA identified eight statistically reliable groups of patients in receipt of psychological treatment from the services, and this profile structure was replicated in two independent samples of over 8000 patients. Exploring the characteristics of the identified profiles shows a number of key differences, as well as similarities, between the sub-groups. The profiles alone may have potential utility in clinical audits, as they may provide a useful grouping of patients to compare between services and treatments, before differential response to treatments is considered.

If the patient characteristics of the LPs are compared to the results of the regression analyses in the previous chapter, then one may start to hypothesise likely outcomes of treatment for the LPs. Higher severity on both the PHQ and GAD were consistently associated with a lower likelihood of recovery following treatment, as was higher functional impairment and receiving welfare benefits. As members of LP7 have the highest average PHQ, GAD and W&SAS scores at T1 as well as a high proportion of individuals receiving benefits, then it appears likely that the probability of recovery will be low in this profile of patients. Conversely, LP1 have low severity, little functional impairment and a very low likelihood of receiving benefits, and therefore the likelihood of recovery would be predicted as high.

The regression analyses on patient characteristics associated with reliable change following treatment suggested that lower scores on the PHQ and GAD, higher functional impairment, receiving welfare benefits, psychotropic medication and being in a non-white ethnic group were associated with a lower likelihood of achieving this outcome. If we look at the distribution of these patient characteristics across the LPs it becomes more difficult to predict this outcome between the profiles. The profiles with the highest mean severity scores (LP7, LP8 and LP5) may be the most likely to show reliable change due to these higher scores, but it is unknown whether the higher proportion of individuals receiving welfare benefits and psychotropic medication, as well as individuals from non-white ethnic groups (especially in LP8) will contribute to a lower likelihood of this outcome being achieved for these profiles, as suggested by the regression analysis. It would therefore be of interest to explore whether differences in the likelihood of achieving this outcome, as well as clinical deterioration and drop out from treatment, differs between the latent profiles identified.

The previous latent class method analyses in CMHD populations found a six (Rosellini & Brown, 2014), and a seven class/profile solution (Unick, Snowden & Hastings, 2009) with their included populations and patient characteristics. The current analysis has found that there were eight profiles identified following the LPA, and the difference may be due to a combination of the variables used and the number of patients in the samples. This analysis has used a range of both clinical and demographic patient characteristics, whereas the previous research has concentrated more exclusively on clinical measures and the presence of diagnostic symptoms. One disadvantage of using diagnoses or clinician rated symptom measures only is that the patients themselves cannot complete them, which may limit the possibility of using shared decision making in the process of treatment selection. It is possible that using a wider range of patient characteristics has resulted in the current analysis identifying a larger number of profiles than previous studies. As this analysis identified an eight-profile solution in two independent samples, and then reproduced this structure in the full sample, one can be confident that the eight-profile solution is the best fit for the included dataset from the IAPT services.

Although one can be confident in the eight-profile solution in this study, it may also be of interest longer term to explore whether this same profile structure exists in datasets from similar services across the country. It is quite possible that different profile structures may exist within different datasets from IAPT services, which may reflect local demographics and communities. Rather than being a limitation of the method, it may instead allow the opportunity to develop locally tailored DSTs that best reflect the population entering treatment. This is beyond the scope of this project, but it may be of value to future research to explore the potential variation in profile structures.

As mentioned previously, LPA (as with all latent class methods) also provides output to calculate the posterior probabilities of profile membership which can be used to allocate patients to the identified profiles. This has huge potential value for the development of a DST, as the results of the LPA presented could be translated into algorithmic formula and provide the probability of profile membership to any patient assessed by the services (or at other services). An algorithm of these posterior probabilities could be built into the services' EPMS resulting in profile information provided to clinicians (as well as to patients), which could be used to inform treatment and care.

The analysis was limited by the patient characteristics available in the IAPT dataset, and characteristics identified in the systematic review (Chapter 3) but not included in the IAPT MDS could change the profile structure if included in future LPA. If these characteristics were considered of value to the services and collected in the future then they could be used to refine the profile structures identified in this analysis. The same could be said for some characteristics such as employment status that was not available in the current analysis but could be of interest to explore in future analysis.

Although the LPA has identified eight distinct profiles of patients attending the services, the potential value of these profiles to a DST to inform treatment selection decisions rests on whether they are able predict differential response to IAPT treatments. The next chapter presents a series of analyses that explore the variation in outcomes following psychological treatments, including recovery, reliable change, deterioration and dropout, between the identified latent profiles.

Chapter 6. Variation in treatment outcomes by latent profile.

Abstract

The identification of eight statistically distinct profiles of patients entering treatment at the IAPT services presented in Chapter 5 can inform both researchers and clinicians about the different types of patients, or casemix attending services. The identification of these profiles also has the potential to aid clinical decisions about treatment selection, and the aim of this chapter was to explore the variation in treatment outcomes between profiles and also whether the probability of outcomes varied between LI and HI interventions within these profiles. The chapter is split into three analyses; the first explored the variation in treatment outcomes between the profiles of patients across IAPT treatments. The second analysis supplemented these findings by exploring the probability of treatment outcomes within an additional dataset of n=4683 patients attending the same services at a later date to validate the findings of Analysis 1. The third analysis explored the differences in outcome between patients receiving LI only or HI only within in each profile, to identify profiles where different intensities of intervention may be associated with an increased probability of a positive outcome. The results revealed a large variation in outcomes between profiles in, for example, recovery (74% to 15%), deterioration rates (5% to 20%), and probability of dropout (17% to 40%), which was maintained in an independent validation sample of patients. Further analysis of differences between the two intensities of psychological intervention found that for some profiles one intensity was significantly more likely to result in better outcomes than the other. The findings from these analyses suggest that the latent profiles could be used to predict the likelihood of a range of outcomes following different intensities of psychological treatment. An algorithm which provides details on profile membership for new patients could then be used to optimise treatment selection decisions in services, and thereby personalise treatment for patients.

Introduction

The previous chapter identified eight statistically distinct profiles of patients who entered treatment in IAPT services, which can provide useful information about the characteristics of patients seeking treatment for CMHDs. This information on the casemix attending the services may be of value to clinicians, managers and commissioners in understanding who attends the services, but it could be of further value to clinical decision making if the profiles indicated which groups of patients are more likely to respond positively to treatment. Due to the different make-up of patient characteristics in the latent profiles, it is likely that there may be differences in the probability of recovery, change in symptoms including deterioration and dropout from treatment between the profiles. Providing information about the probability of treatment outcomes to clinicians might inform decisions about appropriate care.

Understanding the probability of treatment outcomes, for example the likelihood of recovery, for each LP would enable clinicians to have a quantifiable 'expected utility' of each treatment for different groups of patients referred to the services. As expected utility theory proposes that clinicians estimate probabilities about the expected outcome of treatment (Schoemaker, 1982), having this information readily available and built into a healthcare information system could be valuable, especially for less experienced clinicians who may be less able to predict the expected utility of treatments. As previous research suggests both clinicians and the general population have limited abilities when estimating probabilities (Elstein, 1999; Yamagishi, 1997), providing previously estimated probabilities of treatment outcome may be especially helpful.

In addition to providing information that may aid decisions on the selection of specific treatments, differences in outcomes between profiles could also show whether treatments may need to be tailored for particular profiles. For example, if certain profiles have a high incidence of patients dropping out of treatment then it may be of value to incorporate enhanced treatment-engagement strategies into the care plan to increase the likelihood of treatment completion and associated positive outcome. If a profile is at high risk of deterioration during treatment this may suggest that more frequent monitoring of symptoms is required throughout treatment. Having the probability of clinical deterioration available in advance may also be important considering the limited validity of clinicians' predictions of which patients may respond negatively to treatment (Hannan et al., 2005) and may therefore suggest alternative and potentially more effective treatments.

All IAPT services provide either Low Intensity (LI) interventions at Step 2 or High Intensity (HI) interventions at Step 3, therefore it was considered of interest to explore whether there are differences in outcomes within latent profiles between the different intensities of psychological intervention. The analysis of the overall population in Chapter 4 suggested a similar likelihood of recovery for patients receiving either LI (41.40%) or HI interventions (41.59%). However, there may be significant differences in the likelihood of recovery

between different intensities of psychological treatment for certain profiles of patients. Of course, it is also possible that the probability of outcomes were similar between the different intensities of intervention for profiles, for example it may be expected that for some patient profiles the likelihood of recovery is high and therefore little differences between intensity of intervention would be seen. Relaying this information to the clinician would allow them to suggest that LI interventions be considered initially, as these interventions are more cost- and time-effective. If results suggest that a profile has a higher likelihood of positive outcomes in one intervention over the other then the clinician could present this intervention as the better option to the patient. For example, if HI interventions suggest a significantly higher likelihood of recovery than LI interventions for a specific profile then it might be suggested that HI should be the first choice treatment, rather than initiate at LI and potentially have to step up to HI during the episode of care. Although stepping up is an appropriate part of the Stepped Care model, it seems a poor use of resource and time to start an intervention knowing that it is significantly less likely to be beneficial than an alternative.

The aims of this chapter were to explore the differences in the likelihood of outcomes following psychological treatment between the latent profiles, as well as differences within the profiles when different intensities of intervention were received. The chapter is split into three separate analyses. Analysis 1 explored the variation in outcomes between the latent profiles using the initial development sample that informed the original LPA in Chapter 5 (patients referred between September 2008 and March 2012). This analysis is then supplemented with Analysis 2 which compared the same outcomes using an additional dataset of patients attending the same IAPT services between April 2012 and August 2013. The goal of this analysis was to explore whether the probabilities of treatment outcome were maintained in an independent sample of similar patients, who received treatment at the same services. In the final part of the chapter, Analysis 3 focuses on the differences in outcome within each LP when different intensities of intervention were delivered to profile members, to investigate whether there is differential response to the interventions for patients within the same profile.

Analysis 1 – Variation in outcomes between Latent Profiles.

The aim of this analysis was to explore the variation in the likelihood of different treatment outcomes between the eight latent profiles (LPs) identified in Chapter 5.

Method.

Patients

Analysis 1 focused on the same dataset that was previously used for the LPA in Chapter 5, and the regression analysis presented in Chapter 4. This dataset is referred to as the 'development' dataset throughout this chapter. The inclusion criteria for this analysis were identical to those used for the regression analysis in Chapter 4, where included patients were required to have T2 data available and be classed as caseness on either the PHQ-9 or GAD-7 at T1. This resulted in the same n=10693 patients included from the dataset of the n=16636 patients who entered treatment and where used in LPA presented in Chapter 5 (a patient flow diagram is presented in figure 4.1, Chapter 4).

As the aim of this thesis was to create a DST that could be incorporated into clinical practice, an algorithm that could assign any patient to their latent profile was developed using the results of the LPA. An algorithm was built in Microsoft Excel 2013 that was able to take the nine patient characteristics used in the LPA and then provide the probably of membership to each of the eight profiles for each individual patient. The probability of profile membership was calculated using the posterior probabilities and priors from the output of the LPA (see equations presented in chapter 5). Each included patient was assigned to the profile to which they had the highest probability of membership as calculated by the algorithm, and this was their profile for the analyses of outcomes included in this chapter. The number of patients from both the full sample of n=16636 (taken into treatment) and the included sample (Caseness and T2 data) from each LP is presented in Table 6.1.

Table 6.1. Number of patients per LP between datasets.

	Taken into treatment		Caseness and T2 data		Patients excluded
Profile	n	%	n	%	
LP1	3,002	18%	663	6%	2,339
LP2	3,771	23%	2,950	28%	821
LP3	513	3%	66	1%	447
LP4	674	4%	508	5%	166
LP5	1,414	8%	1,106	10%	308
LP6	1,514	9%	1,079	10%	435
LP7	2,070	12%	1,572	15%	498
LP8	3,678	22%	2,749	26%	929
Total	16,636		10,693		5,943

LP1 show the largest loss of patients from the analysis (n=2339, 77.91%), mainly as these individuals were unlikely to meet caseness at T1 (due to the low mean PHQ and GAD scores for this profile). LP3 also have low mean severity scores, which also explains the large number of patients from this profile who were excluded from this analysis. As only 1% of the included sample (n=66 patients) were from LP3, the results of analysis from this profile should be interpreted with caution.

Outcomes

The five outcomes of interest in this analysis were the same five outcomes investigated as the dependent variables in Chapter 4 (recovery, reliable change, recovery or improvement, deterioration and dropout), that are defined and presented in Table 4.2 (Chapter 4). The percentage of patients who achieved each outcome was compared between the LPs. To explore the statistical difference in outcomes between LPs, logistic regression analysis was performed for the five outcomes to calculate the odds ratio of achieving each outcome. The logistic regression analyses were performed by entering only two profiles on each occasion as the independent variable (for example profile 1 vs profile 2) with the outcome (e.g. recovery) the dependent variable, to allow a direct comparison of the odds ratio of each outcome between profiles. Analysis was conducted using STATA 12 (StataCorp LP, 2011).

Results.

Recovery.

Table 6.2. The number and percentage of patients achieving recovery.

Profile	No	Yes	Total	%
LP1	170	493	663	74%
LP2	1,357	1,593	2,950	54%
LP3	18	48	66	73%
LP4	221	287	508	56%
LP5	776	330	1,106	30%
LP6	581	498	1,079	46%
LP7	1,340	232	1,572	15%
LP8	1,947	802	2,749	29%
Total	6,410	4,283	10,693	40%

The percentage of patients in recovery following treatment is presented in Table 6.2. The LPs with the highest percentage of patients in recovery were LP1 and LP3, as nearly 75% of patients in these profiles achieved recovery after receiving psychological treatment from the services. The latent profile with the lowest percentage of patients achieving recovery was LP7, where only 15% of patients achieved this outcomes following treatment.

To determine whether the percentage of patients who achieved these outcomes was significantly different between these LPs, logistic regression analysis was performed between each of the eight profiles and odds ratios (OR) are presented as a matrix in Table 6.3. Looking at the first row of the table, the comparison between LP1 and the other profiles is displayed. Compared to LP2, the odds of recovery for LP1 were 2.47 higher (OR=2.47, p<0.001). Compared to LP7 (with just 15% of patients in recovery), the odds of recovery for patients in LP1 were 16.75 higher (OR=16.75, p<0.001).

Table 6.3. Odds ratios for recovery between LPs.

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8
LP1	x	2.47**	1.09	2.23**	6.82**	3.38**	16.75**	7.04**
LP2	0.4**	x	0.44*	0.9	2.76**	1.34**	6.78**	2.85**
LP3	0.92	2.27*	x	2.05^	6.27**	3.11**	15.4**	6.47**
LP4	0.48**	1.11	0.49^	x	3.05**	1.52**	7.5**	3.15**
LP5	0.15**	0.36**	0.16**	0.33**	x	.5**	2.46**	1.03
LP6	0.29**	0.73**	0.32**	0.66**	2.02**	x	4.95**	2.08**
LP7	0.06**	0.15**	0.06**	0.13**	0.41**	0.2**	x	0.42**
LP8	0.14**	0.35**	0.15**	0.32**	0.97	0.48**	2.38**	x

^ p<0.05, * p<0.01, ** p<0.001

The table shows that the odds ratios between most of the profiles are significantly different at p<0.05, with the majority of odds ratios significant at p<0.001. The odds ratios between LP1 with LP3, LP2 with LP4, and LP5 with LP8 were not significantly different (p<0.05), which can be explained by the similar percentage of patients who achieved recovery in these groups (as displayed in Table 6.2). This suggests that although the patient characteristics of these profiles are statistically different, the constellation of characteristics for these profiles did not result in significantly different probabilities of recovery in these pairwise comparisons, although there may have be difference between these profile pairs for different treatment outcomes.

Reliable change

The percentage of patients who achieved reliable change, defined as a change of 5 or more points on the PHQ-9 and 4 or more points on the GAD-7, following interventions is presented in Table 6.4. The results for this positive outcome following treatment were very different to those for recovery presented above. The LP with the highest percentage of reliable change was LP8, where over 50% of the sample achieved reliable change following

treatment. The second highest was LP5, which was only slightly lower at 49% reliable change. The LP with the lowest percentage of patients achieving this outcome are LP3 with just 9% achieving reliable change. Patients in LP1, of whom 74% where in recovery at T1, were limited to a 17% probability of profile members showing reliable change on both measures following treatment. The low probability of reliable change in these two profiles compared with the high probability of recovery may be explained by the lower mean scores on both the GAD and PHQ for this profile of patients, when compared to the other profiles. Included cases in these analyses were patients who scored above caseness on either symptom measures, but it is possible that included LP1 and LP3 cases would have scored above caseness on one scale and not for the other. As a result, it would be more difficult for these patients to reduce for 5 PHQ points and 4 GAD points respectively, as scores on one scale would likely be just above caseness (but relatively low compared to other profiles), suggesting a floor effect.

As reported in Chapter 4, individuals with higher symptom severity scores at T1 were more likely to achieve reliable change at T2, which may in part be due to regression to the mean and patients scoring lower will likely have a floor effect where their symptoms cannot decrease by as much. However, the highest mean symptom scores were for LP7, yet this profile did not have the highest amount of reliable change and therefore the presence of certain characteristics of this profile may have reduced the likelihood of reliable change, for example the high likelihood of receiving welfare benefits. It may be that LP7 patients were not suited to psychological treatment, especially those delivered by IAPT services and therefore alternative services may be a better option to optimise outcomes for this group of patients.

Table 6.4. The number and percentage of patients achieving reliable change.

Profile	No	Yes	Total	%
LP1	551	112	663	17%
LP2	1,742	1,208	2,950	41%
LP3	60	6	66	9%
LP4	318	190	508	37%
LP5	567	539	1,106	49%
LP6	792	287	1,079	27%
LP7	973	599	1,572	38%
LP8	1,349	1400	2,749	51%
Total	6,352	4,341	10,693	41%

The number of patients who reliably changed was very similar for LP2, LP4 and LP7 at 37-41%. Patients in LP6 had a lower likelihood of reliable change, with only 27% achieving the outcome, suggesting this group may have important clinical differences compared to the other profiles. As the reliable change outcome requires a decrease on both measures, it is possible that individuals in LP6 were more likely to change on one symptom measure only.

As the mean GAD-7 score was considerably lower for LP6 patients than the pre-treatment PHQ-9 values, it is possible that there was a floor effect with the GAD-7 score reducing likelihood of reliable change on this measure. The odds ratios presented in Table 6.5 show that the odds of reliable change on both symptoms scales was significantly different between the majority of profiles. As with Table 6.2.2, there was no significant differences between odds of reliable change between LP1 and LP3, LP2 and LP4 as well as LP5 and LP8, most likely due similar probabilities of achieving this outcomes between profiles.

Table 6.5. Odds ratios for reliable change between LPs.

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8
LP1	x	0.29**	2.03	0.34**	0.21**	0.56**	0.33**	0.2**
LP2	3.41**	x	6.93**	1.16	0.73**	1.91**	1.13	0.67**
LP3	0.49	0.14**	x	0.17**	0.11**	0.28*	0.16**	0.1**
LP4	2.94**	0.86	5.97**	x	0.63**	1.65**	0.97	0.58**
LP5	4.68**	1.37**	9.51**	1.59**	x	2.62**	1.54**	0.92
LP6	1.78**	0.52**	3.62*	0.61**	0.38**	x	0.59**	0.35**
LP7	3.03**	0.89	6.16**	1.03	0.65**	1.7**	x	0.59**
LP8	5.11**	1.5**	10.38**	1.74**	1.09	2.86**	1.69**	x

^ $p<0.05$, * $p<0.01$, ** $p<0.001$

Recovery or improvement

The percentage of patients who either achieved the recovery or improvement outcome following treatment is presented in Table 6.6. The results appear very similar to those for the recovery outcome (Table 6.2). Both LP1 and LP3 contain the highest percentage of patients who achieved this positive outcome. LP7 are again the group with the lowest percentage of patients showing either reliable change or in recovery, although 7% more patients have achieved this positive outcome when compared to the recovery alone outcome (22% vs 15%).

Table 6.6. The number and percentage of patients achieving recovery or improvement.

Profile	No	Yes	Total	%
LP1	168	495	663	75%
LP2	1,143	1,807	2,950	61%
LP3	18	48	66	73%
LP4	199	309	508	61%
LP5	636	470	1,106	42%
LP6	537	542	1,079	50%
LP7	1,232	340	1,572	22%
LP8	1,649	1100	2,749	40%
Total	5,582	5,111	10,693	48%

The odds ratios presented in Table 6.7 suggest that the odds of recovery or improvement were significantly different between most of the groups. The non-significant odds ratios (p>=0.05) were for the comparisons between LP1 and LP3, LP2 and LP3, LP2 and LP4, LP3 and LP4, as well as the comparison between LP5 and LP8. There were slightly less statistically significant differences between profiles achieving this outcome, which may be due to the use of multiple, different criteria in the calculation of this outcome.

Table 6.7. Odds ratios for recovery or improvement between LPs.

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8
LP1	x	1.86**	1.1	1.9**	3.99**	2.92**	10.68**	4.42**
LP2	0.54**	x	0.59	1.02	2.14**	1.57**	5.73**	2.37**
LP3	0.91	1.69	x	1.72	3.61**	2.64*	9.66**	4**
LP4	0.53**	0.98	0.58	x	2.1**	1.54**	5.63**	2.33**
LP5	0.25**	0.47**	0.28**	0.48**	x	0.73**	2.68**	1.11
LP6	0.34**	0.64**	0.38*	0.65**	1.37**	x	3.66**	1.51**
LP7	0.09**	0.17**	0.1**	0.18**	0.37**	0.27**	x	0.41**
LP8	0.23**	0.42**	0.25**	0.43**	0.9	0.66**	2.42**	x

^ p<0.05, * p<0.01, ** p<0.001

Deterioration

Table 6.8. The number and percentage of patients showing deterioration.

Profile	No	Yes	Total	%
LP1	590	73	663	11%
LP2	2,694	256	2,950	9%
LP3	55	11	66	17%
LP4	455	53	508	10%
LP5	1032	74	1,106	7%
LP6	868	211	1,079	20%
LP7	1,493	79	1,572	5%
LP8	2,575	174	2,749	6%
Total	9,762	931	10,693	9%

The percentage of patients who reported clinical deterioration on the either symptom scale following treatment is presented in Table 6.8. The LP with the highest number of patients who showed clinical deterioration was LP6, indicating that for one fifth of patients in this profile there was a negative impact on either depression or anxiety symptoms following treatment. This percentage is high compared to studies of deterioration in psychotherapy suggest around 5-10% deterioration in patients (Chapter 1), and might suggest that there may be important characteristics of this profile that increased the risk of negative outcomes. LP3 had the second highest incidence with 17% of patients showing clinical deterioration.

LP7 were the least likely to show deterioration with just 5% of patients displaying a negative outcome, followed by LP8 with 6%.

Logistic regression analyses showed that the differences in odds of deterioration were significant between a number of the LPs, and is presented in Table 6.9. LP6 showed significantly increased odds of deterioration when compared to all LPs except LP3. For example, the odds of deterioration were over 2.56 higher for patients in LP6 compared to those from LP2 (OR=2.56, p<0.001) and 4.59 higher when compared with LP7 (OR=4.59, p<0.001). LP1, LP2, LP3 and LP4 showed no significant differences in rates of deterioration when directly compared, except the comparison of LP3 with LP2 (OR=2.1, p<0.05). No significant differences were found for the odds of deterioration between members of LP5 when compared to either LP7 or LP8, and the odds did not significantly differ between members of LP7 and LP8

Table 6.9. Odds ratios for deterioration between LPs.

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8
LP1	x	1.3	0.62	1.06	1.73*	0.51**	2.34**	1.83**
LP2	0.77	x	0.48^	0.82	1.33^	0.39**	1.8**	1.41*
LP3	1.62	2.1^	x	1.72	2.79*	0.82	3.78**	2.96*
LP4	0.94	1.23	0.58	x	1.62^	0.48**	2.2**	1.72*
LP5	0.58*	0.75^	0.36*	0.62^	x	0.29**	1.36	1.06
LP6	1.96**	2.56**	1.22	2.09**	3.39**	x	4.59**	3.6**
LP7	0.43**	0.56**	0.26**	0.45**	0.74	0.22**	x	0.78
LP8	0.55**	0.71*	0.34*	0.58*	0.94	0.28**	1.28	x

^ p<0.05, * p<0.01, ** p<0.001

Dropout

The percentage of patients dropping out of treatment is presented in Table 6.10. Both LP7 and LP8 reported the highest number of patients leaving treatment early, with 40% of patients dropping out of treatment in these profiles. LP3 had the lowest percentage of dropout (17%), followed by LP4 (20%) and LP1 (26%).

Table 6.10. The number and percentage of patients dropping out of treatment.

Profile	No	Yes	Total	%
LP1	488	175	663	26%
LP2	2,046	904	2,950	31%
LP3	55	11	66	17%
LP4	405	103	508	20%
LP5	789	317	1,106	29%
LP6	728	351	1,079	33%
LP7	948	624	1,572	40%
LP8	1,655	1094	2,749	40%
Total	7,114	3,579	10,693	33%

The differences in the odds of dropout are presented in Table 6.11. The odds of dropout were significantly higher in LP7 and LP8 when compared to other profiles, as shown by significant odds ratios between these two profiles and the other groups of patients, except when compared these profiles were directly compared. LP3 and LP4 indicated the lowest probability of dropout which may be linked to the mean age of these two profiles which is notably higher than the other profiles. This may provide further evidence for the finding in Chapter 3 that older age was linked to a reduction in dropout highlighted in previously conducted systematic reviews of outcome in CMHDs. However, there may also be other important characteristics of LP3 and LP4 which contribute to the significantly lower odds of dropout compared to most other profiles.

Table 6.11. Odds ratios for dropout between LPs.

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8
LP1	x	0.81^	1.79	1.41^	0.89	0.74*	0.54**	0.54**
LP2	1.23^	x	2.21^	1.74**	1.1	0.92	0.67**	0.67**
LP3	0.56	0.45^	x	0.79	0.5^	0.41*	0.3**	0.3**
LP4	0.71^	0.58**	1.27	x	0.63**	0.53**	0.39**	0.38**
LP5	1.12	0.91	2.01^	1.58**	x	0.83	0.61**	0.61**
LP6	1.34*	1.09	2.41*	1.9**	1.2	x	0.73**	0.73**
LP7	1.84**	1.49**	3.29**	2.59**	1.64**	1.37**	x	1
LP8	1.84**	1.5**	3.31**	2.6**	1.65**	1.37**	1	x

^ $p<0.05$, * $p<0.01$, ** $p<0.001$

Discussion

The results of Analysis 1 suggest that there is significant variation in the probability of treatment outcomes between the LPs identified in the LPA. Table 6.12 displays the percentage of patients achieving each outcome across the LPs. Both visual inspection of the probability of specific treatment outcomes as well as results of the logistic regression analyses suggest there are important differences in outcomes between the majority of latent profiles. These differences have the potential to inform clinicians on the expected utility of psychological treatments per profile and may be of value to consider when deciding on appropriate treatments from the services.

The two profiles with the highest probability of recovery following treatment were LP1 and LP3, and with almost a 75% chance of recovery, it would appear that the services offered a very high likelihood of a positive outcome for this profile of patients. In contrast, the results for LP7 indicate that the probability of recovery is limited to just 15% of patients, indicating that patients were very unlikely to achieve this outcome. Providing a clinician with this information may help them shape treatment in order to maximise the likelihood of recovery. LP7 may be seen as a group for whom IAPT services do not offer the most appropriate treatment and potentially other treatment options should be considered.

The results of the regression analysis in Chapter 4 found that higher baseline severity increased the odds of reliable change. However, the LP with the highest baseline severity scores (LP7) was not the profile with the highest likelihood of reliable change, and instead three other profiles had a higher probability of achieving this outcome. If we consider the patient characteristics of LP7, which include a high probability of receiving welfare benefits, medication and caseness for phobia, and reference these against the results of the regression analyses in Chapter 4, then the results suggest the interaction between these variables may be more powerful at predicting reliable change than severity scores alone. Although much research has focused on the role of severity in predicting treatment response (Blom et al., 2007; Fournier et al., 2010), results presented in this chapter suggest that other patient characteristics such as demographic information may have an important role in predicting outcomes, despite previous evidence for these characteristics identified in the meta-review presented in Chapter 3.

The probability of clinical deterioration also showed significant differences between profiles. Although the differences were not as large as those presented for recovery, considering the negative impact of clinical deterioration on a patient, the differences between profiles presented here are no less important. The results indicated that members of LP6 had the highest probability of clinical deterioration following treatment. The probability was much higher than all other profiles, with the exception of LP3 and it would be of interest to explore whether there were differences between the intensities of IAPT intervention and probability of deterioration for this profile. It could be speculated that the high incidence of deterioration in LP3 may be linked to the older age of this profile, chronicity of the disorder, adverse social circumstances or that physical comorbidities that could have contributed to poorer outcomes for some members of this profile. No information on physical comorbidities was included in the dataset, but future research might consider collecting this information.

Table 6.12. Comparison of treatment outcomes across the latent profiles.

	% of sample	Recovery			Reliable change			Recovery or improvement			Deterioration			Treatment dropout		
		Yes	Total	%	Yes	Total	%	Yes	Total	%	Yes	Total	%	Yes	Total	%
LP1	**6%**	493	663	74%	112	663	17%	495	663	75%	73	663	11%	175	663	26%
LP2	**28%**	1593	2950	54%	1208	2950	41%	1807	2950	61%	256	2950	9%	904	2950	31%
LP3	**1%**	48	66	73%	6	66	9%	48	66	73%	11	66	17%	11	66	17%
LP4	**5%**	287	508	56%	190	508	37%	309	508	61%	53	508	10%	103	508	20%
LP5	**10%**	330	1106	30%	539	1106	49%	470	1106	42%	74	1106	7%	317	1106	29%
LP6	**10%**	498	1079	46%	287	1079	27%	542	1079	50%	211	1079	20%	351	1079	33%
LP7	**15%**	232	1572	15%	599	1572	38%	340	1572	22%	79	1572	5%	624	1572	40%
LP8	**26%**	802	2749	29%	1400	2749	51%	1100	2749	40%	174	2749	6%	1094	2749	40%
Total	**100%**	4283	10693	40%	4341	10693	41%	5111	10693	48%	931	10693	9%	3579	10693	33%

Although the probability of treatment outcomes was statistically different between the majority of profiles, for a few direct comparisons there was no significant difference. For example, comparisons between LP1 and LP3, LP2 and LP4 as well as LP5 and LP8 for recovery, reliable change and deterioration indicated no significant differences. This finding suggests that although the constellation of patient characteristics that comprise these profiles may be statistically different, the likelihood of these outcomes for these stratified groups of patients is very similar. However, it is possible that outcomes vary between these profiles when different intensities of psychological intervention are received, but also the probably of other outcomes (e.g. dropout) was significantly different between these comparisons.

The findings presented in this analysis suggest that there were a number of important differences in the likelihood of treatment outcomes between the profiles when receiving treatment at the IAPT services. However, it was also important to explore whether the likelihood of treatment outcomes was consistent over time, which was the aim of the next set of analyses.

Analysis 2 – Treatment outcomes in the validation sample.

Analysis 2 compared the results of Analysis 1 to the probability of treatment outcomes in an additional sample of patients attending the same IAPT services at a later date. The aim of this analysis was to validate whether the probabilities of treatment outcome were consistent over time and within an independent sample of patients.

Method

The sample used for this analysis was a dataset of patients referred to the same IAPT services between April 2012 and August 2013, and is referred to as the 'validation sample' throughout this chapter.

The same inclusion criteria as used in Analysis 1 was employed for this analysis and resulted in n=4683 patients included. Patients were assigned to the latent profile to which they had the highest probability of profile membership using the algorithm created and described for Analysis 1.

The same five outcomes of interest in Analysis 1 were also used in this analysis, and the percentage of patients achieving each outcome was calculated in the same method as the previous analysis. The results for the validation sample were presented alongside the results of the development sample to allow comparison.

Results

After the probability of group membership had been calculated for each of the validation sample patients, the distribution of the profiles was compared between the two samples. Table 6.13 displays the frequency and percentage of patients in each LP for the development sample and the validation sample. The table suggests that the distribution of patients in each LP was very consistent over time, with limited differences in percentage between the development and validation sample.

Table 6.13. LP distribution: development and validation samples.

	Development sample		Validation Sample	
Profile	**n**	**%**	**n**	**%**
LP1	663	6%	288	6%
LP2	2,950	28%	1,248	27%
LP3	66	1%	35	1%
LP4	508	5%	261	6%
LP5	1,106	10%	495	11%
LP6	1,079	10%	488	10%
LP7	1,572	15%	709	15%
LP8	2,749	26%	1,159	25%
Total	10,693		4,683	

Recovery

The number and percentage of patients in recovery in both the validation and development sample is presented in Table 6.14. The percentage of patients in recovery following treatment for each profile was consistent for the majority of LPs, with a variation of just one or two percentage points. The profile with the largest difference between datasets was LP2, where 54% of patients were in recovery in the development sample but this dropped to 49% in the validation sample. A 3% drop in the likelihood of recovery was indicated in LP8 between the development and validation samples. Overall the differences between the likelihood of recovery was similar between the development and validations samples for each profile.

Table 6.14. Development vs validation sample: recovery

	Validation sample			Development sample		
Profile	**In recovery**	**Total**	**%**	**In recovery**	**Total**	**%**
LP1	218	288	76%	493	663	74%
LP2	613	1,248	49%	1,593	2,950	54%
LP3	26	35	74%	48	66	73%
LP4	152	261	58%	287	508	56%
LP5	140	495	28%	330	1,106	30%
LP6	221	488	45%	498	1,079	46%
LP7	95	709	13%	232	1,572	15%
LP8	303	1,159	26%	802	2,749	29%
Total	1,768	4,683	38%	4,283	10,693	40%

Reliable change

The percentage of patients who achieved reliable change between the two samples is slightly more variable for some profiles between the datasets and presented in Table 6.15. Overall, the number of patients who achieved reliable change following treatment was lower in the validation sample than the development sample by 4% (37% vs 41%). This pattern appears reasonably consistent across the LPs, and the largest difference was for LP5 where 41% of the validation showed reliable change compared to 49% in the development sample. LP3 also showed a 6% difference in probability of reliable change between datasets, as only 3% of LP3 members from the validation sample achieved reliable changed, compared to 9% of the development sample. Again, it should be noted that the number of LP3 patients in the validation sample was very low, with only 35 LP3 patients included, and therefore these results should be viewed with caution and potentially replicated in a large sample with higher representation of LP3.

Table 6.15. Development vs validation sample: reliable change

	Validation sample			Development sample		
Profile	**Reliable change**	**Total**	**%**	**Reliable change**	**Total**	**%**
LP1	41	288	14%	112	663	17%
LP2	457	1,248	37%	1,208	2,950	41%
LP3	1	35	3%	6	66	9%
LP4	94	261	36%	190	508	37%
LP5	202	495	41%	539	1,106	49%
LP6	138	488	28%	287	1,079	27%
LP7	269	709	38%	599	1,572	38%
LP8	546	1,159	47%	1400	2,749	51%
Total	1,748	4,683	37%	4,341	10,693	41%

Recovery or improvement

The number and percentage of patients who were in recovery or improvement in both samples is presented in Table 6.16. The results were very similar to those for recovery presented in Table 6.14 and indicate that the percentage of patients who achieved this outcome was largely consistent between the independent samples. The two profiles showing more than 2% variation are LP2 (4%) and LP5 (7%) with both profiles showing less recovery or improvement in the validation sample than development.

Table 6.16. Development vs validation sample: recovery or improvement

	Validation sample			Development sample		
Profile	Recovery or improvement	Total	%	Recovery or improvement	Total	%
LP1	220	288	76%	495	663	75%
LP2	715	1,248	57%	1,807	2,950	61%
LP3	26	35	74%	48	66	73%
LP4	160	261	61%	309	508	61%
LP5	174	495	35%	470	1,106	42%
LP6	240	488	49%	542	1,079	50%
LP7	150	709	21%	340	1,572	22%
LP8	438	1,159	38%	1100	2,749	40%
Total	2,123	4,683	45%	5,111	10,693	48%

Deterioration

The percentage of patients who reported clinical deterioration following treatment for both samples is presented in Table 6.17. The overall percentage of deterioration was the same for both datasets (9%), and generally the variation between the samples for individual profiles was very low. LP6 had the highest probability of deterioration across both samples, as 20% of patients in this profile deteriorated in both independent samples. The main difference between samples is for LP3 where 17% reported deterioration in the development sample, but this was reduced to 9% in the validation sample. This may suggest improvements by the service in treating this profile of patients, but the very low number of included patients from this sample suggests caution with any interpretation of findings for this group.

Table 6.17. Development vs validation sample: deterioration

	Validation sample			Development sample		
Profile	Deterioration	Total	%	Deterioration	Total	%
LP1	27	288	9%	73	663	11%
LP2	105	1,248	8%	256	2,950	9%
LP3	3	35	9%	11	66	17%
LP4	23	261	9%	53	508	10%
LP5	35	495	7%	74	1,106	7%
LP6	96	488	20%	211	1,079	20%
LP7	52	709	7%	79	1,572	5%
LP8	87	1,159	8%	174	2,749	6%
Total	428	4,683	9%	931	10,693	9%

Treatment dropout

The largest variation in outcome between the development and the validation sample is for dropout of treatment (presented in Table 6.18). Overall there was more dropout in the validation sample than for the development sample (38% vs 31%). Exploration of the individual LPs shows that the increase in percentage dropout was low for LP1, LP4, LP6 and LP7 (less than 3% increase for these profiles) but much higher for the other four groups of patients. It is unclear whether this increase in dropout is due to a change in practice at the service, or potentially the recording of dropout on the system, but it appears that over time the number of treatment dropout has increased. The findings presented here suggest that this increase in dropout is focused more on certain profiles, and potentially these groups of patients could be considered for treatment engagement interventions. It may also be possible that some individuals were allocated to more inappropriate treatments, hence why patients left treatment early, and an exploration of difference in dropout between intensities of intervention may be of value.

Table 6.18. Development vs validation sample: dropout

	Validation sample			Development sample		
Profile	**Dropped Out**	**Total**	**%**	**Dropped Out**	**Total**	**%**
LP1	77	288	27%	175	663	26%
LP2	441	1,248	35%	904	2,950	31%
LP3	9	35	26%	11	66	17%
LP4	54	261	21%	103	508	20%
LP5	185	495	37%	317	1,106	29%
LP6	172	488	35%	351	1,079	33%
LP7	302	709	43%	624	1,572	40%
LP8	534	1,159	46%	1094	2,749	40%
Total	1,774	4,683	38%	3,579	10,693	33%

Discussion

The findings of analysis 2 indicate that the variation in outcomes between groups was largely replicated in an independent sample of patients attending the services. For most profiles, the percentage of patients who achieved each outcome was very similar in the two independent samples and suggests that the latent profiles can predict the likelihood of treatment outcomes across profiles and time.

The probability of recovery following treatment was consistent across profiles over time and suggests that LPs may be useful in predicting the likelihood of a positive outcome. As recovery is the main key performance indicator (KPI) for IAPT services nationally and a likely goal of many patients, a robust method of predicting the likelihood of recovery based on patient characteristics may be of real value to patients, services and clinicians. For commissioners, the differences in overall recovery rates for a service could be further explored by analysing recovery in relation to the distribution of profiles entering treatment locally. Services with higher number of patients from profiles with a low probability of recovery (e.g. LP7) will likely have lower overall recovery rates. This information may be equally valuable comparisons of reliable change, deterioration and dropout between services.

The likelihood of dropping out of treatment showed the largest variation between the validation and development samples, although this variation was still within a few percentage points across the profiles. It is possible that changes to the services over time may have contributed to an increased rate of dropout in the most recent time period. IAPT services are required to achieve an access target of at least 15% of adults with relevant disorders (HSCIC, 2015), and therefore there may have been increasing pressure to take

patients into treatment, even if it was not necessarily the most appropriate treatment. Alternatively, other service-related factors such as a decrease in staffing levels may have had an impact on dropout.

Although the results of Analyses 1 and 2 suggest consistent variation in treatments across the profiles, it would be of further value for clinical decisions on treatment selection if differences between intensities of IAPT treatments were considered. This is the focus of the next analysis.

Analysis 3 – Variation in outcomes between interventions.

The results of Analysis 1 and Analysis 2 indicate that there is significant variation in outcomes between the latent profiles identified, which is largely consistent between independent samples of patients attending the service. This information alone may be important to clinicians, as it may help them predict how likely a patient is to achieve a positive outcome from treatment, or terminate early, and therefore may inform decisions on treatment planning. However, it does not necessarily help decide on which type of treatment would be most appropriate. Therefore, exploring whether the outcomes vary between intensities of intervention within latent profiles could provide further information to inform treatment choice. The following analysis focused on the differences in treatment outcomes between low and high intensity interventions for each latent profile.

Methods

For this analysis, it was decided to combine the development and validation samples to create one single dataset of n=15376 patients who were caseness at T1 and included T2 data. Combining datasets was performed to maximise the number of patients included in the within profile analysis, as some profiles (e.g. LP3 and LP4) had low representation within the dataset. IAPT services essentially provide Low Intensity (LI) and High Intensity (HI) psychological interventions and these were the focus of the current analysis. As the aim of this analysis was to compare individuals receiving specific intensities of interventions, any patient who received more than one intensity of intervention was excluded, for example patients stepped up or stepped down during treatment.

The number of patients who received each intensity of treatment, or who were stepped up/down is displayed in Table 6.19. Approaching half of patients received LI interventions only (45.04%) compared to 28.06% who received only HI treatments. Over 15% of patients were stepped up during contact with the service. The number of individuals who were stepped down, or who attended either Step 1 or Step 4 treatments was very low, as these interventions are most commonly delivered by GPs (Step 1) or inpatient facilities (Step 4).

Information on intensity of treatment was missing for n=1451 (9.44%) patients. A total of n=11240 patients receiving LI or HI interventions only were the focus of this analysis.

Table 6.19. Number of patients by intensity of intervention.

Step of intervention	Number of patients	Percentage of sample
Stepped Down	263	1.71%
LI	6,926	45.04%
HI	4,314	28.06%
Stepped UP	2,367	15.39%
Step 1 or Step 4	55	0.36%
Not known/recorded	1,451	9.44%
Total	15,376	

As the current dataset was collected as part of routine service evaluation, the allocation of each patient to either LI or HI interventions was decided by clinicians in consultation with the patient and potentially with the clinicians' supervisor(s). The decision to allocate to either intervention was made following a clinical assessment and would have considered baseline severity of depression and anxiety as well as other factors. It was possible that there may have been differences between baseline symptom scores between patients allocated to LI instead of HI interventions, as it might be expected that more severe patients would be allocated to HI treatment. In order to test this assumption, independent samples t-tests were performed on the mean PHQ and GAD scores at T1 for patients receiving LI interventions and those receiving HI intervention in each LP.

The results are presented in Table 6.20 and show that there were significant differences in baseline severity in LP1 and LP2 respectively, but not for the other profiles. Individuals in LP1 who received HI interventions had a significantly higher mean PHQ score at T1 than those receiving LI interventions, however the mean GAD score was significantly lower for LP1 patients receiving HI interventions compared to those receiving LI. The opposite was found for LP2, as significantly lower PHQ and higher GAD scores at T1 were found in the HI group, compared to individuals from LP2 who received LI interventions. No other LPs showed significant differences in mean symptom severity scores at T1 between patients receiving LI and those receiving HI interventions. Of course, a range of variables other than scores of a symptom severity scale influence the presentation and course of a mental disorder, and may have influenced the decision to allocate a patient to treatment.

Table 6.20. Comparison of mean symptom scores by intervention

	Initial PHQ9 score					Initial GAD7 score				
	Low Intensity		High Intensity		t-test p-value	Low Intensity		High Intensity		t-test p-value
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
LP1	6.71	3.12	7.33	3.24	**0.01**	8.22	2.16	7.74	2.45	**0.01**
LP2	11.38	3.11	10.86	3.31	**0.00**	12.53	2.95	12.85	2.93	**0.01**
LP3	7.83	3.80	8.41	3.63	0.47	5.87	3.39	5.00	3.22	0.23
LP4	10.59	3.45	11.11	3.61	0.07	11.16	3.15	10.91	3.55	0.36
LP5	17.64	3.36	17.95	3.45	0.13	15.92	2.70	15.81	2.95	0.51
LP6	13.92	3.05	13.62	3.27	0.10	8.13	2.65	8.30	2.70	0.27
LP7	22.73	2.74	22.84	2.93	0.43	18.28	2.52	18.33	2.62	0.69
LP8	18.69	2.98	18.73	3.38	0.76	16.29	2.73	16.47	2.96	0.11

The percentage of patients achieving each of the five outcomes was compared within each LP for patients who received LI interventions only and HI interventions only. Logistic regression analysis was conducted to explore whether the difference in percentage of patients was statistically significant between interventions within each profile, with intensity of treatment included as the independent variable. Odds ratios and the associated p-value were presented.

Results

The distribution of patients receiving LI interventions only and those receiving HI interventions only is presented in Table 6.21. With the exception of LP7, LI was the dominant choice of intervention for the profiles, although there was variation between LPs. Nearly 70% of individuals in LP1 and LP2 were allocated to LI interventions. The LP with the third highest allocation of LI interventions was LP8 which is interesting considering the high mean severity scores for this group of patients (means of 18.85 and 16.43 on the PHQ and GAD respectively at T1).

Table 6.21. Distribution of LI and HI patients by Latent Profile

	Total	Receiving LI		Receiving HI	
		n	%	n	%
LP1	742	511	68.87%	231	31.13%
LP2	3109	2192	70.50%	917	29.50%
LP3	86	47	54.65%	39	45.35%
LP4	608	372	61.18%	236	38.82%
LP5	1181	739	62.57%	442	37.43%
LP6	1200	675	56.25%	525	43.75%
LP7	1634	675	41.31%	959	58.69%
LP8	2680	1715	63.99%	965	36.01%
Total	11240	6926	61.62%	4314	38.38%

Recovery

The percentage of patients who were in recovery after receiving either LI or HI interventions only for each LP is presented in Table 6.22. This table also presents the Odds Ratio (OR), 95% confidence intervals and *p*-value for recovery between LI and HI. The percentage of LP1 patients in recovery following LI interventions was 75% compared 79% after HI interventions, suggesting a slightly increased chance of recovery if more intensive treatment was delivered for this profile of patients. Logistic regression analysis indicated that the increase in odds for HI interventions was not statistically significant (*p*=0.227) over LI. The results for LP2 suggest that there was a 7% increase in the likelihood of recovery following HI interventions compared to LI (59% vs 52%), and this difference was statistically significant (OR=1.32, *p*=0.001). This indicated a significant benefit of HI interventions over LI for LP2, although the probability of recovery was over 50% following either type of intervention.

Table 6.22. LI vs HI comparison by LP: recovery

Profile	Low Intensity		High Intensity		OR	95% CIs	p-value
	Total cases in LP	% recovery	Total cases in LP	% recovery			
LP1	511	75%	231	79%	1.26	0.87, 1.84	0.227
LP2	2,192	52%	917	59%	1.32*	1.13, 1.54	0.001
LP3	47	72%	39	72%	0.97	0.38, 2.51	0.955
LP4	372	58%	236	56%	0.89	0.64, 1.24	0.493
LP5	739	30%	442	31%	1.05	0.81, 1.35	0.718
LP6	675	45%	525	53%	1.39*	1.11, 1.75	0.004
LP7	675	11%	959	17%	1.66*	1.24, 2.22	0.001
LP8	1,715	29%	965	32%	1.11	0.94, 1.32	0.215
Total	6,926	42%	4,314	41%	0.98	0.9, 1.06	0.551

The odds of recovery increased for patients who received HI interventions instead of LI in all LPs except LP3 and LP4, where results suggested a slight increase in the probability of recovery when LI interventions where received, although this difference was not statistically significant (*p*>=0.05). The reason for the decrease in odds of recovery for LP3 and LP4 when receiving HI interventions compared to LI may be due to treatment preferences for these groups of patients. However, the very low number of patients from LP3 compared to all other profiles suggests that these results should be interpreted with caution and that further analysis with more LP3 patients may be of value. Although HI interventions increased the odds of recovery for most LPs, the difference was only statistically significant for LP6 (OR= 1.39, *p*=0.004) and LP7 (OR= 1.66, *p*=0.001) in addition to LP2.

Reliable change

The percentage of patients who achieved reliable change on both symptom measures is presented in Table 6.23. Overall the results suggest that the odds of achieving reliable change increased when patients received HI rather than LI interventions for all LPs except LP3 and LP4, although this difference is not significantly different for any LP except LP1. The percentage of patients showing reliable change increased by 11% (13% vs 24%) when HI interventions were received by members of LP1, suggesting the likelihood of achieving reliable change more than doubled if a HI intervention was received instead of LI (OR=2.18, p<0.001). An increase in reliable change would likely be expected when HI interventions were received instead of LI, mainly due to the increase number of sessions delivered. However, as this difference was not statistically significant in any LPs other than LP1, it suggests that the number of sessions alone does not significantly predict reliable change.

Table 6.23. LI vs HI comparison by LP: reliable change

Profile	Low Intensity		High Intensity		OR	95% CIs	p-value
	Total cases in LP	% reliable change	Total cases in LP	% reliable change			
LP1	511	13%	231	24%	2.18*	1.46, 3.26	<0.000
LP2	2192	41%	917	43%	1.08	0.92, 1.26	0.331
LP3	47	9%	39	5%	0.58	0.1, 3.36	0.544
LP4	372	39%	236	35%	0.85	0.61, 1.19	0.345
LP5	739	48%	442	51%	1.12	0.89, 1.42	0.333
LP6	675	28%	525	30%	1.09	0.84, 1.4	0.525
LP7	675	37%	959	40%	1.13	0.93, 1.39	0.225
LP8	1715	51%	965	53%	1.1	0.94, 1.29	0.227
Total	6926	40%	4314	42%	1.09	1.01, 1.17	0.037

Recovery or improvement

The results for recovery or improvement outcome were very similar to those for the recovery alone outcome and are presented in Table 6.24. The odds of achieving this positive outcome increased when patients received HI interventions in all profiles except LP3 and LP4 (although this was not statistically different). Statistically higher probabilities of recovery or improvement were found for LP2 (OR=1.28, p=0.002), LP6 (OR=1.29, p=0.027), LP7 (OR=1.33, p=0.022) and LP8 (OR=1.19, p=0.036) when individuals from these profiles received HI interventions instead of LI.

Table 6.24. LI vs HI comparison by LP: recovery or improvement

Profile	Low Intensity		High Intensity		OR	95% CIs	p-value
	Total cases in LP	% recovery or improvement	Total cases in LP	% recovery or improvement			
LP1	511	75%	231	79%	1.25	0.86, 1.82	0.249
LP2	2192	60%	917	66%	1.28*	1.09, 1.51	0.002
LP3	47	72%	39	72%	0.97	0.38, 2.51	0.955
LP4	372	64%	236	59%	0.82	0.58, 1.14	0.234
LP5	739	41%	442	44%	1.11	0.87, 1.4	0.403
LP6	675	49%	525	56%	1.29*	1.03, 1.63	0.027
LP7	675	19%	959	24%	1.33*	1.04, 1.69	0.022
LP8	1715	39%	965	44%	1.19*	1.01, 1.4	0.036
Total	6926	49%	4314	49%	0.97	0.9, 1.04	0.382

Deterioration

A comparison of the percentage of patients deteriorating following treatment is presented in Table 6.25. Typically, the odds of deteriorating from treatment increased when individuals received HI interventions instead of LI, which may in part be explained the higher complexity expected for patients allocated to HI treatments. However, the only statistically significant increase in odds of deterioration was for LP7 (OR=1.73, p=0.023). LP5 was the sole profile where less deterioration was indicated following HI treatments, although the difference was not statistically significant (p=0.07). A large increase in the odds of deterioration was suggested within LP3 when HI interventions were received compared to LI (21% vs 6%), however this difference was not statistically significant, most likely due to the low number of patients from this profile entering treatment at the services.

Table 6.25. LI vs HI comparison by LP: deterioration

Profile	Low Intensity		High Intensity		OR	95% CIs	p-value
	Total cases in LP	% Deterioration	Total cases in LP	% Deterioration			
LP1	511	10%	231	10%	1.09	0.65, 1.83	0.735
LP2	2,192	7%	917	8%	1.11	0.84, 1.49	0.458
LP3	47	6%	39	21%	3.78	0.93, 15.41	0.063
LP4	372	7%	236	11%	1.51	0.86, 2.68	0.154
LP5	739	7%	442	5%	0.61	0.36, 1.04	0.07
LP6	675	17%	525	21%	1.26	0.94, 1.69	0.116
LP7	675	4%	959	6%	1.73*	1.08, 2.76	0.023
LP8	1,715	6%	965	7%	1.2	0.87, 1.67	0.261
Total	6,926	8%	4,314	9%	1.19*	1.04, 1.36	0.014

Dropout

The results of the comparison between LI and HI interventions for percentage of patients dropping out of treatment is presented in Table 6.26. Overall, there was significantly less dropout in HI interventions compared to LI interventions for the full sample (OR=0.73, p<0.001). LP3 was the only profile where the opposite was indicated, although the difference in the likelihood of dropout was not statistically significant (p=0.496). There was a significant decreased likelihood of dropout during treatment for LP2 (OR=0.59, p<0.001), LP4 (OR=0.44, p=0.001), LP6 (OR=0.66, p=0.002), LP7 (OR=0.74, p=0.004), and LP8 (OR=0.75, p=0.001) when patients are in HI interventions rather than LI.

Table 6.26. LI vs HI comparison by LP: dropout

Profile	LI interventions		HI Interventions		OR	95% CIs	p-value
	Total cases in LP	% Dropout	Total cases in LP	% Dropout			
LP1	511	24%	231	21%	0.85	0.58, 1.23	0.383
LP2	2192	31%	917	21%	0.59*	0.49, 0.71	<0.000
LP3	47	15%	39	21%	1.47	0.48, 4.51	0.496
LP4	372	22%	236	11%	0.44*	0.28, 0.72	0.001
LP5	739	28%	442	24%	0.8	0.61, 1.05	0.109
LP6	675	32%	525	23%	0.66*	0.51, 0.85	0.002
LP7	675	39%	959	32%	0.74*	0.6, 0.91	0.004
LP8	1715	39%	965	32%	0.75*	0.64, 0.89	0.001
Total	6926	32%	4314	26%	0.73*	0.67, 0.8	<0.000

These findings suggest that for the majority of patients, there is likely to be a significantly higher incidence of dropout from LI interventions. One reason for this may be that patients receiving self-help treatments may not feel they are supportive enough, or prefer more contact with a clinician, and therefore leave treatment. It is also possible that individuals in LI who are completing self-help materials in their own time at home may feel they are able to manage this this treatment themselves and so withdraw from the service. Although their symptoms may have improved, as they did not make any further contact with the services they were coded as dropping out of treatment, although this single code does not capture reasons for leaving treatment.

Table 6.27. Treatment outcomes by latent profiles and intervention

Profile	Recovery			Reliable change			Recovery or improvement			Deterioration			Dropout		
	LI - %	HI - %	Odds ratio	LI - %	HI - %	Odds ratio	LI - %	HI - %	Odds ratio	LI - %	HI - %	Odds ratio	LI - %	HI - %	Odds ratio
LP1	75%	79%	1.26	13%	24%	2.18*	75%	79%	1.25	10%	10%	1.09	24%	21%	0.85
LP2	52%	59%	1.32*	41%	43%	1.08	60%	66%	1.28*	7%	8%	1.11	31%	21%	0.59*
LP3	72%	72%	0.97	9%	5%	0.58	72%	72%	0.97	6%	21%	3.78	15%	21%	1.47
LP4	58%	56%	0.89	39%	35%	0.85	64%	59%	0.82	7%	11%	1.51	22%	11%	0.44*
LP5	30%	31%	1.05	48%	51%	1.12	41%	44%	1.11	7%	5%	0.61	28%	24%	0.8
LP6	45%	53%	1.39*	28%	30%	1.09	49%	56%	1.29*	17%	21%	1.26	32%	23%	0.66*
LP7	11%	17%	1.66*	37%	40%	1.13	19%	24%	1.33*	4%	6%	1.73*	39%	32%	0.74*
LP8	29%	32%	1.11	51%	53%	1.1	39%	44%	1.19*	6%	7%	1.2	39%	32%	0.75*
Total	42%	41%	0.98	40%	42%	1.09*	49%	49%	0.97	8%	9%	1.19*	32%	26%	0.73*

* Indicates significant odds ratios at $p<0.05$

Outcomes by profile

The results from the comparison between LI and HI interventions suggest that for a number of profiles there are significant differences in outcomes when different intensities of treatment are received by members of the same profile. The findings are summarised in Table 6.27 and a summary for each profile is set out below:

LP1: Patients in this profile have a high likelihood of recovery following treatment for both LI and HI interventions, which at 75% is considerably higher than the IAPT national recovery figure of 46% (NHS Digital, 2016). However, the probability of reliable change is significantly higher following HI interventions than LI and this is the only profile where there is a significant difference in favour of HI for reliable change. The odds of deterioration and dropping out of treatment were not significantly different between the interventions. When selecting an intensity of interventions for patients in this LP, the limited differences between outcomes would suggest that LI interventions would be sufficient to provide a good outcome for the majority of patients. However, there may be situations where LI would not be an appropriate intervention, for example for patients with PTSD for whom there is no evidence based LI treatments.

LP2: This profile also show a high percentage of patients recovering with LI treatment, with over 50% of patients moving to recovery. However, there was significantly more recovery for LP2 patients who received HI interventions compared to HI. There were no significant differences between intervention intensities for achieving reliable change or deteriorating for members this LP, although there was significantly less dropout in HI interventions. When selecting an appropriate treatment for this LP, HI interventions will significantly increase the likelihood of recovery and keeping patients in treatment, although a high percentage of patients in this profile (52%) will meet criteria for recovery following LI interventions delivered by the services. One suggestion may be to initiate LI interventions, but be prepared to step up to HI if there is limited improvement after three sessions as recommended by some researchers in the field (Lambert, 2013).

LP3: Analyses on this profile included only a small number of patients and therefore the results should be interpreted with caution. The findings of the analyses suggested that although differences are not statistically significant, LI interventions typically increased the likelihood of recovery and decreased the likelihood of deterioration and dropout for this profile of patients. The findings for most profiles indicated that HI treatments improved the likelihood of recovery and reliable change, even if the difference was not statistically significant, and therefore the results for LP3 suggest that these patients may better more for LI treatment.

LP4: Members of LP4 also had an increased likelihood of recovery and reliable change following LI treatments when outcome to members of this profile who received HI treatments, although the differences were not statistically significant. This profile were

therefore the only other profile, in addition to LP3, to show a benefit of LI treatments over HI. However, the incidence of dropout is significantly different between the interventions, and was in favour of HI interventions (22% LI vs 11% HI). When selecting treatments for this profile, unless it is felt that the patient is at risk of dropping out, the lack of significant difference in response to treatment outcomes, and the trend supporting increased benefits of LI interventions, would suggest that lower intensity treatment be recommended as an initial step.

LP5: The profile of patients in LP5 suggest moderately severe symptoms, and this profile were found to have a low likelihood of recovery following both LI and HI treatment. Interestingly, there was no significant difference in the likelihood of any of the five outcomes between members of this profile who received LI and HI interventions. This suggests that LI interventions may be an appropriate treatment, despite the low probability of positive outcomes (30% recovery) and therefore consideration might be given to other services or more intensive interventions that could provide more appropriate treatment.

LP6: Individuals in this profile show significantly increased odds of recovery and reduced dropout following HI interventions compared to LI. The probability of recovery following LI interventions was very similar to that of the national average (45% vs 46%), and therefore treatment recommendations may be either to initiate HI interventions as the first step, or potentially an LI intervention with an increased number of sessions. The major risk with this profile of patients is the relatively high probability of deterioration following treatment (20%), and this level of deterioration was notably higher than the estimated average of 5-10% (Boisvert & Faust, 2003; Crawford et al., 2016; Rozental et al., 2016). This may suggest that clinicians consider alternative non-IAPT interventions or monitor progress carefully with patients of this profile. Further analysis of this profile using information not collected routine by IAPT services may help predict which patients are most at risk of deterioration.

LP7: Overall, LP7 reported the lowest percentage of patients who were in recovery following treatment, although there were significantly increased odds of recovery in HI treatments compared to LI (17% vs 11%). When an improvement in symptoms was considered alongside recovery (‘recovery or improvement’) the probability of a positive outcome increased to 25% following HI treatments. There was significantly decreased risk of dropout during HI interventions, although the risks of deterioration where higher when HI treatment was received compared to LI. With consideration of these findings, it might be suggested that IAPT services were not the most appropriate setting to treat this profile of patients, and that referral to alternative services may increase the likelihood of a positive outcome. If treatment was to be delivered in IAPT services, HI would be recommended as the initial treatment step.

LP8: The findings for these patients suggested that although the odds of recovery were not significantly different between LI and HI interventions, but the HI treatments were found to

significantly increase the likelihood of the 'recovery or improvement' outcome being achieved. As there was no significant difference between interventions for deteriorating or reliable change, and as HI treatment were significantly associated with reduced odds of dropout, it might be suggested that HI treatment would be recommend for LP8 patients as initial IAPT treatment.

Stepped up patients.

Although the main comparison of outcomes following different intensities of treatment is focused on individuals receiving LI or HI interventions only, a description of the probability of treatment outcomes for individuals who were stepped up from LI to HI interventions is included in this section. As the number of stepped up patients meeting inclusion criteria within some of the profiles was very low, logistic regression analyses comparing the probability of outcome between 'Stepped up' and patients who received LI or HI only within profiles were not performed due to a lack of power. A total of n=2367 patients were stepped up during their episode of care with the services, compared to n=6926 who received LI interventions only, and n=4314 who received only HI interventions (Table 6.19). Instead the percentage of patients achieving each outcome is compared between patients in each of these three intervention groups.

Recovery

The percentage of patients who were in recovery after being stepped up is compared to individuals who received LI or HI interventions only in Table 6.28. Overall, individuals who were stepped up during the course of treatment at the services had a lower probability of being in recovery than individuals who received only LI or HI interventions (37% compared to 42% and 41%). LP4 were the only LP where the probability of recovery was higher for stepped up individuals than for individuals receiving HI or LI treatments only. As stepped up individuals were not responding to the initial LI intervention, there may have been important patient characteristics that were contributing to this lack of response, not available in the IAPT MDS. Stepping up to HI treatment may have no improved the situation for a number of patients as these patients were unlikely to benefit from either intensity of IAPT treatment. This may explain why the probability of recovery was generally lower for stepped up patients than those for patients who received HI treatment only.

Table 6.28. Probability of recovery across LI, HI and Step up interventions.

Profile	Low intensity			High Intensity			Stepped Up		
	Total	Yes	%	Total	Yes	%	Total	Yes	%
LP1	511	384	75%	231	183	79%	99	74	75%
LP2	2,192	1,145	52%	917	541	59%	602	321	53%
LP3	47	34	72%	39	28	72%	5	3	60%
LP4	372	217	58%	236	131	56%	86	53	62%
LP5	739	225	30%	442	139	31%	267	73	27%
LP6	675	304	45%	525	280	53%	189	83	44%
LP7	675	76	11%	959	167	17%	381	59	15%
LP8	1,715	501	29%	965	304	32%	738	217	29%
Total	6,926	2,886	42%	4,314	1,773	41%	2,367	883	37%

Reliable change

Individuals who were stepped up during contact with the services did not appear to have a greatly increased probability of achieving reliable change either, as presented in Table 6.29. As HI interventions did not significantly increase the likelihood of reliable change compared to LI treatments for any profile except LP1, it is not surprising that Stepped up patients appeared to have a similar probability of this outcome to patients allocated to LI or HI treatments only. For LP5, the probability of reliable change was considerably lower for Stepped up patients compared to LI (42% vs 48%), but as these patients were already not benefiting from LI treatments it may not be surprising that changing treatment did not significantly improve the situation, and instead allocating to HI initial for this profile of patients may have improved outcomes.

Table 6.29. Probability of reliable change across LI, HI and Step up interventions.

Profile	Low intensity			High Intensity			Stepped Up		
	Total	Yes	%	Total	Yes	%	Total	Yes	%
LP1	511	64	13%	231	55	24%	99	18	18%
LP2	2,192	891	41%	917	390	43%	602	246	41%
LP3	47	4	9%	39	2	5%	5	0	0%
LP4	372	145	39%	236	83	35%	86	32	37%
LP5	739	353	48%	442	224	51%	267	112	42%
LP6	675	188	28%	525	155	30%	189	51	27%
LP7	675	253	37%	959	388	40%	381	148	39%
LP8	1,715	870	51%	965	513	53%	738	374	51%
Total	6,926	2,768	40%	4,314	1,810	42%	2,367	981	41%

Recovery or improvement

The percentage of individuals achieving recovery or improvement across the three intervention groups is presented in Table 6.30. Stepping up increased the likelihood of recovery over LI only for LP1, LP6, LP7 and LP8 but all were very marginal increases in the probability of the outcome.

Table 6.30. Probability of recovery or improvement across LI, HI and Step up interventions.

Profile	Low intensity			High Intensity			Stepped Up		
	Total	Yes	%	Total	Yes	%	Total	Yes	%
LP1	511	385	75%	231	183	79%	99	77	78%
LP2	2,192	1,317	60%	917	604	66%	602	364	60%
LP3	47	34	72%	39	28	72%	5	3	60%
LP4	372	237	64%	236	139	59%	86	54	63%
LP5	739	306	41%	442	194	44%	267	100	37%
LP6	675	332	49%	525	292	56%	189	94	50%
LP7	675	131	19%	959	232	24%	381	87	23%
LP8	1,715	677	39%	965	421	44%	738	302	41%
Total	6,926	3,419	49%	4,314	2,093	49%	2,367	1,081	46%

Deterioration

The likelihood of clinical deterioration is presented in Table 6.31 and the analysis suggested that individuals who were stepped up had a slightly increased likelihood of deterioration when compared to patients who only received LI treatments. Poorer outcomes may be expected here as stepped up patients would be considered a sub-group of LI only patients who were not showing significant benefits from treatment and may have already have displayed a worsening of symptoms, hence the decision to step up. The probability of deterioration between stepped up and HI only patients was very similar, except for LP3 where only n=5 Stepped up patients were included, therefore making it difficult to draw conclusions from this analysis

Table 6.31. Probability of deterioration across LI, HI and Step up interventions

Profile	Low intensity			High Intensity			Stepped Up		
	Total	Yes	%	Total	Yes	%	Total	Yes	%
LP1	511	49	10%	231	24	10%	99	11	11%
LP2	2,192	160	7%	917	74	8%	602	65	11%
LP3	47	3	6%	39	8	21%	5	2	40%
LP4	372	27	7%	236	25	11%	86	9	10%
LP5	739	53	7%	442	20	5%	267	17	6%
LP6	675	116	17%	525	109	21%	189	40	21%
LP7	675	26	4%	959	62	6%	381	23	6%
LP8	1,715	97	6%	965	65	7%	738	54	7%
Total	6,926	531	8%	4,314	387	9%	2,367	221	9%

Dropout

The likelihood of dropping out of treatment seems higher for individuals who were stepped up than patients who attended either LI or HI interventions only, as presented in Table 6.26. Across all profiles except LP3, the percentage of individuals dropping out of treatment was much higher for individuals who were stepped up than for patients who received LI or HI treatments only. This may have been linked to non-response to the initial LI treatment, and patients who were stepped up my have believed that IAPT treatments were unlikely to help them, and therefore terminated treatment.

Table 6.32. Probability of dropout across LI, HI and Step up interventions.

Profile	Low intensity			High Intensity			Stepped Up		
	Total	Yes	%	Total	Yes	%	Total	Yes	%
LP1	511	121	24%	231	48	21%	99	31	31%
LP2	2,192	676	31%	917	191	21%	602	194	32%
LP3	47	7	15%	39	8	21%	5	0	0%
LP4	372	81	22%	236	26	11%	86	25	29%
LP5	739	207	28%	442	105	24%	267	94	35%
LP6	675	214	32%	525	123	23%	189	77	41%
LP7	675	260	39%	959	303	32%	381	168	44%
LP8	1,715	662	39%	965	310	32%	738	313	42%
Total	6,926	2,228	32%	4,314	1,114	26%	2,367	902	38%

Outcomes for stepped up patients.

The probability of positive outcomes from treatment was typically lower for stepped up patients than for patient who received HI treatments, and similar to patients who received LI treatments only across profiles. However, as stepped up patients could be considered a subgroup of LI patients who were considered to be non-responders to treatment, it might be

expected that this group of patients would continue to show a lack of response even when stepped up, and instead argues that allocating these patients to HI treatments as the initial step could have improved outcomes.

The sub-analysis by treatments showed that stepped up patients were at higher risk of dropout than patients who received only one intervention. It is possible that the lack of response to the LI treatment, followed by a wait until the HI treatment was commenced may have disheartened patients, resulting in them leaving treatment rather than waiting for the HI treatment. Allocating to a more appropriate HI treatment in these situations could reduce the risk of dropout, as it could provide a positive experience of treatment earlier, rather than using patient and service resource on a LI treatment that is unlikely to yield benefits.

Discussion

The findings from Analysis 3, which has compared outcomes between different intensities of treatment by LP suggests that for certain profiles there may be increased benefits associated with different intensities of intervention. Considering the patient's LP following initial assessment may have clinical utility in informing decisions about appropriate IAPT treatment, and used aid treatment selection decisions. For example, as the likelihood of positive outcomes was significantly higher following HI interventions than LI interventions for patients in LP8, it might therefore be recommended that HI be considered as the initial treatment for patients with this profile.

When there was a difference in outcomes between the intensities of intervention for a profile, typically it was found that HI either increased the odds of a positive outcome, or reduced the risks of dropping out of treatment, over LI interventions. As HI interventions typically involve more face-to-face contact, this finding may not be surprising. Instead it is potentially more surprising that HI interventions did not increase the magnitude of positive outcomes for more profiles. The likelihood of recovery was only significantly higher in LP2, LP6 and LP7 when HI treatments were received compared to LI, and reliable change was only significantly more likely in HI for one profile (LP1). However, there may have been important reasons for the initial selection of HI treatment, including clinical presentations such as PTSD that are not captured in the nine included patient characteristics.

Although the difference in outcomes between interventions was very similar for some profiles, the difference between 2% or 3% would be significant at the service or national level as it would indicate a significant absolute number of patients who would have benefitted from treatments. However, at the individual level the expected utility of outcomes would be similar and might suggest that the decision could be made incorporating patient preferences, but also worth considering that the LI intervention might be more appropriate

due to the reduced burden on both the patient service resource, as well as reduced waiting time in services.

Overall dropout was higher in LI interventions than for HI, but this may in part the nature and means of delivery of these interventions. HI interventions are typically face-to-face sessions of one-to-one therapy and therefore dropping out of treatment may be less likely as regular contact with a clinician may strengthen the alliance with the patient and treatment (Strunk et al., 2012). As LI interventions have more limited face to face contact, they may leave the patient more vulnerable to dropping out of treatment. A recent meta-analysis of treatment dropout in CBT found the probability of dropout for individual CBT was 24.6% during treatment, which is very similar to the percentage found for HI interventions in the current analysis, but was 34.2% for computer/technology delivered interventions (Fernandez et al., 2015) which is perhaps more similar to the self-help interventions offered in IAPT. It is also possible that for some patients LI treatment was successful and as a result they no longer felt the need to stay in contact with services, but this was still recorded as dropout due to lack of precision in coding as well as a low likelihood of follow up by the service.

It is of interest that for both LP3 and LP4 the LI interventions increased the odds of a positive outcome following treatment compared to HI treatment, even though these differences were not significantly different. For all other profiles HI was generally favourable, and therefore this finding suggests that there is something within these profiles that may contribute to the different findings. The main similarity between LP3 and LP4 is that patients in these profiles are typically older and more likely to be from a white ethnic group than other profiles. Due to the low numbers of patients who attend the services from these profiles, especially the number of LP3 patients who were caseness, it is difficult to draw conclusions and further research with increased numbers of patients from these profiles would be of value.

When individuals who were stepped up during contact with services were compared to patients in each LP who only received either LI or HI interventions, the results suggested minor increases in positive outcomes over LI interventions alone for some profiles, although poorer outcomes found for others. Findings also suggested that dropping out of treatment is higher for individuals who were stepped up than those who received either LI or HI alone. It is quite possible that some patients who did not benefit from LI interventions and then were offered additional HI support declined further treatment as they disheartened with their experience of treatment from the services. They may also have had a negative experience of treatment which would make the patient less likely to benefit from further treatment provided by the same service. As stepping up patients did not appear to greatly increase the likelihood of positive outcomes than receiving LI interventions alone, identifying the most appropriate treatment following initial assessment may provide the best opportunity to improve patient outcomes. It may be that additional patient characteristics not included in the current IAPT MDS could help refine the profiles and help identify patients who are likely to benefit from being stepped up.

This analysis has provided considerable information about the probability of treatment outcomes for LPs that may help inform clinical decisions around the most appropriate treatment for a new patient entering the services. For some profiles of patients, the results suggest that an allocation to HI treatment may significantly increase the chance of positive outcomes, and therefore allocating to LI could be considered inappropriate. A recommendation that displays the potential value of HI over LI for certain profiles of patients could be included as part of a clinical DST to aid treatment choices by providing the information on LP outcomes presented in this chapter. Such information could inform the decision based on the expected utility of treatment options. For some profiles, there was no significant difference in outcomes between interventions, which may suggest that the use of the least intrusive and resource intensive intervention (LI) as initial treatment for patients of these profiles. This would support the stepped care model and for some profiles of patients such as LP7 and LP8, allocating to HI as the initial treatment may improve outcomes quicker, increasing service efficiency.

Summary

The analyses presented in this chapter suggest that there are significant variations in the likelihood of treatment outcomes between latent profiles that may be important to consider for patients entering treatment. These differences were maintained in an independent sample of patients attending the same services at a later date and provide evidence of the reliability of the predictions of likely treatment outcome for the profiles. Identifying the profile for each patient at initial assessment may provide valuable information about the likely response to treatment as well as deterioration and dropout, which may be used to inform treatment selection decisions.

In addition to providing a description of stratified groups of patients with similar characteristics attending the services, the latent profiles can also be used to identify differential probability of treatment outcomes. For example, the characteristics of members of LP1 are low levels of depressive and anxiety symptom severity, limited phobic symptoms and relatively high levels of social and occupational functioning compared to the average for the full sample of patients. These patients appear to have a high probability of being in recovery following treatment at the services, whether LI or HI interventions were received, relative to the probability of recovery for the full dataset of patients (74% in LP1 compared to 40% for the full sample). With regard to treatment planning, these patients may be viewed as a profile for whom LI interventions are likely to be sufficient for a good outcome. As 31% of LP1 received HI interventions, this could represent an over use of healthcare resource and an unnecessary burden for patients in some scenarios (although there may be reasons to allocate to HI treatments, such as when a diagnosis of PTSD has been established). LP1 and LP3 share a number of similar characteristics, such as low levels of symptom severity

and a low likelihood of being in receipt of welfare benefits, being prescribed medication or having phobic symptoms. However, the mean age of LP3 is twice that of LP1 (33 compared to 67), as well as being less likely to be from a non-white ethnic group (11% compared to 17%), but these differences appear to have little impact on the likelihood of recovery, which is very similar for both profiles.

Considering patients with poorer outcomes, LP7 show relatively high levels of both depression and anxiety symptoms which may account for the high proportion prescribed psychotropic medication (73%) compared the full dataset (39%). These characteristics may also contribute to the low number in work (74% were in receipt of welfare benefits). The outcomes for this group are relatively poor for both low and high intensity interventions, which suggest that alternative non-IAPT treatment options might be considered. As 73% of LP7 patients have been prescribed psychotropic medication, it may be that reviewing this treatment to ensure patients are actually taking medication could improve outcomes. However, the significant difference in recovery between LI and HI interventions suggests that if they were to be treated in the services included in this study, then a HI intervention should be considered. The outcomes for recovery or improvement also suggest that HI interventions might also be considered as the initial treatment for patients in LP8, as the outcomes for HI interventions were significantly lower for this profile of patients.

Identifying patients at risk of deterioration is important as it may reduce the likelihood of ineffective or harmful treatments being offered. For example, the deterioration rates for LP6 were four times higher (20%) compared to LP7 (5%), and more than double the mean deterioration for the whole sample (9%). The mean deterioration rate for all patients entering treatment is within the range typically seen in psychotherapy research at 5-10% (Boisvert & Faust, 2003; Crawford et al., 2016; Rozental et al., 2016), but it would be of interest to see whether there were a subgroup of patients within other psychotherapy data sets similar to LP6, who may also have a higher risk of negative outcomes. As the probability of recovery for this profile at 46% was similar to the overall recovery rate for the whole sample (and IAPT nationally), this suggests that if treatment were to be offered to this group of patients then a number of additional factors may be worth considering, such as offering HI interventions (as there was a significant difference between LI and HI interventions) and the use of additional interventions (for example, medication). Similarly, as the probability of dropout is very high in LP7 and LP8 patients who were in receipt of LI interventions (39%), this information could be used by clinicians either to consider treatment engagement as part of the treatment, or consider HI interventions as the initial treatment to reduce the incidence of dropout.

One limitation of these analysis is that, although the LPs show differential outcomes, in some cases between different intensities of interventions, there may be important patient characteristics not currently available in the IAPT dataset that could inform current treatment decisions. Although mean PHQ-9 and GAD-7 were not significantly different between

members of each LP attending LI or HI treatments for all profiles except LP1 and LP2, there may have still been additional reasons for allocation to one intensity that were not captured as part of the dataset, for example specific diagnoses such as PTSD or social anxiety disorder for which there are no evidence-based LI interventions recommended for IAPT. The inclusion of anxiety disorder specific measures in future IAPT datasets and analysis may further inform the predictive ability of the LPs, but also contribute to an IAPT-relevant DST, as significant PTSD or social anxiety disorder symptoms could result in a recommendation for HI treatment.

There is potential for the results of these analyses to inform clinical decision-making aids, and for recommendations to be provided to clinicians. The algorithm developed for this analysis uses the nine included patient characteristics described in Chapter 5 and calculates the probability of membership to each of the eight profiles using the posterior probabilities and priors from the LPA.

This algorithm could easily be hosted in the EPMS used by the services, and as the nine patient characteristics are mandatorily collected as part of IAPT assessment, this information could automatically be entered into the algorithm through the record system. Information about profile membership and probability of treatment outcomes could then be provided to clinicians in real time, as this has been shown to import the utility of clinical DSTs (Kawamoto et al., 2005). The algorithm could also be hosted and accessed through mobile phone or tablet apps to allow patients to enter their details and receive a description of their profile, likely response to treatments and a recommendation that they could discuss with their IAPT clinician. As IAPT services have a target to achieve recovery in 50% of patients entering treatment (HSCIC, 2015), the overall percentage recovery of 40% for this sample argues that the more effective allocation of patients to treatment could improve recovery rates in order to achieve the national target.

The algorithm described above could therefore be developed into a treatment selection decision aid and used during initial assessment to inform whether low or high intensity interventions would be the most appropriate. It may also be possible to provide further information to clinicians on the predicted change in symptoms during treatment, rather than just likely prognosis. Even profiles with a high probability of a positive outcome, such as LP1, did not have a 100% probability of recovery and therefore there will be patients who during the course of treatment are show little benefit. It would be of value to identify these potential treatment non-responders as soon as possible to provide an opportunity for the patient and clinician to discuss and consider alternatives, therefore aiding treatment monitoring decisions.

The next section of this thesis explores the expected trajectories of change in symptoms during treatment for patients, using latent growth methods to calculate the mean change in symptoms between sessions. Trajectories of change for individual profiles will also be

estimated to provide a more tailored expected response trajectory. In addition, latent class growth analysis (LCGA) (Muthén, 2001; Muthén et al., 2002) will be performed to identify sub-populations of patients within the data who have differing forms of change, for example those who display a continual reduction in symptoms and patients show little response to psychological treatments.

Chapter 7. Latent Growth Curve & Latent Class Growth Analysis.

Abstract.

The eight latent profiles (LPs) of patients attending IAPT presented in the previous two chapters could already be used to inform treatment selection decisions between different intensities of psychological intervention. However, as profile membership is based wholly on data collected at an initial screening or assessment session, it may be of further interest and clinical utility to understand how patients' symptoms change during the course of treatment, with potential to aid clinical decisions made during the delivery of an intervention. In this chapter, latent growth methods (McArdle, 1986) and latent class growth analysis (LCGA) (Muthén, 2001) were used to identify the expected trajectories of change in symptom severity for patients attending low intensity guided self-help treatment (n=3334) and high intensity (n=4394) psychological interventions at the IAPT services. The mean change in both depression and anxiety symptoms was similar between the two intensities of intervention, with greater change in symptoms recorded during the initial sessions of treatment, which then levelled out as the number of sessions increases. Growth curves were also performed by LP, and it was found that although the form of change was similar across most profiles, distinct trajectories of change existed between them. LCGA was used to identify groups (classes) of patients with different trajectories of change for both depression and anxiety symptoms. The distribution of these trajectories varied between LPs, indicating that individual profiles were more likely to be associated with just two or three of the trajectories. Providing information on expected trajectories of change for individual LPs could inform a more personalised approach to both treatment selection and treatment monitoring for patients entering treatment. The identification of different trajectories has potential to be used in clinical practice to aid clinical decisions during treatment, by monitoring change and providing feedback about whether the patient's progress in treatment is or is not indicative of positive outcomes at the end of therapy.

Introduction.

Being able to predict whether patients will respond or not to psychological interventions from baseline assessment information has potential benefit in aiding clinical decision making in routine care. The latent profiles (LP) previously identified provide information on the probability of different treatment outcomes when either LI or HI interventions were received at IAPT services. This could be used by clinicians to inform treatment selection at initial assessment. However, as baseline information is necessarily limited, with potentially

important variables either not measured, or not feasible to measure (e.g. bio-markers), there may be additional clinical value in providing information about how a patient is responding to a given treatment (Lambert, 2013). This could be used to determine whether or not they are following an expected trajectory that would indicate a positive outcome at the end of treatment. Within the context of IAPT services, this may involve decisions about whether individuals would benefit from being stepped up from LI to HI interventions, whether the focus or modality of the psychological intervention could be changed, if a discussion with a GP about medication may be appropriate, or the need to refer the patient to secondary care.

Routine outcome measurement (ROM), the collection of information on a patient's symptoms and other clinically relevant information at multiple time points during the course of an intervention, can provide valuable information to both the clinician and patient on the progress of treatment (Harmon et al., 2007; Lambert & Shimokawa, 2011). IAPT services promote the use of ROM, with scores on both the PHQ-9 and GAD-7 collected at each treatment contact with the service as part of the IAPT service specification (CSIP, 2007).

Research evidence suggests that providing clinicians with information on their patient's progress during treatment can have a positive effect on treatment outcomes. For example, the QA system designed by Lambert and colleagues (2001) which uses patient ROM information to feedback to the clinician when change in symptoms is indicative of a poor outcome from treatment (see Chapter 2). Patients whose reduction in symptoms is less than expected are flagged as being 'Not on track' and the identification of these patients enables the clinician to consider potential adaptions to treatment, for example changing treatment modality or referring to secondary care services. A meta-analysis of treatment trials that have compared the use of the QA system to control groups without feedback have found it to be associated with better patient outcomes, especially for 'Not on tack' patients (Shimokawa et al., 2010). The positive impact of feedback on patient outcomes in routine treatment services has also been found in studies using alternative feedback methods to the QA system (Knaup et al., 2009), including the use of pencil and pen charting (Anker, Duncan, & Sparks, 2009).

The use of feedback in psychological interventions in this way has typically involved 'expected response curves', graphical representations indicating the likely trajectory of change in symptoms over time for a typical patient (Bybee, Lambert, & Eggett, 2007). Patients who do not appear to be following the expected response are therefore 'Not on track'. The typical pattern of change in psychological interventions generally shows a greater impact of treatment on symptom reduction in the first few sessions, displayed by a steep curve, which then levels out as the number of treatment sessions increases (Cuijpers et al., 2013; Kopta et al., 1994). The rate of change in the first few sessions is also associated with eventual treatment outcome (Lewis, Simons, & Kim, 2012) and it has been suggested that up to 40% of the variance in final outcome can be attributed to change in symptom scores by the third session of psychological intervention (Lambert, 2013). Meta-analyses of

pharmacological interventions have reported similar findings, with two reviews of antidepressant trials suggesting that relatively large decreases in symptom severity within the first two weeks are predictive of eventual treatment response (Gorwood et al., 2013; Szegedi et al., 2009).

Longitudinal structural equation modelling approaches such as latent growth curve (LGC) modelling provide a valuable method of describing the process or form (trajectory) of change for a sample of individuals with data from repeated time points, and has shown utility in studies of psychological interventions (Clapp et al., 2013; Henderson, Rowe, Dakof, Hawes, & Liddle, 2009; Stulz, Lutz, Kopta, Minami, & Saunders, 2013). In the context of the current IAPT dataset, ROM data from both the PHQ-9 and GAD-7 could be used to identify the expected response curve of symptoms scores for all patients entering treatment. The expected curves could then be presented to clinicians and patients as part of the EPMS, and this information could be made available for any new patient entering treatment. The potential benefit to clinical practice is that change in symptoms scores can be compared to the identified expected response curve for similar patients attending the same treatment, and clinicians could use this information to decide whether the current treatment was having the expected positive effect or not. The use of LGC modelling could be combined with information from the previously identified LPs to tailor the expected response curve to specific profiles of patients, and so increasing both the precision of the estimates and their potential utility in clinical practice.

LGC analysis can be extended to identify different subgroups, or classes of curves/trajectories of change within the sample using growth mixture modelling (GMM) approaches, such as latent class growth analysis (LCGA) (Muthen, 2001; Muthen et al., 2002). The aim of these methods is to identify homogeneous groups (classes) of individuals with similar trajectories within a heterogeneous sample of individuals. For example, there may be two statistically distinct classes in the sample – one displaying a general decrease in symptoms over time points, whereas a second 'curve' (class) may display a flat line indicating a lack of change over time (non-response to treatment). These approaches have been used in modelling treatment outcomes and symptom change in CMHDs in routine treatment (Rubel et al., 2015; Stulz, Lutz, Leach, Lucock, & Barkham, 2007), the probability of mental health diagnosis over time (Paksarian et al., 2016) and trajectories of service utilisation (Musliner et al., 2016).

A GMM analysis on a dataset of 192 patients that received six sessions of psychological interventions in routine treatment services by Stulz et al (2007) using the CORE-OM symptom scale identified five distinct classes of patients over first six sessions of treatment. Four of the identified classes showed relatively flat response trajectories, suggesting little change in symptoms over the first six sessions, and the main variation between these trajectories was the different levels of initial symptom severity or intercept values. However, one class, defined as 'early improvement' indicated a group of patients who demonstrated a

steady and consistent decrease in symptoms over the first six sessions of treatment. Although this analysis was conducted on only the first six sessions, rather than through to the end of treatment, this number of sessions is around the average number received in IAPT services, especially low intensity interventions (Chan & Adams, 2014; Radhakrishnan et al., 2013). As HI interventions are expected to include more than six sessions of treatment (NHS Digital, 2016), analysis using an increased number of time points would be more valuable to a potential DST in IAPT services. Other authors have found less distinct classes of change during analysis of CMHD datasets from clinical trial samples. For example, Gueorguieva, Mallinckrodt and Krystal (2011), performed GMM methods on a dataset of patients receiving antidepressant treatment for depression and the results indicated two classes of patients, which could be classified as responders and non-responders to the intervention.

Although clinicians could compare symptom scale scores for an individual patient to create a graphical representation of change in symptoms during treatment (e.g. using pen and paper), there are a number of reasons why using all available data from comparable patients through computer-based technology could have advantages. Firstly, research has shown that clinicians are not always able to reliably predict which patients will benefit from treatment (Hannah et al., 2005; Walfish et al., 2012), and therefore an objective system based on comparable patients that could present the expected trajectory and likely outcome may be more instructive. Secondly, creating manual reports for individual patient trajectories would be demanding for clinicians to process, therefore the use of computer-based decision support through the graphical reports on patient progress can be a more efficient use of resource (Gibbons et al., 2015). Thirdly, entering ROM information in the healthcare system can contribute to a quality assurance process, which could aid service-level decision making and commissioning of treatments (Lambert, 2013).

Using the previously identified LPs could be a useful method of stratifying patients entering treatment and could contribute to more personalised expected response curves for patients receiving treatment at IAPT services, as it was theorised that trajectories could differ by LP. The LPs could therefore be used both to personalise clinical decisions about treatment selection but also inform decisions during the course of treatment, such as whether the intervention is having the expected effect on clinical symptoms. If the estimation of the expected response curves could be built into the LP algorithm developed in Chapter 6, then this could produce a DST for personalising both treatment selection and treatment monitoring decisions for IAPT services.

The aim of this chapter was to explore the expected response curves, as measured on both the PHQ-9 and GAD-7, of individuals receiving treatment in the IAPT services. LGCs were estimated for the full sample of included patients, but also by LP to identify the expected response curve of each profile. In addition, GMM methods were used to identify distinct

classes of patients who differed with regard to their trajectories of symptom change during psychological interventions.

Method.

Sample.

The sample used for this analysis was taken from the same n=15376 patients attending the two IAPT services as described in Chapter 6, including all patients reporting caseness and with any post-treatment assessment (T2 data) from the discovery and validation datasets.

As the analysis proposed to explore the form of change in clinical symptom scores within treatment, it was decided to focus on sub-samples of patients who received similar types of psychological intervention. Researchers advocate a minimum sample size for structural equation modelling, including latent growth methods, and typically around 100 to 200 cases are the suggested minimum sample size (Barrett, 2007; Curran, Obeidat, & Losardo, 2010). Therefore, it was decided to keep intervention types at the higher order rather than breaking them down into more specific types (such as IPT, couples therapy) with limited numbers of available patients. The majority of HI treatments could be easily grouped together for the analysis, and specific low intensity interventions such as guided self-help and computerised-CBT were grouped together as 'Guided self-help'.

IAPT services may deliver more than one intervention type during an episode of care for a particular patient, such as providing group-based interventions in addition to individual psychological therapy, as well as stepping up patients from low to high intensity interventions. For the current analysis therefore, the 'main intervention' was identified for each patient as the type of intervention they received most frequently during their episode of care with the IAPT services. In situations where two or more interventions were received in equal amounts, the more intensive intervention was selected as the main intervention. Table 7.1 displays the frequency of each main intervention in the dataset. The majority of patients received either guided self-help (GSH; 44.7%) or high intensity (HI; 39.9%) interventions.

The next most common interventions were non-facilitated or 'pure' self-help (6.3%) followed by group-based interventions (4.4%). No intervention was recorded for 2.2% of patients and 1.9% of patients received signposting to services outside of IAPT as their main intervention. Less than 1% of the sample received either exercise-based interventions or other interventions that were labelled as non-standard by clinicians at the services.

Table 7.1. Frequency of main intervention.

Main Intervention	Number of Patients	Percentage of total
Guided SH	6,870	44.68%
HI	6,136	39.91%
Pure SH	965	6.28%
Group Intervention	671	4.36%
Not recorded	338	2.20%
Signposting	296	1.93%
Non-standard HI	89	0.58%
Exercise	6	0.04%
Non-standard LI	5	0.03%
Total	15,376	

Information about the non-standard interventions was limited and as such a small number of patients received these interventions, patients receiving these as their main interventions were excluded from the analyses, as were the n=6 patients for whom exercise was the main intervention. Patients who had no intervention recorded were excluded (n=338), as were patients who received 'sign-posting' as their main intervention (n=296), as this was not considered a psychological intervention. There were n=671 patients for whom 'groups' was the main intervention received. However, group-based interventions delivered across services may have very different functions, in particular relapse prevention, trauma-focused work or psychoeducation, and can therefore vary in duration, purpose and the types of patients for whom the groups are focused. As a result, it was decided to exclude 'group-based' interventions from latent growth analysis due to the potential variation in the actual intervention received between patients.

Patient flow for inclusion in the analysis is displayed in figure 7.1. Latent growth modelling requires observations from at least three time points to model change in symptoms, and therefore patients who had just two contacts with the services were excluded from the analysis. In addition, patients were only included if they received at least three sessions of their 'main intervention'. Otherwise there would be a risk that any additional interventions may have had a major effect in the change in symptom scores which could not be attributed to the main intervention.

By using this exclusion criteria, only n=182 patients who received pure self-help as their main intervention could still be included from an original sample of n=965 patients receiving this as their main intervention. This loss of over 80% of all patients receiving pure self-help was mainly due to patients having just two time points typically when receiving this intervention, which is most likely due to the nature of self-help interventions and follows both NICE guidance and the design of IAPT services (CSIP, 2007; NICE, 2011). As less than 200 included patients received pure self-help, and because there would be limited time

points available for this sample due to the nature of the intervention, all patients with this as their primary intervention were excluded.

This resulted in a sample of n=3465 patients who received low intensity GSH interventions and n=4903 patients who received HI interventions as their main intervention type. It was observed that a small number of patients (n=640, 7.49%) across both samples received quite a large number of different additional interventions to their main intervention type. As the analysis was focused on two specific interventions (HI and GSH), additional interventions may confound the trajectories of change. As a result, patients who received three or more additional intervention sessions that were not their main intervention type were excluded from the analyses described in this chapter[1]. The final dataset of included patients consisted of n=3334 patients who received GSH and n=4394 who received HI. The distribution of included patients by Latent profile (LP) is displayed in Table 7.2. Guided self-help and HI patients were modelled separately throughout the analyses described in this chapter. LP3 included only n=10 & n=14 patients for GSH and HI respectively in the dataset, and due to this very low number, growth curves could not be specified for this LP and it was therefore excluded from the analysis.

Table 7.2. Number of patients receiving GSH or HI, by latent profile.

Number of patients included (GSH)		Number of patients included (HI)	
All cases	3334	All cases	4394
LP1	133	LP1	135
LP2	1112	LP2	1052
LP3	10	LP3	14
LP4	167	LP4	180
LP5	363	LP5	453
LP6	326	LP6	487
LP7	325	LP7	924
LP8	898	LP8	1149

Measures.

Sessional routine outcome measurement (ROM) data is collected and entered into the healthcare information system at each contact with the IAPT services, and this monitoring information can provide cumulative session by session reporting. The current dataset used in this analysis included symptom measure total scores at each contact.

[1] Analysis was conducted on the sample which included patients with three or more additional intervention sessions, and is presented in Appendix E. Model fit statistics for this sample were very similar to those for the included sample.

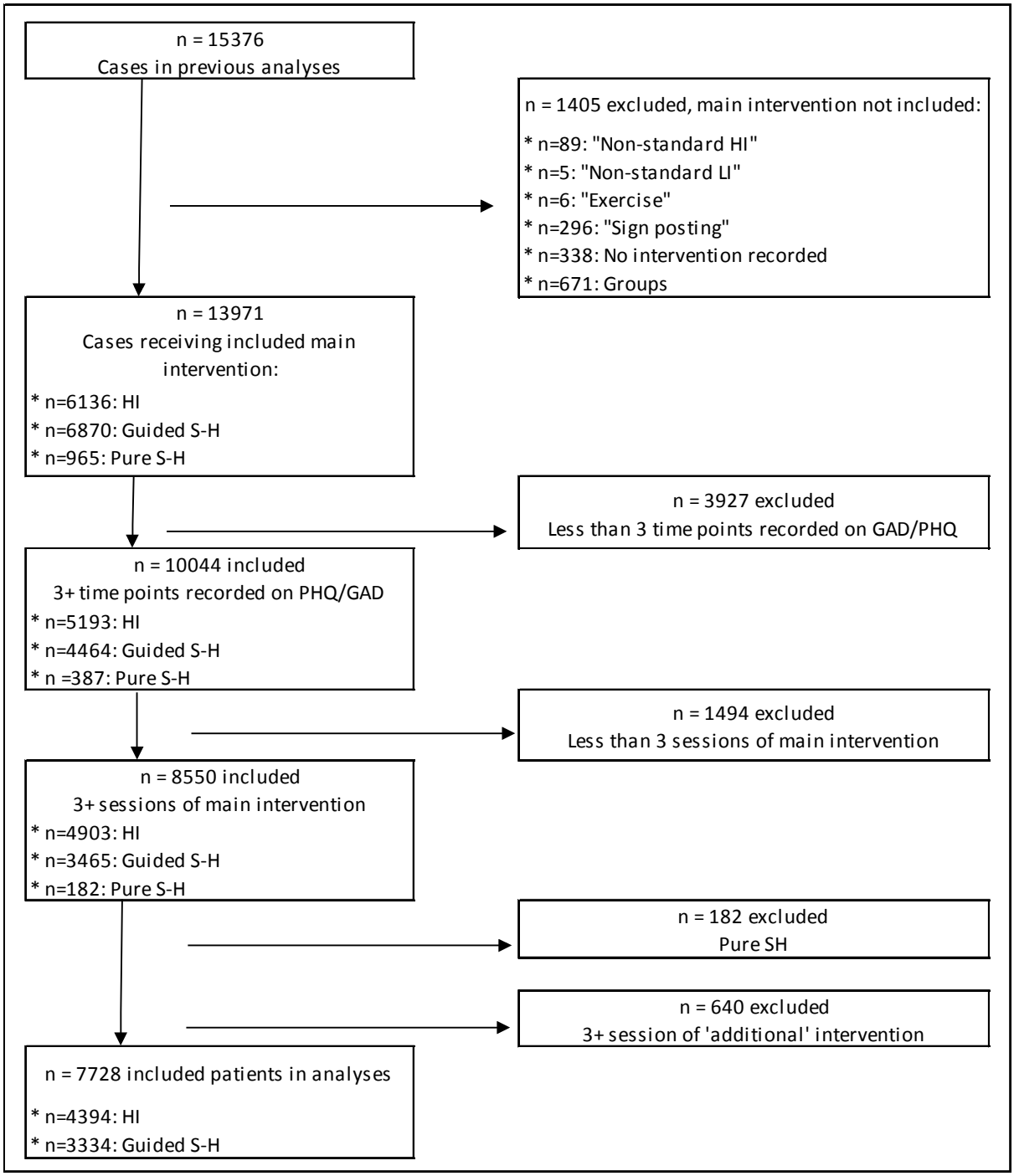

Figure 7.1. Patient flow diagram - growth modelling

It was decided to use up to eight time points for the analysis of GSH interventions, which was consisted of the initial (0) baseline time point and up to seven additional symptom measures collected during contact with the services. Previous researchers have advocated using upper limits on the number of included time points that are close to the mean number of sessions in growth curve modelling, as there is a risk the slope in the LGC could be

distorted when relatively few patients are providing data for later time points (Lutz et al., 2005). The mean number of GSH sessions received by included patients was five (SD=1.72) in the current dataset, and clinicians from the services suggested that up to six sessions were typically provided as standard. NICE guidance for GSH interventions for the treatment of depression and anxiety recommends between five and eight sessions of GSH for CMHDs (NICE, 2009; NICE, 2011) and therefore the eight time points included in the current analysis should cover assessment and evidence based treatment received for the majority of patients.

As HI interventions were expected to be more intensive and therefore more sessions would be received by patients, it was decided to use an upper limit of 13 time points in the analysis, which was made up of a baseline assessment symptom score and 12 additional contacts with the service. Analysis of the current IAPT dataset found a mean of 10.9 sessions (SD = 4.31) which appears very close to the average suggested in the IAPT service specification of 12 sessions (CSIP, 2007).

Analysis plan.

The analyses presented in this chapter are split into two separate studies. Analysis 1 uses 'latent growth curve analysis' (LGC) on the two intervention datasets (GSH and HI) to estimate the mean change in symptom during treatment for both the full intervention sample and the sub-samples of LP members. Analysis 2 uses 'Latent class growth analysis' (LCGA) to identify classes of trajectories in the datasets. Both analyses are described in further detail below. Data analysis was conducted in Mplus version 7 (Muthén & Muthén, 2012).

Analysis 1 - Latent growth curve analysis

The initial phase of the analysis was to specify the form of change in both depression and anxiety symptoms using the PHQ-9 and GAD-7 outcome scores at each time point, for both GSH and HI interventions separately. The analyses would first specify a linear slope as the form of change, then compare model fit statistics when a quadratic slope is specified to observe where this is a better fit for the data. Previous research has typically identified that a quadratic slope provides the best fit for change during psychotherapy treatment (Clapp et al., 2013; Gueorguieva et al., 2011). In the final stage of this analysis, the residuals of each time point would be allowed to correlate to explore whether the inclusion of a potential time-varying factor improved model fit, as is often suggested in structural equation modelling of longitudinal data (Cole, Ciesla, & Steiger, 2007; Landis, Edwards, & Corina, 2009; Muthén & Muthén, 2008).

To compare the LGC models, the following model fit statistics and thresholds were used. Firstly, the Comparative Fit Index (CFI) and Tucker-Lewis Index (TFI) values were

considered with values >0.95 suggesting good model fit, and >0.97 excellent fit (Geiser, 2013; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Secondly, the root mean square of error of approximation (RMSEA) and the standardised root mean square residual (SRMR) were compared, with values <0.05 indicating good model fit on both statistics (Berlin, Parra, & Williams, 2014; Hu & Bentler, 1999).

After establishing the best model fit for change in PHQ-9 and GAD-7 scores for all included individuals receiving either GSH or HI interventions, the LGCs for each Latent Profile (LP) were specified to allow a comparison of expected response curves between LPs.

Analysis 2 - Latent class growth analysis

In order to identify statistically distinct classes of response curves within the IAPT samples, latent class growth analysis (LCGA) was selected. LCGA is a specific type of GMM method that fixes both the slope and intercept to equality across individuals within a trajectory class (Andruff, Carraro, Thompson, & Gaudreau, 2009). Fixing the within-class variance to zero in this manner provides a clearer identification of classes, as recommended as the preliminary step with exploratory group trajectory models (Jung & Wickrama, 2008).

LCGA modelling uses the same model fit statistics that are considered in the latent profile analysis (described in Chapter 5), and for this analysis the Vuong-Lo-Medell-Rubin Likelihood Ratio test (VLMR-LRT; Lo, et al., 2001), the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and entropy value were compared between models. The VLMR-LRT is a comparison between the K model (current model with K number of classes) and the K-1 model (model with one less class), with a p-value <.05 indicating that the K model fits the data better than the K-1 model. A p-value >=0.05 (no longer statistically significant) suggests that the model with one less class is a better fit for the data, and this more parsimonious K-1 model would be preferred over the K model. Lower AIC and BIC values for one model compared to another indicate better model fit, whereas higher entropy values indicate higher accuracy in classification for the model.

As there were no prior hypotheses on the number of classes, the LCGA was conducted with a two-class model, assessing fit statistics and increasing the number of classes until the VLMR-LRT became non-significant or any of the AIC or BIC values increased compared to the previous class solution as is standard for GMM/LCGA methods (Geiser, 2013; Musliner et al., 2016; Rubel et al., 2015). It is also common practice to apply restrictions on the minimum size of a class in GMM methods, with the commonly applied standard being that a class must contain at least 5% of the sample for it to be considered meaningful and numerically stable (Gueorguieva et al., 2011; Spinhoven et al., 2016).

LCGA was performed on the included samples of patients receiving GSH or HI interventions, and conducted separately for the PHQ and GAD total score change over time. Following the

identification of a class solution for each symptom measure and intervention type, the conditional probability of class membership was extracted for each patient. This allowed every patient to be assigned to a class following the LCGA and meant an exploration of differences in the likelihood of each class between LPs could be performed.

Results.

Analysis 1.

Latent growth curve analysis - full sample (Guided self-help)

In the first stage of the analysis of patients receiving GSH, a linear LGC model was applied to the both the PHQ and GAD data. Model fit statistics suggested a poor fit for both these models (PHQ: CFI=0.908, TFI=0.917, RMSEA=0.078, SRMR=0.166; GAD: CFI=0.877, TFI=0.889, RMSEA=0.082, SRMR=0.182). The quadratic was then selected for both symptom measures and model fit was shown to improve, with a good fit indicated for both symptom measure models (PHQ: CFI=0.979, TFI=0.979, RMSEA=0.04, SRMR=0.032; GAD: CFI=0.977, TFI=0.976, RMSEA=0.038, SRMR=0.046).

In the final stage of analysis, the residuals of each time point were allowed to correlate. Adding correlated residuals to the quadratic model resulted in an excellent fit for the data (PHQ: CFI=0.994, TFI=0.992, RMSEA=0.025, SRMR=0.023; GAD: CFI=0.991, TFI=0.987, RMSEA=0.028, SRMR=0.039). Figure 7.2 presents the final curves for the PHQ and GAD for patients attending GSH. The graphs suggest a similar form of change on both symptom measures following GSH treatment.

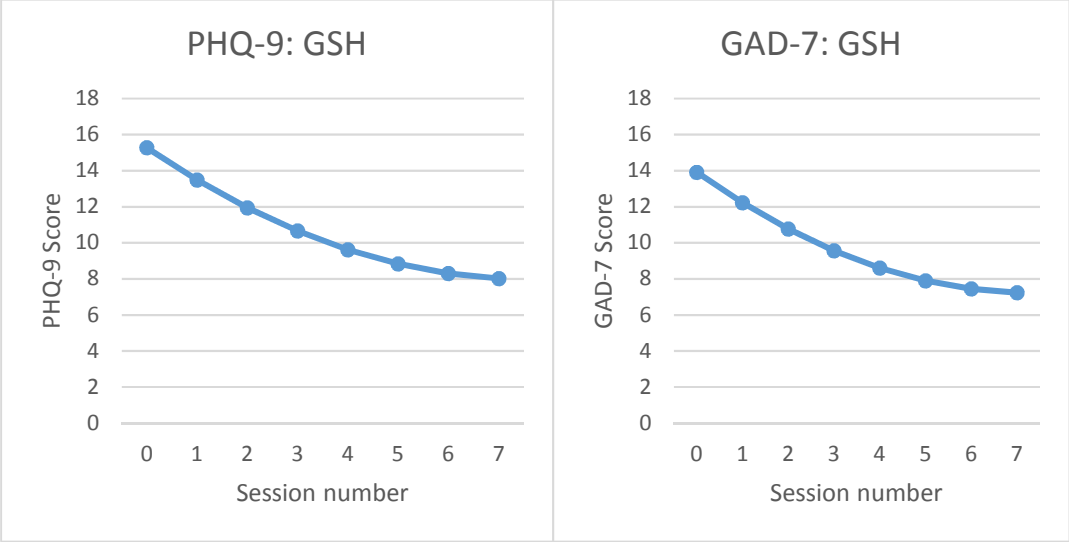

Figure 7.2. LGCs for PHQ-9 and GAD-7 - all patients (GSH).

Latent growth curve analysis - full sample (High Intensity)

Analysis was also performed for both the PHQ and GAD for all patients who received HI as their main interventions. The linear LGC displayed poor model fit, as seen previously with the GSH data (PHQ: CFI=0.915, TFI=0.923, RMSEA=0.07, SRMR=0.1; GAD: CFI=0.9, TFI=0.909, RMSEA=0.073, SRMR=0.118). The introduction of a quadratic curve to the model greatly improved fit (PHQ: CFI=0.963, TFI=0.965, RMSEA=0.047, SRMR=0.04; GAD: CFI=0.96, TFI=0.962, RMSEA=0.047, SRMR=0.045) and the further inclusion of correlated residuals provided excellent model fit (PHQ: CFI=0.988, TFI=0.987, RMSEA=0.029, SRMR=0.028; GAD: CFI=0.987, TFI=0.985, RMSEA=0.029, SRMR=0.034). The curves for HI patients are displayed in figure 7.3 and appear very similar between the PHQ and GAD.

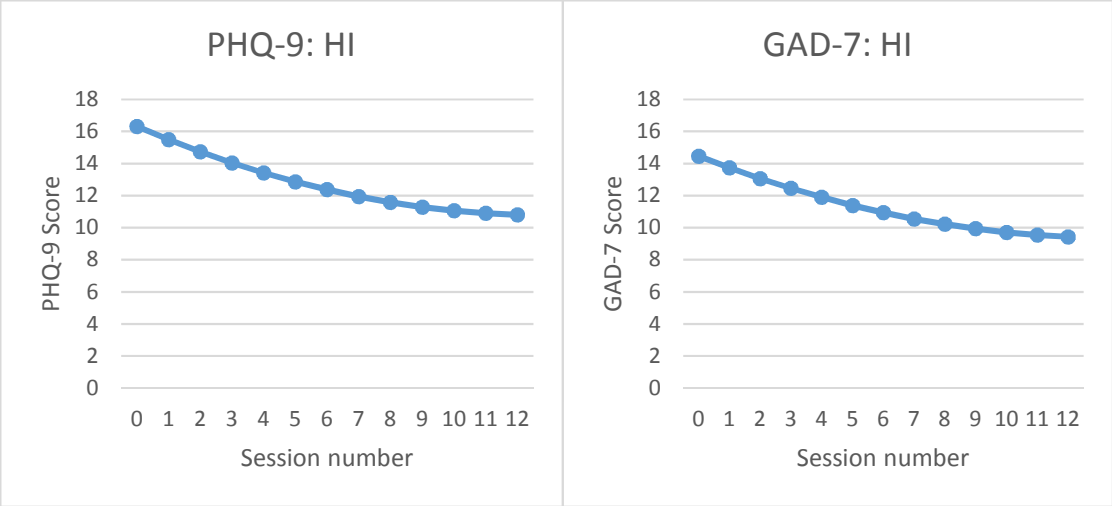

Figure 7.3. LGCs for PHQ-9 and GAD-7 - all patients (HI).

Latent growth curve analysis – Latent profiles (Guided self-help)

In the next stage of the analysis, the full sample was split into LP subgroups with LGCs modelled for each LP. As a quadratic curve with correlated residuals provided excellent fit for the full sample, LGC for each LP was performed specifying curves of the same structure. LP3 was not included due to the low number of patients available for inclusion in the analysis.

The LGC for each LP, as well as for the full sample, when GSH was the main intervention are presented in Figure 7.4, and information on the intercept, slope and quadratic value of each LP curve, are provided in Table 7.3.

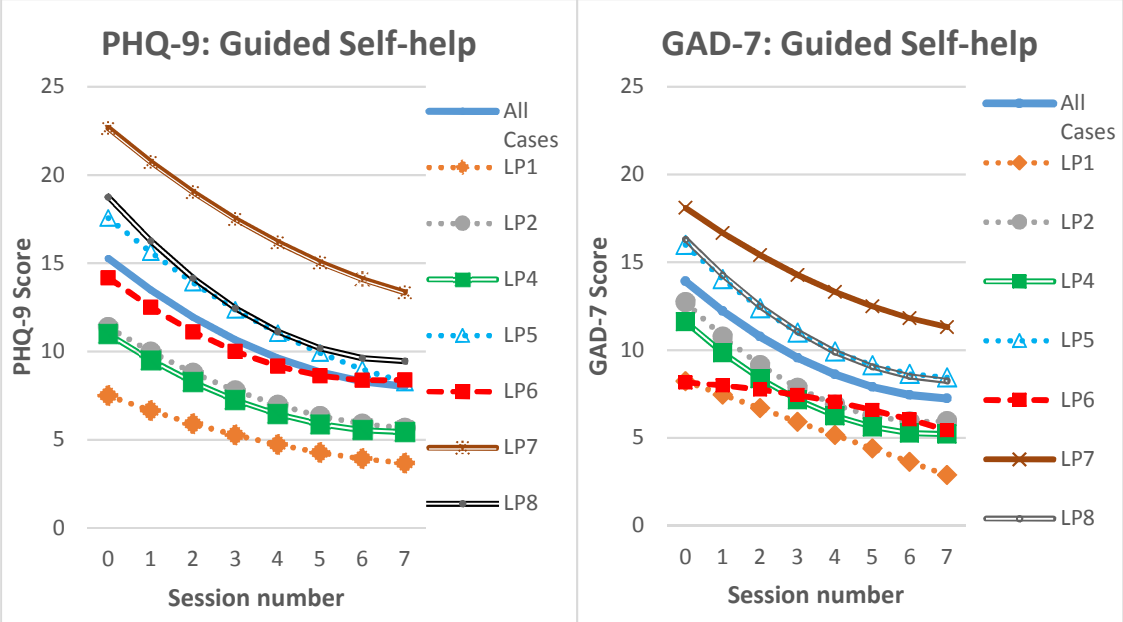

Figure 7.4. LGCs for PHQ-9 and GAD-7 – by latent profile (GSH)

Table 7.3. Intercept, slope and quadratic of LGC by LP (GSH)

PHQ							
LP	n	Intercept		Slope		Quadratic	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
All Cases	3334	15.249	*(15.07,15.43)*	-1.941	*(-2.04,-1.84)*	0.126	*(0.11,0.14)*
LP1	133	7.501	*(6.87,8.13)*	-0.896	*(-1.3,-0.49)*	0.05	*(-0.01,0.11)*
LP2	1112	11.373	*(11.19,11.56)*	-1.481	*(-1.63,-1.33)*	0.095	*(0.07,0.12)*
LP4	167	10.979	*(10.45,11.51)*	-1.583	*(-1.94,-1.2)*	0.113	*(0.05,0.17)*
LP5	363	17.558	*(17.21,17.91)*	-2.03	*(-2.35,-1.71)*	0.1	*(0.05,0.15)*
LP6	326	14.178	*(13.86,14.5)*	-1.82	*(-2.16,-1.48)*	0.142	*(0.08,0.2)*
LP7	325	22.659	*(22.36,22.96)*	-2.008	*(-2.34,-1.68)*	0.097	*(0.05,0.15)*
LP8	898	18.745	*(18.54,18.95)*	-2.678	*(-2.89,-2.46)*	0.193	*(0.16,0.23)*
GAD							
LP	n	Intercept		Slope		Quadratic	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
All Cases	3333	13.902	*(13.76,14.05)*	-1.848	*(-1.94,-1.75)*	0.126	*(0.11,0.14)*
LP1	133	8.223	*(7.74,8.71)*	-0.771	*(-1.18,-0.37)*	0.001	*(-0.07,0.07)*
LP2	1112	12.721	*(12.54,12.9)*	-2.117	*(-2.27,-1.96)*	0.164	*(0.14,0.19)*
LP4	167	11.611	*(-11.1,12.12)*	-1.908	*(-2.35,-1.46)*	0.142	*(0.07,0.22)*
LP5	363	15.998	*(15.71,16.28)*	-2.102	*(-2.4,-1.8)*	0.146	*(0.1,0.2)*
LP6	326	8.159	*(7.85,8.47)*	-0.129	*(-0.4,0.14)*	-0.037	*(-0.08,0.01)*
LP7	324	18.102	*(17.81,18.4)*	-1.502	*(-1.8,-1.21)*	0.076	*(0.03,0.12)*
LP8	898	16.312	*(16.13,16.5)*	-2.209	*(-2.4,-2.02)*	0.151	*(0.12,0.18)*

Observation of the graphs indicates that the general shape of LGCs are similar between LPs, with the main difference being the intercept and the gradient of the slopes. By examining Table 7.3, the largest variation between LPs is with regard to the intercept value, which would be expected as the probability of LP membership is derived from both the PHQ and GAD score. Therefore, the intercept for each LGC would be expected to be similar to the mean baseline symptom measure score for that profile. It should be noted that although caseness was criteria for inclusion in the analysis, the intercept for PHQ in LP1 was below the clinical cutoff (10), which would be explained by patients scoring caseness on the GAD-7 but not the PHQ-9 still being eligible for inclusion. The slopes, or gradients of change, appear to be larger for LPs with higher intercepts, which could in part be explained by regression to the mean as described in Chapter 4, as well as these groups of patients having more available change on the symptom measures.

The quadratic functions follow a very similar pattern for all LPs across symptom measures for patients receiving guided SH interventions, with a decline in symptoms followed by a levelling out of the slope as time increases. However, the curve for LP6 GAD change stands out as the only LGC where the quadratic is reversed – the curve starts flat initially before showing a decrease (improvement) towards later time points. This is interesting when considered alongside the deterioration findings presented in Chapter 6, as LP6 showed around 20% deterioration, significantly higher than all other LPs. It would suggest that change in anxiety symptoms is different for this profile of patients.

Latent growth curve analysis – latent profiles (high intensity)

The LGC for each LP when HI is the main intervention are presented in Figure 7.5, and information on the intercept, slope and quadratic value of each LP curve, are provided in Table 7.4.

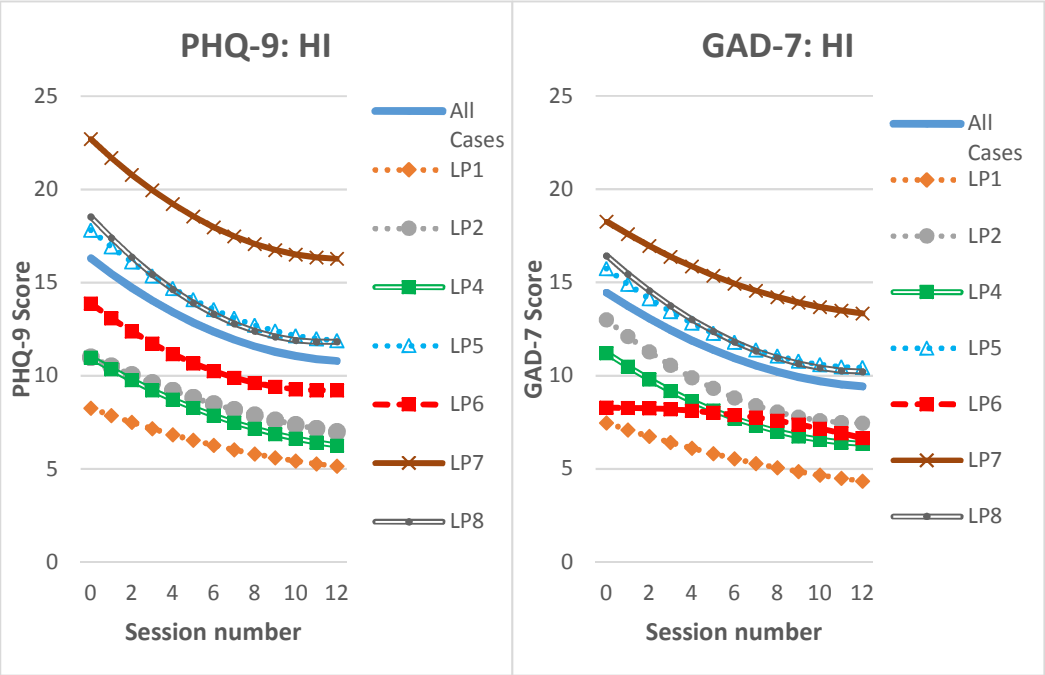

Figure 7.5. LGCs for PHQ-9 and GAD-7 – by latent profile (HI).

The curves presented for HI interventions in Figure 7.5 appear very similar to those of GSH in Figure 7.4, despite the potentially more intensive intervention being received in the Step 3 treatment. It would be predicted that individuals receiving HI interventions would have more severe or complex clinical presentations and therefore require the more intensive intervention and more sessions to achieve a positive outcome. The variation in intercepts between LP is similar between the GSH and HI analysis, as is the same shape of quadratic curve of each LP, which suggests that the available symptom measures are limited as a measure of severity or complexity. LP6 displays the same reversed quadratic change curve with GAD measurement over time as found for LP6 patients in receipt GSH. The main difference between the results of the GSH and HI LGCs, however, is that the change in slope is shallower for HI interventions compared to GSH. This could be because more available time points were included in this analysis compared to GSH, resulting in a shallower curve of mean change. If one were to assume that more complex and severe patients received the more intensive HI interventions then more shallow curves of change might be expected as the patient's response to treatment may be slower. However, it is also possible that a number of patents, for example in LP7, would not benefit for IAPT delivered interventions and therefore the choice of treatment may not have been optimal, with referral

to alternative secondary care services potentially increasing the chance of positive outcomes.

Table 7.4. Intercept, slope and quadratic of LGC by LP (HI)

PHQ							
LP	n	Intercept		Slope		Quadratic	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
All Cases	4394	16.309	*(16.14,16.48)*	-0.855	*(-0.91,-0.81)*	0.033	*(0.03,0.04)*
LP1	135	8.269	*(7.65,8.89)*	-0.403	*(-0.61,-0.19)*	0.012	*(-0.00,0.03)*
LP2	1052	11.003	*(10.8,11.21)*	-0.502	*(-0.59,-0.41)*	0.014	*(0.01,0.02)*
LP4	180	10.984	*(10.47,11.5)*	-0.645	*(-0.86,-0.43)*	0.021	*(0.00,0.04)*
LP5	453	17.816	*(17.48,18.15)*	-0.926	*(-1.09,-0.76)*	0.036	*(0.2,0.05)*
LP6	487	13.876	*(13.59,14.17)*	-0.819	*(-0.95,-0.68)*	0.036	*(0.03,0.05)*
LP7	924	22.698	*(22.5,22.9)*	-1.039	*(-1.15,-0.93)*	0.042	*(0.03,0.05)*
LP8	1149	18.531	*(18.33,18.74)*	-1.182	*(-1.29,-1.08)*	0.052	*(0.04,0.06)*
GAD							
LP	n	Intercept		Slope		Quadratic	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
All Cases	4394	14.46	*(14.33,15.6)*	-0.755	*(-0.8,-0.7)*	0.028	*(0.02,0.03)*
LP1	135	7.472	*(6.96,7.99)*	-0.38	*(-0.6,-0.17)*	0.01	*(-0.00,0.03)*
LP2	1052	12.996	*(12.81,13.18)*	-0.93	*(-1.02,-0.86)*	0.039	*(0.03,0.05)*
LP4	180	11.224	*(10.72,11.73)*	-0.766	*(-0.99,-0.53)*	0.03	*(0.01,0.05)*
LP5	453	15.745	*(15.47,16.02)*	-0.874	*(-1.01,-0.74)*	0.036	*(0.03,0.05)*
LP6	487	8.293	*(8.04,8.55)*	0.008	*(-0.11,0.12)*	-0.12	*(-0.02,-0.00)*
LP7	924	18.267	*(18.09,18.44)*	-0.698	*(-0.79,-0.61)*	0.024	*(0.02,0.03)*
LP8	1149	16.439	*(16.26,16.62)*	-1.022	*(-1.11,-0.93)*	0.042	*(0.04,0.05)*

The LGC analyses describe the expected response curves for both PHQ and GAD change for the full sample of patients receiving either GSH or HI interventions, as well as by individual LP. There is potential for these graphs to be used by clinicians to compare against the known session by session change for current patients. Generally, the LGCs all showed a similar structure, and a reasonable rate of change was estimated initially, which then flattened out over time points (sessions), which replicated the form of change typically reported in other studies of psychological interventions (e.g. Kopta, et al., 1994). However, LP6 displayed a different trajectory for anxiety symptoms change during both GSH and HI interventions, and the LGC of this profile indicated that initial anxiety change is limited and that a reduction in anxiety symptoms was more likely to occur at later sessions for this profile. The identification of LP6 at initial assessment could therefore be used to inform the clinician that a reduction in anxiety symptoms will likely take longer for this profile of patients. It could be that members of LP6 represent a more avoid group of patients entering IAPT treatment, and that the mean expected response curves is influenced by a number of

patients who may display an initial increase in anxiety symptoms, possibly in response to behavioural experiments.

The LGCs presented in this analysis estimate that the mean change in symptom scores during LI and HI IAPT treatments are suggestive of a reduction in symptoms over the course of treatment. However, as shown in Chapter 6 not all members of profiles responded to treatment, and a number of patients reported a significant increase in symptoms that would be follow the expected response curves presented in these LGC. As a result, it could be theorised that there may be two or more different trajectories of change, representing groups of patients who respond differently to the same treatment, whether that be responders or non-responders (Gueorguieva et al., 2011) or potentially a more varied collection of patient trajectories (e.g. such as identified by Stulz et al., 2007). The next stage of the analysis was to explore whether different classes of trajectories existed in the IAPT dataset using LCGA.

Analysis 2.

Latent class growth analysis – Guided self-help

The LGC analysis for both the intensity of intervention and symptom scores found that a model specifying a quadratic curve and the inclusion of correlated residuals was the best fitting model in all scenarios. As a result, this model was specified within all LCGA analyses, as these are an extension of the previous LGC analyses.

Model fit statistics for the LGCA including GSH patients with both PHQ and GAD repeated measures are presented in Table 7.5. The first section of the table displays fit statistics for the PHQ analysis, and the VLMR-LRT p-value remained <.05 until the 5-class model (p=0.276), suggesting the 4-class model was a better fit for the data. For the GAD analysis, the 5-class model was a better fit compared to the 4-class solution (VLMRT-LRT p<0.001), but the 6-class model was not a better fit than the 5-class model (p=0.266). The decrease in AIC and BIC value became shallower by the 5-class solution and therefore this model was selected for GAD change in GSH interventions.

Table 7.5. Model fit statistics: LCGA patients receiving GSH.

PHQ-9						
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**	**% individuals in per class**
k = 2	90549	90684	90614	<0.001	0.809	65/35
k = 3	89769	89928	89845	<0.001	0.748	16/47/37
k = 4	**89478**	**89662**	**89566**	**<0.001**	**0.701**	**31/9/24/35**
k = 5	89369	89577	89469	0.276	0.668	9/33/31/14/13
GAD-7						
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**	**% individuals in per class**
k = 2	86899	87034	86964	<0.001	0.79	36/64
k = 3	86382	86541	86459	<0.001	0.7	43/22/35
k = 4	86216	86399	86304	0.02	0.644	15/25/25/35
k = 5	**86133**	**86341**	**86233**	**0.007**	**0.62**	**26/16/23/22/14**
k = 6	86057	86289	86169	0.258	0.602	14/20/25/15/16/10

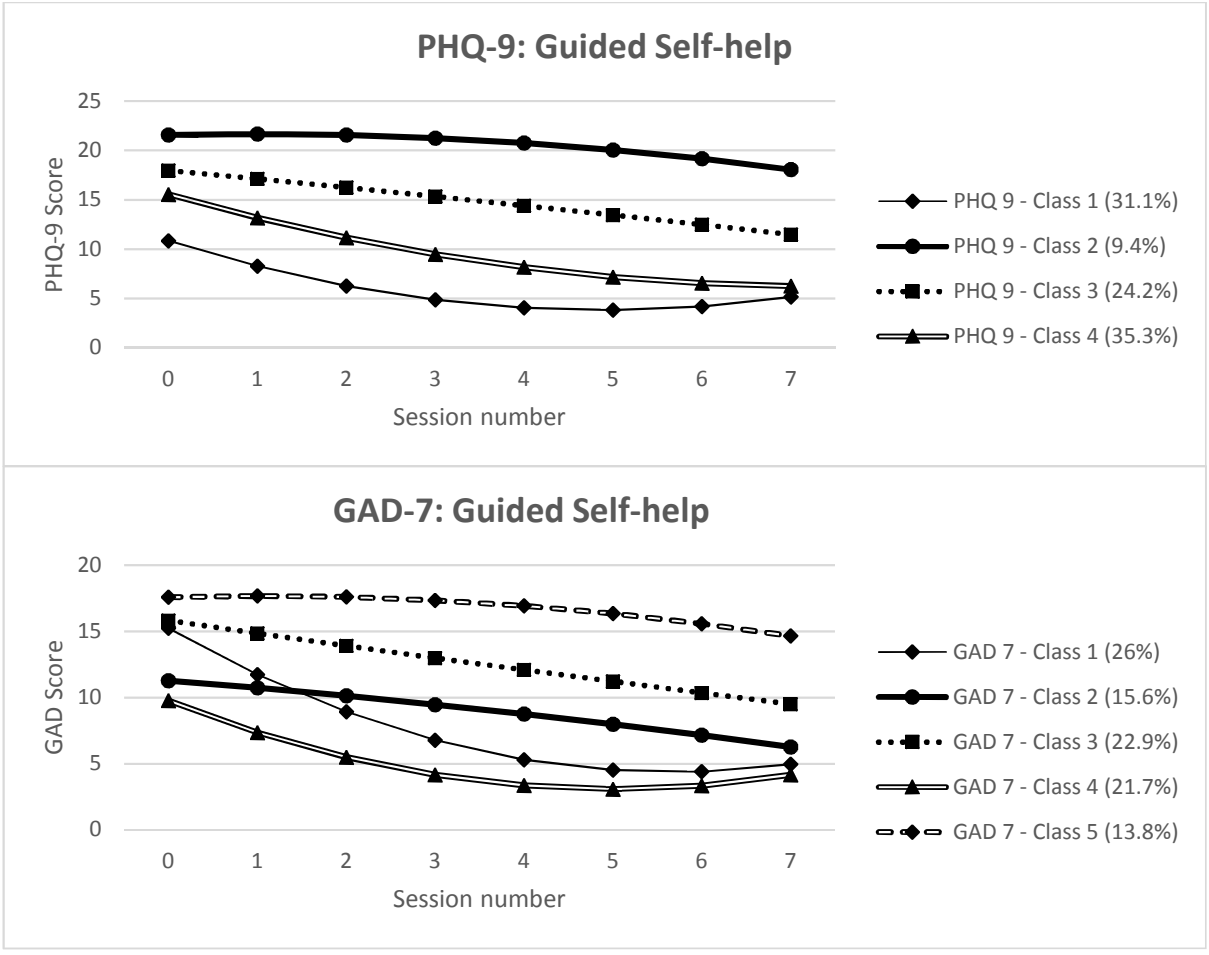

Figure 7.6. LCGA curves for PHQ and GAD - all patients receiving GSH

The identified curves are presented in figure 7.6, with four classes identified for PHQ change, and five classes of GAD score change.

The PHQ classes can be described as follows:

-Class 1 (PHQ) – Early initial response, levelling out during treatment.

-Class 2 (PHQ) – Limited or no response to treatment

-Class 3 (PHQ) – Gradual improvement to treatment

-Class 4 (PHQ) – Early initial response, continued improvement during treatment

With the GAD classes described as:

-Class 1 (GAD) – Early initial response, continued improvement during treatment

-Class 2 (GAD) – Gradual improvement to treatment (less severe)

-Class 3 (GAD) – Gradual improvement to treatment (more severe)

-Class 4 (GAD) – Early initial response, levelling out during treatment.

-Class 5 (GAD) – Limited or no response to treatment

The description of the classes appears similar between the two symptoms measures. The analysis suggests that there is one additional GAD trajectory class (class 3), which could be described as being similar to another GAD class (class 2) but with a higher intercept (higher symptom severity at initial assessment). This would suggest there are two statistically different classes of patients who show a similar form of change, but with different intercept values.

Both symptom measures include a trajectory class showing little, if any improvement in clinical symptoms during treatment (class 2 PHQ, class 5 GAD), which could be assumed to be non-responders. Although stepped-up patients were not included in the analysis, some LI patients may have benefited from being stepped up by the service from LI GSH to HI interventions during treatment, and it would be of interest to know why this decision was not made during the episode of care for patients included in this dataset. Previous researchers have suggested that a lack of change within the first three sessions of treatment are predictive of eventual treatment outcome (Gorwood et al., 2013; Lambert, 2013), and the graphs presented in figure 7.6 would indicate that these two non-responding classes could be identified within the first three sessions. A potential DST for use in IAPT could flag patients who are following these non-responding classes, and this information could be

provided to clinicians to inform a discussion with the patient as to whether stepping up to a HI intervention might be preferable, as it may increase the probability of a positive outcome.

Following the identification of the classes, the association between LP membership and the prevalence of trajectory classes was explored. Table 7.6 presents the proportion of individuals from each LP that were members of the PHQ classes identified. The results suggest that each LP is predominately made up of members of two PHQ classes each for patients receiving GSH. For example, 54% of LP2 patients are members of class 1 (early improvement, levelling out), and 36% class 4 (early improvement, continuing during treatment). LP2 membership to either remaining classes was limited to 10% members of class 3 (gradual improvement). The results for LP5 suggest the majority of patients in this profile are members of either PHQ class 3 or class 4. Most individuals from LP1 and LP3 were members of PHQ class 1, which may be due to this class having the lowest intercept value of all the PHQ classes.

Table 7.6. Proportion of members to each PHQ class by LP (GSH).

LP	PHQ - Class 1		PHQ - Class 2		PHQ - Class 3		PHQ - Class 4		Total
	n	%	n	%	n	%	n	%	n
1	108	81%	0	0%	0	0%	25	19%	133
2	597	54%	4	0%	109	10%	402	36%	1,112
3	8	80%	0	0%	0	0%	2	20%	10
4	100	60%	1	1%	16	10%	50	30%	167
5	44	12%	33	9%	141	39%	145	40%	363
6	94	29%	8	2%	77	24%	147	45%	326
7	2	1%	150	46%	117	36%	56	17%	325
8	77	9%	104	12%	339	38%	378	42%	898
Total	1,030	31%	300	9%	799	24%	1,205	36%	3,334

It is of interest that PHQ class 2, a group of patients showing little if any response to treatment, is the identified trajectory of nearly 50% of LP7, whereas all other LPs contain less than 12% members of this class. As LP7 have such a low probability of recovery (11% to LI interventions) the finding that so many patients from this profile are members of class 2 may not be surprising and may suggest that this profile of patients are a clinically relevant group to identify early in the assessment process in order to inform clinical decisions. The finding that 9% of LP5 and 12% of LP8 were also members of PHQ class 2 suggests that it is not just members of LP7 who have poor response to treatment but that there are some similarities with patients from other profiles. These common characteristics between members of different LPs could include characteristics not currently included in the dataset but identified as having a potential association with outcome such as substance misuse or personality disorders discussed in chapter 3.

The proportion of each LP that are members of each GAD class are displayed in Table 7.7. Membership to GAD classes shows more variation within and between LPs, which could be explained by the range of anxiety disorders (for example OCD, social phobia, GAD) that may be present within the dataset. Certain LPs show more than 15% membership of at least 4 classes, for example LP2, suggesting that there are a number of different likely trajectories for this profile, which may be of value to explore further. Class 5, the group of patients showing little response to treatment, rarely includes members of any profile except LPs 5, 7 and 8, which happen to be the profiles with poorest outcomes. The findings suggest that for other profiles the GAD scores are expected to change during GSH treatment, either gradually over sessions or showing early improvement that levels out when symptom scores fall into the non-clinical range.

Table 7.7. Proportion of members to each GAD class by LP (GSH).

LP	GAD - Class 1		GAD - Class 2		GAD - Class 3		GAD - Class 4		GAD - Class 5		Total
	n	%	n	%	n	%	n	%	n	%	n
1	7	5%	34	26%	4	3%	88	66%	0	0%	133
2	340	31%	199	18%	195	18%	343	31%	35	3%	1,112
3	1	10%	1	10%	0	0%	8	80%	0	0%	10
4	43	26%	39	23%	19	11%	62	37%	4	2%	167
5	123	34%	20	6%	132	36%	23	6%	65	18%	363
6	20	6%	96	29%	24	7%	176	54%	10	3%	326
7	59	18%	5	2%	98	30%	1	0%	161	50%	324
8	308	34%	45	5%	326	36%	46	5%	173	19%	898
Total	901	27%	439	13%	798	24%	747	22%	448	13%	3,333

Results for some LPs were similar to those for the PHQ curves, as these LPs show a higher likelihood of membership to one or two specific classes of GAD curve. For example, over 80% of individuals from LP1 and LP6 are members of either class 2 (gradual improvement, low severity) or class 4 (early improvement, levelling out). As with the PHQ classes, it may be that these profiles are more likely to have a high frequency of membership to trajectory classes that have lower intercept values and are therefore closer to the mean GAD score of these profiles.

Latent class growth analysis – High intensity

Model fit statistics comparison for the LGCA of HI patients with both PHQ and GAD repeated measures is presented in Table 7.8. The first section of the table shows fit statistics for PHQ and suggests that the VLMR-LRT remained significant until the 6-class model, indicating that the 5-class model may be a better fit for the data. However, class 1 of the 5-class solution was made up of less than 3% of the included sample, which suggests it may not represent a sub-group which are meaningfully differentiated from the other classes in the LCGA analysis. As a result, it was decided to retain the 4-class solution. The results for the GAD LCGA analysis showed that the 5-class model was a significantly better fit compared to the 4-class model VLMRT-LRT (*p*=0.039), and as the decrease in AIC and BIC values was becoming shallower, the 5-class solution was selected.

Table 7.8. Model fit statistics: LCGA patients receiving HI.

PHQ-9						
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**	**% individuals in per class**
k = 2	198731	198942	198837	<0.001	0.556	39/61
k = 3	198254	198490	198373	0.002	0.557	33/45/22
k = 4	**197868**	**198130**	**197999**	**<0.001**	**0.598**	**45/13/14/28**
k = 5	197759	198046	197903	0.011	0.629	3/14/43/13/28
k = 6	197701	198014	197858	0.438	0.585	22/28/13/13/20/3
GAD-7						
k model	**AIC**	**BIC**	**Adj-BIC**	**VLMR-LRT (p=)**	**Entropy**	**% individuals in per class**
k = 2	195067	195271	195169	<0.001	0.847	53/47
k = 3	191764	191994	191879	<0.001	0.824	41/27/32
k = 4	190762	191018	190891	<0.001	0.772	21/30/30/19
k = 5	**190440**	**190721**	**190581**	**0.039**	**0.739**	**21/29/17/21/12**
k = 6	190064	190371	190218	0.375	0.715	20/12/13/20/18/17

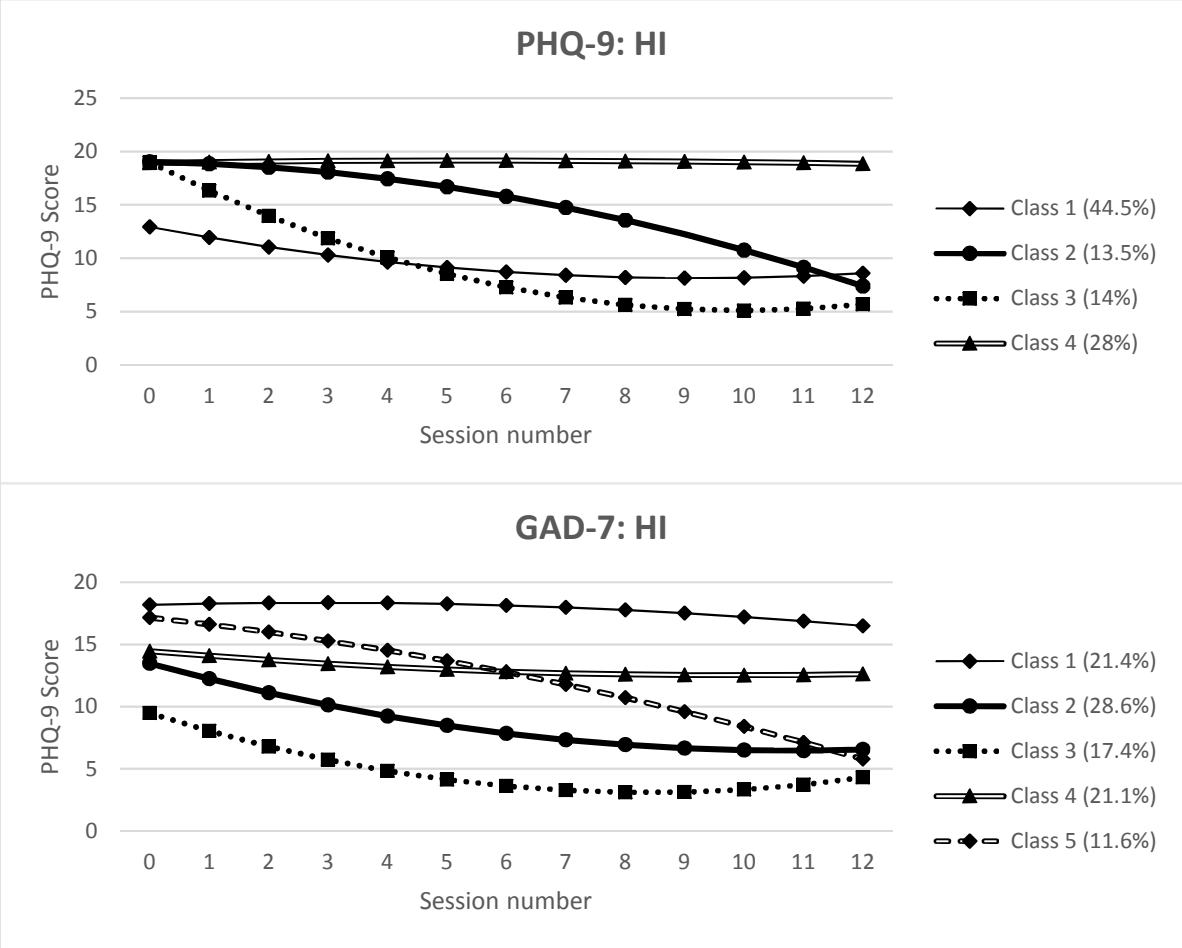

Figure 7.7. LCGA curves for PHQ and GAD - all patients receiving HI.

The identified curves are presented in figure 7.7 and show four statistically differing groups of patients based on PHQ change over time, and five classes identified for GAD scores. The PHQ classes can be described as follows:

-Class 1 (PHQ) – Early initial response, levelling out during treatment.

-Class 2 (PHQ) – Slow initial response, large response later in treatment.

-Class 3 (PHQ) – Rapid improvement, levelling out during treatment.

-Class 4 (PHQ) – Limited or no response to treatment.

With the GAD Classes described as:

-Class 1 (GAD) – Limited or no response to treatment (more severe).

-Class 2 (GAD) – Early initial response, levelling out. (more severe)

-Class 3 (GAD) – Early initial response, levelling out. (less severe)

-Class 4 (GAD) – Limited or no response to treatment (less severe).

-Class 5 (GAD) – Slow initial response, large response later in treatment.

Although the same number of trajectory classes were identified in both the GSH and HI samples for each symptom measure (4 and 5), there are some differences in the trajectories identified. For PHQ change, the four classes identified for GSH indicate four curves with increasing intercept values, described as 'early response that levels out', 'early response that continues', 'gradual improvement' and finally 'limited improvement'. Of the four classes identified for PHQ change in HI, three classes have a similar intercept value but statistically distinct forms of change. Compring the five trajectory classes for GAD change between HI and GSH interventions suggests a few differences in the forms of change identified. The HI trajectory classes include two 'limited response' curves with differing intercept values, whereas there appears to be only one non-responding class in GSH interventions.

A further difference is that the early improving group who continue to improve during treatment during GSH do not appear in HI, and instead a group of patients who show little change symptoms at first, then appear to show response in the later part of HI treatment was identified. The difference between interventions in these two curves may be linked to the number of included time points, and it is possible that if more time points were included for GSH then a similar curve may have been identified. Alternatively, if patients receiving HI interventions were likely to have more complex clinical presentations then they may have taken a little longer to engage and therefore show improvement, which may be evident in HI PHQ class 2.

Table 7.9 presents the proportion of individuals from each LP who were members of the PHQ classes identified when receiving HI interventions. The proportion of individuals from each LP who are members of Class 1 (early improvement, levelling out) is very high for LPs 1, 2, 3, 4 and 6, which may be partly explained by the lower intercept value of this class compared to the other three classes 2, 3 and 4. The mean T1 PHQ scores of these LPs are similar or lower than the intercept of class 1 (PHQ-9 score of 12.9), which may increase the likelihood of patients from these profiles being members of this trajectory. Class 1 is also the most prevalent class for LP5 and LP8, although these profiles also contain members of the other three classes in similar levels. Patients in LP7 are predominately members of Class 4,

the non-responding group, which is likely explained by the low probably of recovery for this profile following HI treatment (18% recovery).

Table 7.9. Proportion of members to each PHQ class by LP (HI).

LP	PHQ - Class 1		PHQ - Class 2		PHQ - Class 3		PHQ - Class 4		Total
	n	%	n	%	n	%	n	%	n
1	125	93%	2	1%	1	1%	7	5%	135
2	847	81%	32	3%	57	5%	116	11%	1,052
3	11	79%	1	7%	2	14%	0	0%	14
4	146	81%	7	4%	11	6%	16	9%	180
5	176	39%	52	11%	67	15%	158	35%	453
6	332	68%	28	6%	34	7%	93	19%	487
7	103	11%	142	15%	134	15%	545	59%	924
8	412	36%	151	13%	221	19%	365	32%	1149
Total	2,152	49%	415	9%	527	12%	1,300	30%	4,394

The proportion of each LP who are members of the identified GAD classes when receiving HI interventions are presented in Table 7.10. More variation in class membership within and between LPs seems to exist for the GAD results, compared to the PHQ trajectories. LP2, LP5 and LP8 show fair representation of three classes (>=20% prevalence) which would indicate that clinicians couldn't necessarily predict the expected trajectory of change for these individuals at the start of treatment, and monitoring symptoms during treatment would be needed. Class 3 (early response, levelling out) occured in 17% of the full sample but LP1, LP3 and LP6 show a disportionately high number of members of this class, suggesting an increased incidence of treatment response for these profiles.

Table 7.10. Proportion of members to each GAD class by LP (HI).

LP	GAD - Class 1		GAD - Class 2		GAD - Class 3		GAD - Class 4		GAD - Class 5		Total
	n	%	n	%	n	%	n	%	n	%	n
1	0	0%	34	25%	91	67%	9	7%	1	1%	135
2	72	7%	443	42%	257	24%	211	20%	69	7%	1,052
3	0	0%	0	0%	13	93%	1	7%	0	0%	14
4	6	3%	82	46%	59	33%	25	14%	8	4%	180
5	104	23%	129	28%	34	8%	140	31%	46	10%	453
6	11	2%	166	34%	214	44%	89	18%	7	1%	487
7	504	55%	112	12%	20	2%	178	19%	110	12%	924
8	285	25%	342	30%	66	6%	288	25%	168	15%	1149
Total	982	22%	1308	30%	754	17%	941	21%	409	9%	4,394

Class 5 (slow initial response, greater response in later treatment) occured relatively infrequently (9% of the full sample) but is more common in LP5, LP7, LP8, which may be due to the higher mean GAD score at T1 for these profiles. The non-responding high severity class (class 1) also show representation in LP5, LP7 and LP8, with little or no representation in any of the other 5 LPs, which again may be due to the intercept value. These two trajectories result in very different outcomes for patients, and being able to predict which patients of these three profiles are more likely to follow the poor outcome could be used to inform both treatment selection and treatment monitoring decisions. Class 4 (limited response, less severe) is spread relatively evenly across LPs, except for LPs 1 and 3 where it is very infrequent. Class 2 (gradual sustained response) is also relatively spread amongst the LPs, except LP3 and LP7.

Discussion

The findings from this chapter provide detailed information about the trajectories of symptom score change during the course of GSH or HI treatment in IAPT services. The LGC analyses could be adapted to provide expected response curves for patients receiving IAPT delivered treatment. Following the work of Lambert et al (2001) the LGC analyses could be used to identify patients that are 'not on track' to recovery, and this information could be feed back to the treating clinician so they can consider adjusting the treatment plan accordingly. By modelling the LGCs by individual LP, a more refined expected response curve can be provided, potentially increasing the accuracy of the estimates for individual patients and the utility of the LGCs. The LCGA identified four statistically distinct classes of patients based on their PHQ response curves, and five classes of GAD response curves for both intensities of intervention. The information provided could be used to further inform decisions during treatment, as they extend the information provided by single LGCs to show the likely outcome for different trajectories.

The LGC analyses showed that for both symptoms measures, across both treatment intensities, the expected response indicated a greater effect of treatment (decrease in symptoms) in the first few sessions, with the effect reducing further into treatment. The shape of these curves is very similar to other expected response curves identified in psychological interventions (Clapp et al, 2013; Kopta et al, 1994). The analysis of LGC by each LP indicated that the shape of the curves was similar between profiles. Only the GAD curve for LP6 patients suggested a different shape to the mean trajectory across all patients, whereby the change was shallow initially and then decreased more during later treatment sessions. As LP6 showed a very high probability of deterioration (17% LI, 21% HI) compared to the other LPs (full sample 8% LI, 9% HI), it may be that this expected response is influenced by a significant number of patients whose anxiety increased initially and either

reduced in later sessions or continued to deteriorate during treatment. This could be due to heightened anxiety in the first few sessions due to initial exposure work that is integral to some interventions for anxiety disorders, for example phobia (Rapee & Heimberg, 1997). As 46% of LP patients recovered from treatment, a large number of patients benefitted from treatment and therefore identifying the sub-group of LP6 patients who are at risk of deterioration could inform clinical decisions for this profile. Although researchers have often suggested that change in the first few sessions is associated with eventual outcome from psychological interventions (Lambert, 2013; Lewis, et al., 2012), the findings for LP6 suggest that initial change may not always be predictive of eventual outcome.

The main difference between the LGCs for individual LPs was the intercept of the curves, which can be explained by the different mean PHQ-9 or GAD-7 scores at T1 associated with each LP. The intercept value of the LGC for each LP was very similar to the mean PHQ and GAD score for that LP, which would be expected. Any variations between the measure symptom measure scores (Chapter 5) and intercept values (Chapter 7) is likely due to the reduced sample of included patients in the latent growth analyses, for example in LP1 where many patients who were not caseness at assessment were excluded from the analysis. The potential utility of these individual curves for each LP is that the expected response curves could be combined with information from the estimation of profile membership. In the first stage of a decision support algorithm, the patient can be allocated to their appropriate LP using the calculation of posterior probabilities described in Chapter 5. In the second stage, the relevant expected response curves for PHQ and GAD change can be provided to the clinician (potentially through the EPMS) once the LP is identified, and the DST could provide the expected trajectory information throughout the episode of care. This could support the development of a personalised treatment plan, whereby a DST could aid both patient and clinician in the treatment selection and treatment monitoring decisions.

The identification of statistically distinct trajectory classes from the LCGA could be used to inform further decisions during the monitoring of treatment progress, especially if feedback on an active patient's progress could be provided alongside the identified trajectories. The difference between the identified classes of PHQ change for GSH and HI interventions is that for GSH there are the four classes which had distinct intercept values that could be seen to equate to their eventual trajectory, whereas a difference in intercepts is not clear with three of the HI PHQ classes. The LCGA results suggest that at around session three of GSH treatment, most patients will either be following one of the early improving classes (class 1 or class 4), the trajectory with more gradual change (class 3) or show limited response to treatment. This supports findings by researchers that eventual outcome can be predicted by the third treatment session in psychological interventions (Lambert, 2013), and the first two weeks exposure to pharmacological intervention (Gorwood et al., 2013; Szegedi et al., 2009). This could be used in practice to inform decisions during treatment, for

example if by session three of GSH treatment there has been limited change (PHQ class 2, GAD class 5) then potentially stepping up might increase the likelihood of positive outcomes.

However, the identified classes for PHQ change when HI interventions were received included three classes with a very similar intercept values but differentiated forms of change, suggesting that although patients with these three trajectories start with similar levels of severity, their response to treatment is very different. The graph in figure 7.7 shows that members of class 3 are likely to be distinguishable within the first few sessions due an immediate decline in depression symptom severity, but the trajectories of class 2 and class 4 appear similar until around the 4th or 5th session. The graph suggests that at this point, members of class 2 are more likely to have a sustained decline in depression symptoms, whereas individuals in class 4 will likely continue to report a limited change in depression symptoms. These findings therefore do not support the 'three session' rule, and instead suggest that the expected trajectory of patients receiving HI treatments may not be distinguishable until at least the fourth sessions. It may be of interest to further explore class 2 individuals to understand what in treatment contributed to this sudden improvement in functioning. Identifying treatment factors that are associated with these sudden gains has the potential to inform the development and refinement of psychological interventions, and it may be worth exploring additional IAPT datasets to determine this finding is generalizable to other settings.

The findings of these analyses suggest that there are potential differences in the likely trajectory of symptom change between the two types of intervention delivered. For some patients it is likely that the chosen intervention was not appropriate, and for others, especially from profiles with low probabilities of recovery both at LI and HI in IAPT (e.g. LP7) it is possible that treatment at alternative treatments should be considered. Combining information from the probability of outcomes by profile with likely trajectories of change once treatment has commenced could be used to inform both treatment choice but also care planning through routine monitoring.

The results of the LCGA analyses presented in this chapter have similarities to previous GMM analyses performed in CMHD treatment sample, which might suggest that the results of the current analysis could be of value to mental health services beyond the IAPT services who provided data. Stulz et al (2007) identified five distinct classes from a dataset of patients receiving a range of psychological interventions over six sessions of treatment, the same number of classes identified for the current LCGA of GAD score change. The response curves presented by Stulz et al are similar to the graph of GSH GAD presented in figure 7.6, with one rapidly improving class and four other classes with more level change curves, although limited decrease in symptoms is suggested by the graph presented by Stulz et al.

The current analyses have some advantages over the analysis conducted by Stulz et al. The included sample for their study was limited to just n=192 patients receiving more than 7

sessions and with three or more outcome measures completed, which is considerably smaller than the samples used for the analysis presented in this chapter. The lack of change observed in four of the five classes identified by the previous researchers may be due to the inclusion of a wide range of therapeutic models compared to the mainly CBT based approach used in the IAPT dataset. The psychological interventions delivered by the mental health services included in the Stulz et al (2007) analysis included a wide range of treatments (including cognitive therapy, psychodynamic therapy, gestalt therapy and transactional therapy in addition to other integrative treatments) and it may be that these interventions have different efficacy from GSH delivered in IAPT services. As the authors' inclusion criteria required patients to have had at least seven sessions, and number of patients who benefitted from treatment early and therefore completed by the sixth session may have been excluded, which may have altered the findings of this study.

Despite the differences in the samples used in the analyses, as the same number of trajectories were identified, this may suggest a prognostic ability across a range of psychological treatment settings. Further research across multiple datasets of patients receiving routine psychological services may validate these differential responses to treatment. Nevertheless, the results of the current analysis are quite different to those from the GMM presented by Gueorguieva and colleagues (2011) who indicated that depression change followed two distinct classes in their dataset compared to four depression trajectory classes identified in the current analysis, for both GSH and HI change. The similarities between the results of the current analysis and those of Stulz et al (2007) and the differences with Gueorguieva et al (2011) may be explained by the samples and the range of interventions included in the analyses. The dataset used by Gueorguieva et al was taken from controlled trials of antidepressants and therefore strict inclusion criteria was used for patients to be eligible to receive the treatment. As a result, the initial symptom severity levels for all patients was similar, which is indicated by the similar intercept value of the two classes the authors identified. By comparison, the samples used by the current analysis, and the analysis by Stulz et al, used routine data from mental health services and therefore there was more variation between patients on symptom severity that was used to model change, as well as diagnoses. These analyses may therefore better inform routine clinical practice, having more ecological validity than the selected samples from randomised controlled trials, but the variation in initial symptom scores may be important when considering the use of GMM type analyses, as this may contribute to the number of curves identified in the dataset.

When the distribution of LCGA classes was explored by LP, it was found that there were usually two main trajectories of PHQ change, and two or three GAD curves associated with each LP. For certain LPs, there was one class that was endorsed by the majority of profile members, for example LP7 where the majority of individuals were members of the limited response to treatment classes, and LP1 where most patients were members of the early improving classes with good outcomes. Other LPs have members of many classes, which

suggests even the identification of LP at initial assessment could not be used inform the likely trajectory of change, and instead the monitoring of symptoms would be required to determine prognosis.

Although the distribution of each LGCA class varied between LPs, it also appears that a number of the differences in class membership may be due to the mean symptom scores (intercept values) of the trajectory classes. For example, if the graph of PHQ classes in HI treatment patients (figure 7.7) is observed, there are three classes with a similar intercept value (around a score of 19 on the PHQ) and one class (class 1) with a much lower intercept value (around 13). Class 1 was the most likely class for 49% of the entire sample of patients receiving HI treatment, and the most likely class for LP1, LP2, LP3, LP4 and LP6. This is not surprising as the intercept value for this class is around the mean, or below, for each of these five classes. It may be of more value to conduct LCGA by latent profile, instead of using the full sample, as this may provide a more patient-centred information for supporting treatment monitoring decisions.

One potential limitation of the analysis is that diagnostic specific measures, especially for anxiety disorders such as OCD and social phobia, were not available. Although the GAD-7 measures a range of anxiety symptoms and is considered to have some sensitivity in identifying patients with panic disorder, social phobia and PTSD, it does not address some of a range of specific symptoms found in other disorders (e.g. those of PTSD or OCD). This may explain why there were more classes identified for GAD change than PHQ, as there are more potential anxiety diagnoses influencing GAD scores. The use of anxiety disorder specific measures (ADSMs), for example the Social Phobia Inventory (SPIN) (Connor et al., 2000) could increase the accuracy in the predictions of likely change from initial assessment and would be of value to explore in further analyses. The current dataset did not include ADSMs as these were not routinely collected in IAPT until more recently, these measures could replace the use of the GAD-7 to guide treatment decisions. Newer datasets with better completion of these additional measures could provide more information to inform expected change for specific anxiety disorders.

The analysis presented in this chapter explored the change in symptoms during GSH or HI interventions, and although these were the main intervention for 85% of the original sample (n=15376), the results of this analysis cannot be extrapolated to other IAPT delivered interventions. Some interventions such as pure self-help and signposting would unlikely benefit from modelling trajectories of change due to the limited number of time points received in routine treatment, as only one or two contacts would be expected. However, there may be clinical utility in exploring the trajectories of symptom change during ‘group-based’ treatment in IAPT. Group interventions are a recommended psychological treatment for CMHDs such as GAD, OCD and depression (NICE, 2011), but were the main intervention for less than 5% of the full sample. Group treatments in IAPT can have a range of functions, for example relapse prevention or trauma-focused work, and can vary in

purpose as well as duration. Due to the potential variation in actual intervention received, it could be argued that an analysis of change in symptoms by different types of IAPT delivered group treatments would be more clinically beneficial, but this was not possible in the current analysis due to low number of patients who attended specific types of groups. Instead, future analysis could supplement this dataset with data from additional IAPT services, or could use the national IAPT data available (NHS Digital, 2016) to explore trajectories of change in a significant cohort of patients who received group-based treatments.

A further limitation of the current analysis is that a number of potential characteristics or factors that could impact trajectories of change were not included. For example, therapist characteristics have shown to explain around 9% of the variation in treatment outcomes (e.g. Green et al., 2014) and clinician information could be included in growth analysis to refine the predicted trajectories. However, this would result in information about the clinician being required as part of the DST used by the service, and therefore any potential algorithm generated would need to be locally adjusted for the clinicians working in the service, and updated with staffing changes. In addition, patient characteristics at baseline could also be included in latent growth methods to refine the predicted trajectories of change. As the majority of the patient characteristics available in the IAPT dataset are already included in the probably of profile membership, conducting more focused LCGA by profiles could result in more stratified trajectory information being made available. The next chapter presents a series of LCGA conducted by latent profile.

Summary

The findings from this chapter provide significant information that has potential utility to aid decisions that need to be made by clinicians during the course of treatment. For example, whether the intervention is having a positive effect, or whether it needs to be modified to improve the likely outcome. The LGC analyses present the expected response curve for patients attending treatment to either GSH or HI, which can be tailored to show more relevant information for individuals of specific LPs. This information could be used by clinicians as a reference to compare their current patients against the trajectories of change expected from similar patients (i.e. members of the same LP). If the current patient is showing a change in symptoms which is less than would be expected from the active treatment, then this information could be used to inform a discussion between the patient and clinician as to whether they feel the treatment is working as desired, or whether an alternative treatment should be considered. The trajectory curves identified by the LCGA presented in this chapter could be built into an EPMS such as those used by IAPT services to provide real-time comparison between the current trajectory of a patients and the classes identified. However, it was found that individuals from certain LPs were understandably more

likely to be members of classes with intercept values closer to the mean symptom score at initial assessment. As a result, LCGA performed by LP may be of more value for informing a DST to support treatment monitoring decisions. The next chapter focuses on LCGA conducted by individual LPs, which may provide a more refined and personalised approach to support treatment monitoring decisions in IAPT services.

Chapter 8. Development of a decision support tool for IAPT services.

Abstract.

Identifying differing classes of patients with specific trajectories of change in symptoms during the course of treatment could be fed back to clinicians to inform on treatment progress and aid treatment monitoring decisions. However, identifying the likely trajectories of change for each of the latent profiles could combine the major findings from this thesis to develop a decision support tool that can aid both treatment selection and treatment monitoring decisions in IAPT services, and create a more personalised approach to psychological treatment in the UK. Latent class growth analysis (LCGA) was performed by latent profile, for patients who received either guided self-help (GSH) or high intensity (HI) interventions. The results of the LCGA found a range of one to four identified trajectories of change in depression or anxiety symptoms per profile. The chapter then demonstrates how patient characteristics could be used by an algorithm to provide a probability of profile membership at assessment, a recommendation for appropriate IAPT intensity of treatment and then provide information about the likely trajectories of change for the patient. This algorithm could be hosted within an IAPT service's EPMS, where patient information collected pre-treatment could be automatically entered into the algorithm to identify the latent profile, before providing a potential treatment recommendation based on likely response to IAPT treatments. Once treatment has been selected, the algorithm could then support treatment monitoring decisions by providing information about the likely change in symptoms for patients of the same profile, and a decision support tool combining all these elements could be used personalise treatment decisions in IAPT services to contribute to a more efficient healthcare system.

Introduction.

The results of the LCGA presented in the previous chapter indicate that there are a number of distinct trajectories of change in both PHQ-9 and GAD-7 scores during Guided Self-Help (GSH) and High Intensity (HI) interventions. This information could be used to determine whether a positive or negative outcome was more likely given the current trajectory of change in symptom scores, aiding decisions about the effectiveness of the intervention being delivered. The distribution of these classes between the previously identified latent profiles (LPs, see Chapter 5) suggests that there are predominantly one or two trajectories

that are likely for members of each LP. This information could be useful to clinicians, as once a patient is identified as a member of a particular LP then the LCGA results could be cross-referenced and the likely trajectories of change known as treatment is started. However, the results also suggest that many of the LPs have the same likely trajectory, for example trajectory class 1 for PHQ change during HI treatment (early reduction in symptoms, levelling out with increasing sessions) was the primary class for five or the eight LPs (LP1, LP2, LP3, LP4, LP6). It could argued that this type of treatment response (early reduction, levelling out) is independent of LPs, and instead is just a typical response to interventions across the dataset of patients.

An alternate reason for this finding could be linked to the intercept (starting) value of the identified classes. It was typically found the most frequent trajectory identified for an LP was the trajectory class with an intercept value closest to the mean PHQ-9 and GAD-7 score for that LP. This would seem logical, as the outcome variable in the LCGA is also included as a characteristic used to identify the patient's LP. As a result, it could be argued that conducting LCGA by individual LP may provide more valuable information to inform the likely trajectory of change for sub-groups of patients attending the IAPT services. This would provide a more tailored approach to the predicting of change during treatment, and it would be expected that performing LCGA by LP would result in a more accurate prediction of outcome, increasing potential clinical utility.

It was expected that the number of classes identified in LCGAs conducted for each profile would be lower than the number identified for the full sample of patients presented in Chapter 7 (4 PHQ-9 classes, 5 GAD-7 classes). This hypothesis is based on results from previous growth modelling on CMHD samples, where patient data collected from routine psychological treatment services across a range of disorders identified five classes (Stulz et al., 2007) and analyses of more homogeneous clinical trial samples identified just two classes (Gueorguieva, et al., 2011). As the LP samples to be used in the current analysis will include patients who are expected to share similar patient characteristics (hence why they are members of the same LP), the likely change in symptoms would also likely be less variable. For some profiles, it is possible that only one expected response curve would be identified, and therefore the results of the latent growth curve (LGC) analyses presented in Chapter 7 would be the only available estimate of change in symptoms during treatment.

It would also be of interest to observe whether the differentiation in trajectories is apparent by session three as suggested by researchers in psychotherapy outcomes (Lambert, 2013). The results of the LCGA conducted on the full sample of patients receiving HI treatment (see Chapter 7) suggested that PHQ-9 change was not differentiated between patients in class 2 and class 4 until at least the fourth session of treatment, and therefore conducting LCGA by profile might uncover which patients will need longer monitoring than others before change in the treatment approach may need to be considered.

Once the different likely trajectories of change have been identified, a decision support tool (DST) could be developed to aid decision making in IAPT services. An algorithm could be constructed to provide profile membership information for any new patient attending IAPT services based on patient characteristics assessed, and probability of treatment outcomes for both LI and HI treatments could be provided as part of the tool. This could then be used to inform a discussion between the clinician and patient about appropriate treatment and likely response. Once treatment has commenced the DST could also provide the possible trajectories of change to allow the clinician to compare change for the current patient against that which would be expected. If progress is not following a trajectory indicative of a positive outcome then the DST could prompt the clinician to consider alternative treatments. Providing support from DSTs during clinician workflow through computer-based systems has been linked to a greater impact of tools in supporting clinical decisions (Kawamoto et al., 2005) and therefore an algorithm which is incorporated into the EPMS used by IAPT services could provide real time information that will likely be the most beneficial.

The development of such as tool would be a prototype but by combining both treatment selection and treatment monitoring decision support it would be able to support more clinical decisions than other available systems in mental health treatment. For example, the OQ-45 based QA system (Lambert et al, 2001) has been developed to provide support for treatment monitoring decisions during psychotherapy by flagging patients who are 'Not on track' but does not support decisions about which type of treatment is likely to result in the best patient outcomes. The system also relies of the use of the OQ-45 which requires 45 questions to be collected at each session, which may make this system too burdensome for routine use in IAPT. DSTs for treatment selection decisions in CMHDs are currently in the development phase, with systems yet to be evaluated in routine care. However, methods such as the 'PAI' (DeRubeis et al., 2014) and the 'LRI' (Delgadillo et al., 2016) have potential to support treatment decisions in IAPT, but do not at present include any information about within treatment change.

IAPT services provide a unique opportunity in mental health to evaluate the use of DSTs due to the large number of patients receiving treatment per year, but also the IAPT minimum dataset (MDS) that provides standardised pre-treatment and within-treatment patient data. The aim of this chapter is to explore whether there are statistically distinct trajectory classes of patient response to IAPT treatment within each LP, providing a more personalised estimate of likely change. Once differing trajectories of change have been identified, a DST that combines latent profile information and the trajectories of change for each profile can be developed for potential use in IAPT services to personalise treatment.

Method

Sample

This analysis used the same dataset described in the previous chapter, and included all individuals receiving GSH (n=3334) or HI (n=4394) intervention who met the previously used inclusion criteria. These two datasets were split by latent profile, and the distribution of LPs by intervention is presented in Table 7.2 (Chapter 7). The number LP3 members very was low, with n=10 individuals receiving GSH and n=14 receiving HI interventions. These numbers were considered too low to include in LCGA modelling, and therefore LP3 was excluded from subsequent analyses.

Measures

The LCGA analysed used the PHQ and GAD symptom scale scores at each contact, collected as part of the IAPT services' ROM practices. As described in Chapter 7, the number of time points included in the analyses was different between the two intervention types. Eight time points were included for GSH, and 13 time points for HI interventions, as were used in analyses presented in the previous chapter.

Analysis

LCGA was performed using the same method as described in Chapter 7. The Vuong-Lo-Medell-Rubin Likelihood Ratio test (VLMR-LRT; Lo, et al., 2001), the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and entropy value statistics were compared between models for each analysis by LP. A VLMR-LRT with a p-value >=0.5 was used to identify when increasing the number of classes no longer improved model fit, whereas lower AIC and BIC values for a model indicated better model fit. Higher entropy values indicated better classification accuracy for the model.

There was no prior hypothesis about the number of classes, and therefore the LCGA was performed stating with a two-class model, assessing fit statistics and then increasing the number of classes with the LCGA until the VLMR-LRT was no longer statistically significant or either the AIC or BIC values increased compared to the previous model (Geiser, 2013). As with the LCGA conducted in Chapter 7, classes were required to include at least 5% of the sample for them to be considered numerically stable and meaningful (Spinhoven et al., 2016).

LCGA were conducted by LP, with patients assigned to the LP to which they had the highest probability of membership. Separate LCGA were conducted on PHQ-9 and GAD-7 measures, for patients within the LP who received GSH and for patients who received HI interventions. This resulted in four LCGA performed for each LP.

Results

Due to the large amount of analyses performed in this chapter, model comparison tables are presented in Appendix F, and the LCGA results are summarised by LP below.

LP1

Guided Self-Help (PHQ = 1 class, GAD = 2 classes)

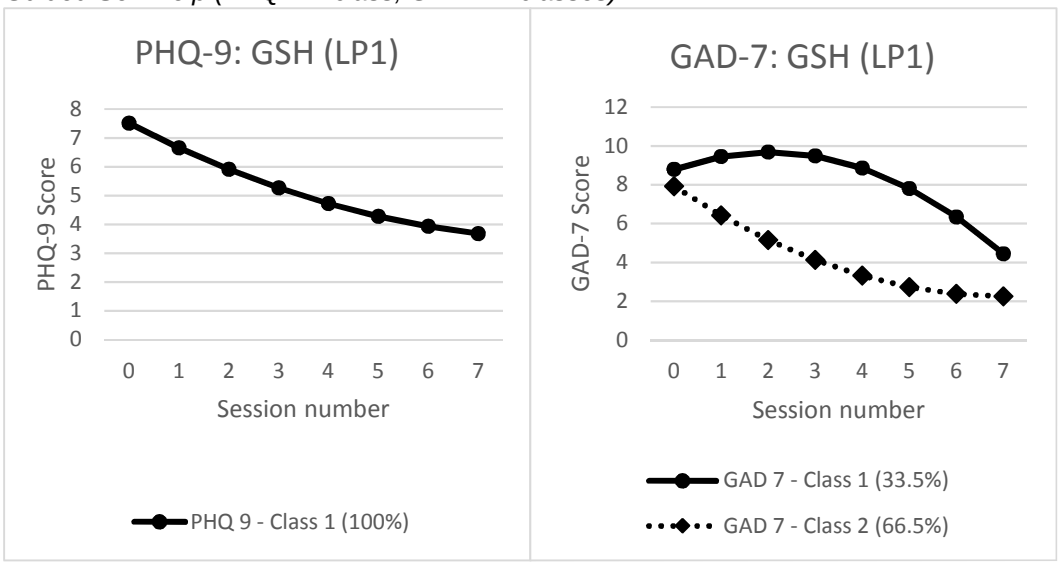

Figure 8.1. Classes identified in LCGA for LP1 (GSH).

The LCGA model comparison statistics for individuals in LP1 who received GSH are presented in Table E1.1 (Appendix F). The K=2 model was the first LCGA performed for PHQ change during treatment. However, the VLMR-LRT reported a non-significant p-value (*p*=0.134) for the K=2 model, which would indicate that a K=1 model would be a better fit for the data. The K=1 model (one class) was the same as the latent growth model presented in the previous chapter for LP1. The AIC and BIC values were also compared between the K=1 and K=2 models, and although values were slightly lower in the K=2 model, the difference was small and therefore the K=1 model was selected as the best fit for the data. This suggests that there were no statistically different curves of PHQ change identified in the sample in addition to the initial latent growth curve, implying that change in depression scores for this profile of patients generally follows the expected response curve. The trajectory curve is presented in the left section of figure 8.1.

The results of the LCGA for GAD change showed that the K=2 model was a better fit than the K=1 as the VLMR-LRT p-value was less than 0.05, but the VLMR-LRT become non-significant at K=3 model. The K=3 model also included a class of patients that represented just 3% of the included sample, and therefore the K=2 model was selected for GAD change in LP1 patients receiving GSH, indicating two statistically distinct classes of patients. The graph presented in the right panel of figure 8.1 shows these two classes; the first (class 1)

initially shows little change before a more rapid decrease in symptoms, whereas class 2 indicate a group of patients with larger decreases in symptoms early in treatment that levels out with increased sessions.

High Intensity (PHQ = 2 classes, GAD = 2 classes)

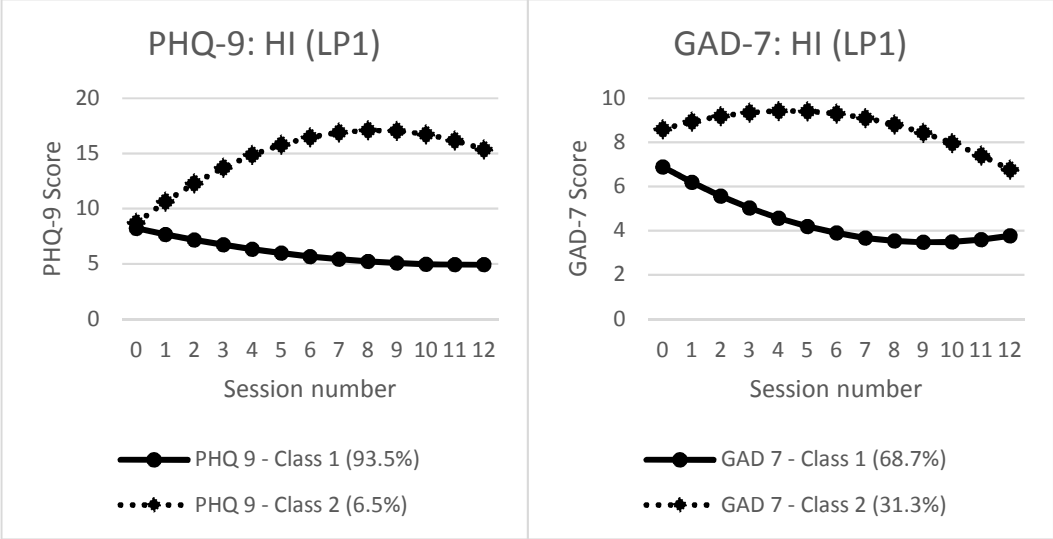

Figure 8.2. Classes identified in LCGA for LP1 (HI).

The results of the LCGA for patients in LP1 who received HI interventions are presented in Table E1.2. Model comparison statistics indicated that two classes were the best fit for both PHQ and GAD measures. For PHQ change, the VLMR-LRT remained statistically significant for the K=2 model, but non-significant for the K=3 model (p=0.082). The K=3 model also showed an increased BIC value and the third class included only 1% of the sample, which meant the K=2 model was selected for PHQ change in this sample. The VLMR-LRT test showed a similar pattern with GAD measures, as K=2 was significant whereas K=3 moved into non-significance (p=0.13).

The two graphs presented in figure 8.2 show the classes for individuals receiving HI interventions in LP1. Although two classes were identified in both the PHQ and GAD scales, the shapes of the graphs were different. The graph in the right panel of figure 8.2 shows that GAD change in HI were very similar to GSH; one class indicated more early change that levels out and a second class with limited change at first before a more rapid decrease in later sessions.

The graph for PHQ change in HI presented in the left panel of figure 8.2 shows two independent trajectories. Class 1 displays a group of patients with large decreases in symptoms that levels out (as seen in the three previous graphs) but class 2 have a very different trajectory, where depression symptoms increase in severity over the first few sessions. This trajectory would indicate a deterioration of symptoms, and although this class

make up just 6.5% of LP1 patients receiving HI interventions, providing this information to clinicians could benefit patients by indicating that the current treatment may be having a harmful affect.

LP2

Guided Self-Help (PHQ = 3 classes, GAD = 4 classes)

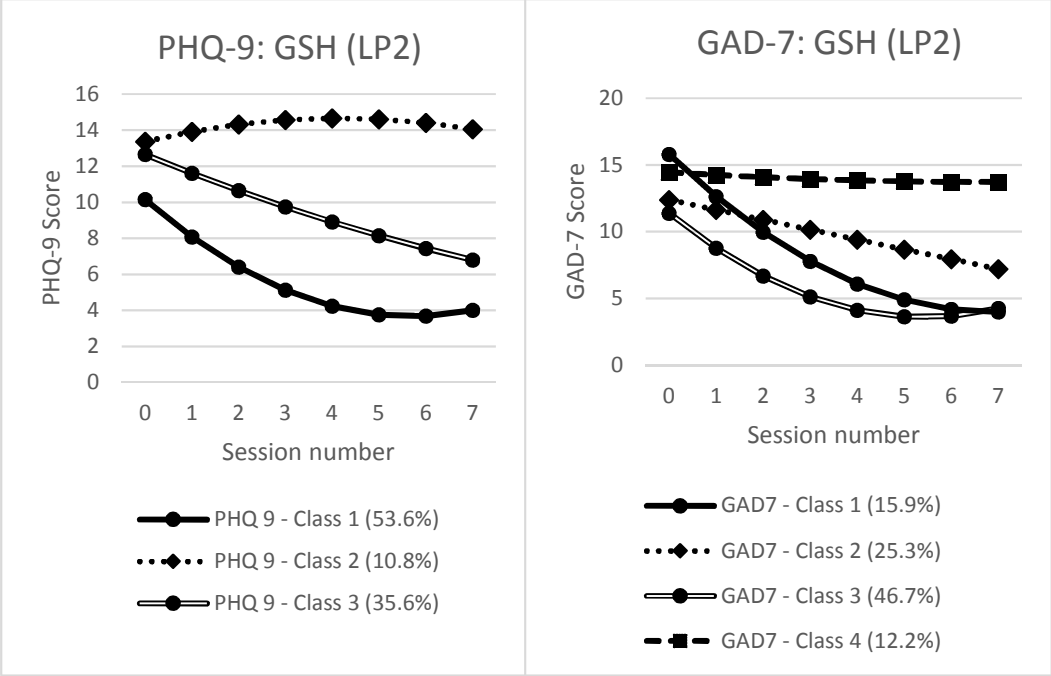

Figure 8.3. Classes identified in LCGA for LP2 (GSH).

The results of the LCGA for LP2 individuals receiving GSH is presented in Table E1.3. The VLMR-LRT remained statistically significant until the K=4 model for PHQ score change, indicating the K=3 model was the best fit for the data, especially when compared alongside the limited BIC change and greatly decreased entropy value for the K=4 model. The K=3 model was therefore selected, and the trajectories presented in the left of figure 8.3. Class 1 show a group of patients who show a rapid initial decrease in symptoms which levels out, therefore is very similar to the expected response curve for this profile, and accounts for over half of LP2 patients (54%). Class 2 accounts for just over 10% of patients and indicates a group of patients with a slight increase in symptoms which flattens out, suggesting little change in depression symptoms for this class. The final class (class 3) show a steady and relatively linear decrease in symptoms from the start of treatment to the seventh contact.

The model comparison statistics for the LCGA of GAD change (Table E1.3) suggest that a K=4 model was the best fit for the data, represented by a non-significant VLMR-LRT value for the K=5 model (p=0.431). The K=4 model was therefore selected, and the trajectories presented in the right of figure 8.3. Both classes 1 and 3 show very similar shaped curves,

with a steep initial decrease in symptoms that levels out. The main difference between the trajectories is the intercept values which show that initial severity is higher for class 1 patients compared to class 3. Together these GAD classes account for over 60% of patients in LP2 attending GSH. Class 2 appear to be a group of patients with a steady and linear decrease in anxiety symptoms during treatment, and seem similar to class 3 identified in the PHQ LCGA for LP2. The final class (class 4) indicates a group of patients with little change in symptoms during treatment, and account for just over 12% of patients.

High Intensity (PHQ = 3 classes, GAD = 4 classes)

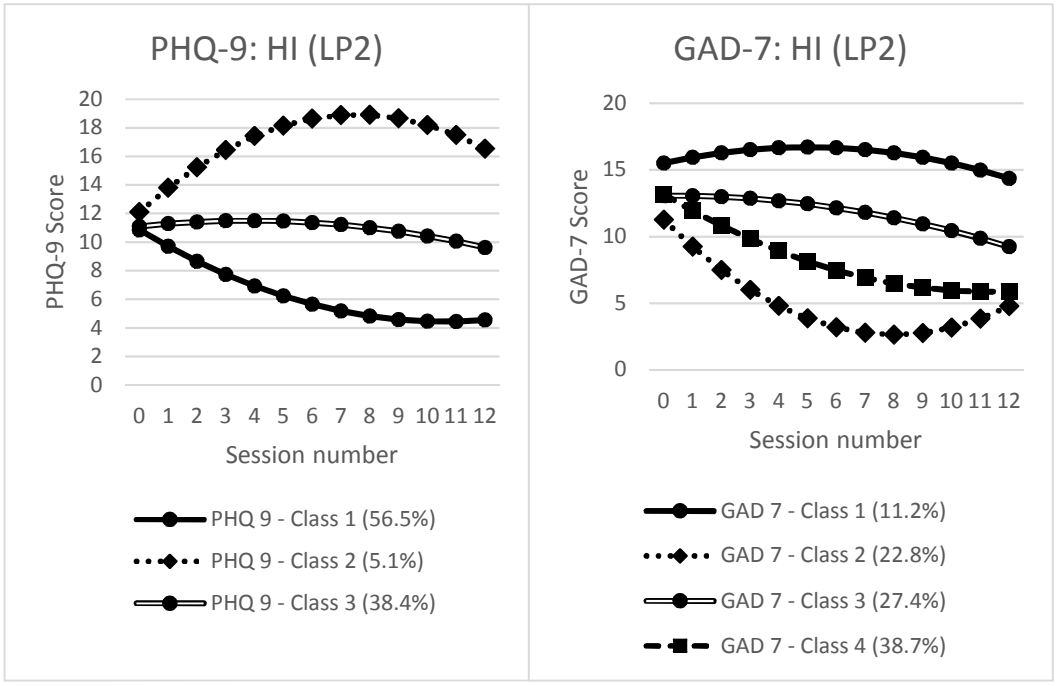

Figure 8.4. Classes identified in LCGA for LP2 (HI).

The results of the LCGA for patients in LP2 who received HI interventions are presented in Table E1.4. The model fit comparison statistics indicate that for PHQ change, the VLMR-LRT remained statistically significant until the K=4 model (p=0.504), suggesting the K=3 model was a better fit for the data. K=4 model also include a class of patients making up less than 5% of the sample, and therefore the K=3 model was selected for PHQ change. The LCGA results for GAD change suggested that the VLMR-LRT was significant until the K=5 model, suggesting K=4 was a better fit for the data. The entropy value was also higher for the K=4 model, and therefore K=4 was selected for GAD change.

The graphical representations of the identified curves are presented in the two graphs in figure 8.4. The three PHQ curves show one group (class 2) who appear to show an initial deterioration of symptoms and is very similar to class 2 in the LP1 PHQ HI LCGA (figure 8.2). Class 1 suggests a group of patients with high initial response to treatment that levels

out, and accounts for over half of the sample (56.5%). The trajectory of class 3 suggests a very limited reduction in depression symptoms during the course of HI treatment, although by session 12 the PHQ score appears to be below caseness (<10).

The four GAD curves presented in 8.4 show one group with a very rapid reduction in symptoms during initial sessions that levels out (class 2), a group with early response to treatment that levels out (class 4), a group with a slow gradual decrease in anxiety symptoms (class 3) and a class of patients with an initial deterioration in symptoms that decreases slightly over time (class 1).

LP4

Guided Self-Help (PHQ = 2 classes, GAD = 2 classes)

The model fit comparison for LP4 patients receiving GSH is presented in Table E1.5, and shows that for both PHQ and GAD change, a K=2 solution was the best fit for the data. In both scenarios the VLMRT-LRT was statistically significant for the K=2 model, but not K=3, and the reduction in BIC values between K=2 and K=3 models was limited.

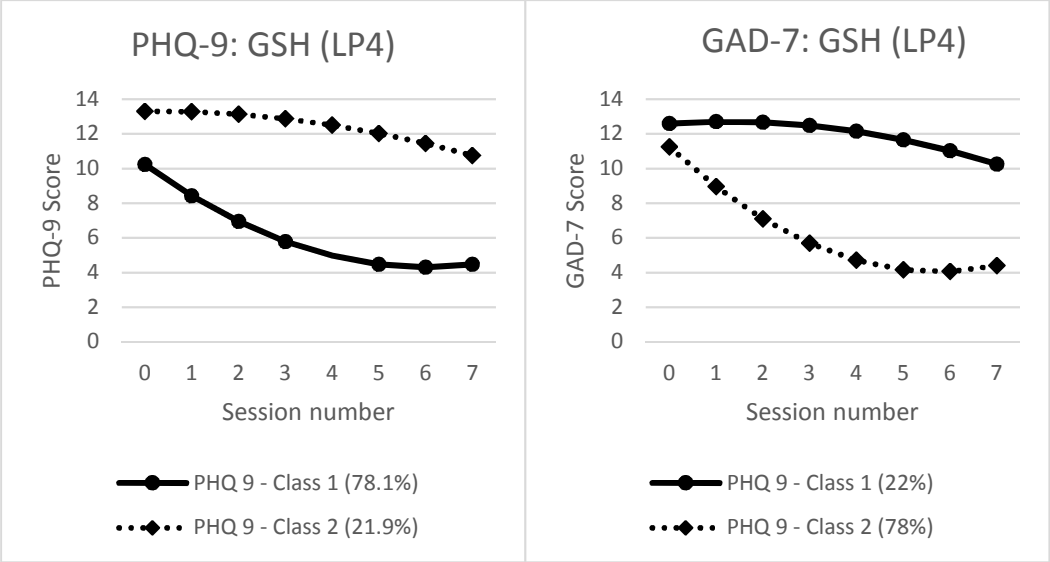

Figure 8.5. Classes identified in LCGA for LP4 (GSH).

The graphs for the two identified classes in both PHQ and GAD change are presented in figure 8.5. The graphs between the two symptom measures look very similar, one class being an early responding group who follow the expect response curve identified in the growth model (see Chapter 7) and a second class of patients who show a slight, gradual reduction in symptoms over time, and account for around 22% of the sample in both symptom measures.

High Intensity (PHQ = 1 class, GAD = 3 classes)

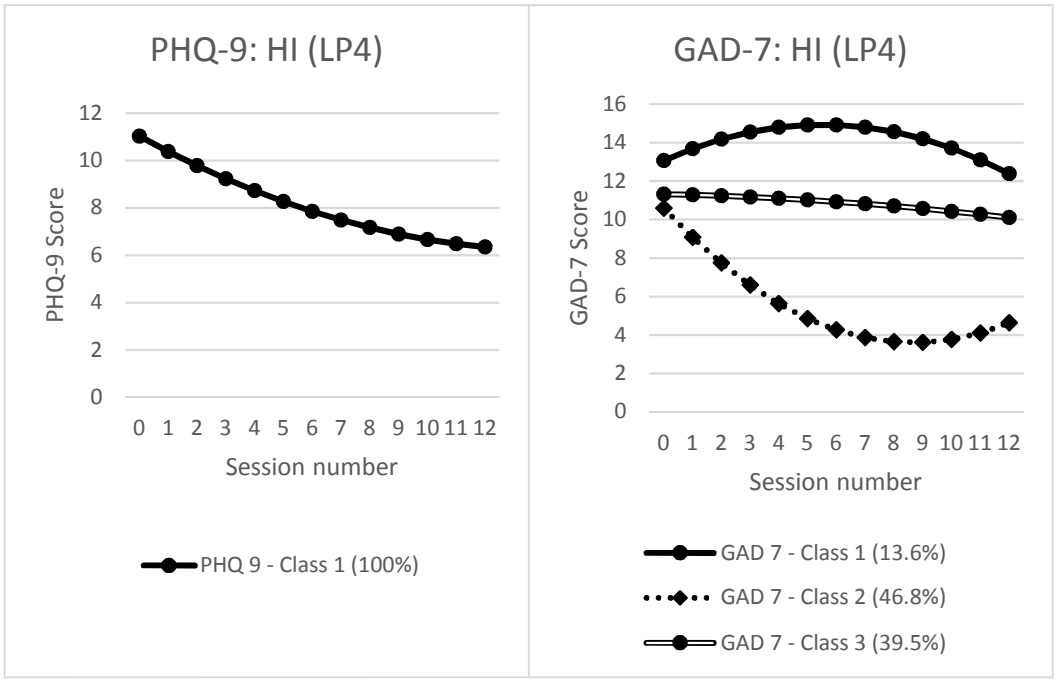

Figure 8.6. Classes identified in LCGA for LP4 (HI).

The results of the LCGA for LP4 patients who received HI interventions are presented in table E1.6. The LCGA for PHQ change found a non-significant VLMR-LRT for the K=2 model, which indicated that the K=1 (the expected response curve) was a better fit for the data. The results of the LCGA for GAD change found a statistically significant VLMR-LRT values up to K=4 (p=0.812). The K=3 model was therefore selected for GAD change.

The left panel of figure 8.6 shows expected response curve for PHQ in LP4 patients receiving HI, and the three LCGA classes for GAD change in the right-hand panel. GAD classes 1 and 3 both appear to indicate patients who do not respond to treatment, and class 1 suggest patients who are at risk of deterioration during treatment. The trajectory of class 2 is very similar to the expected response curve of this LP, showing larger change in the first sessions with levels out over time.

LP5

Guided Self-Help (PHQ = 1 class, GAD = 2 classes)

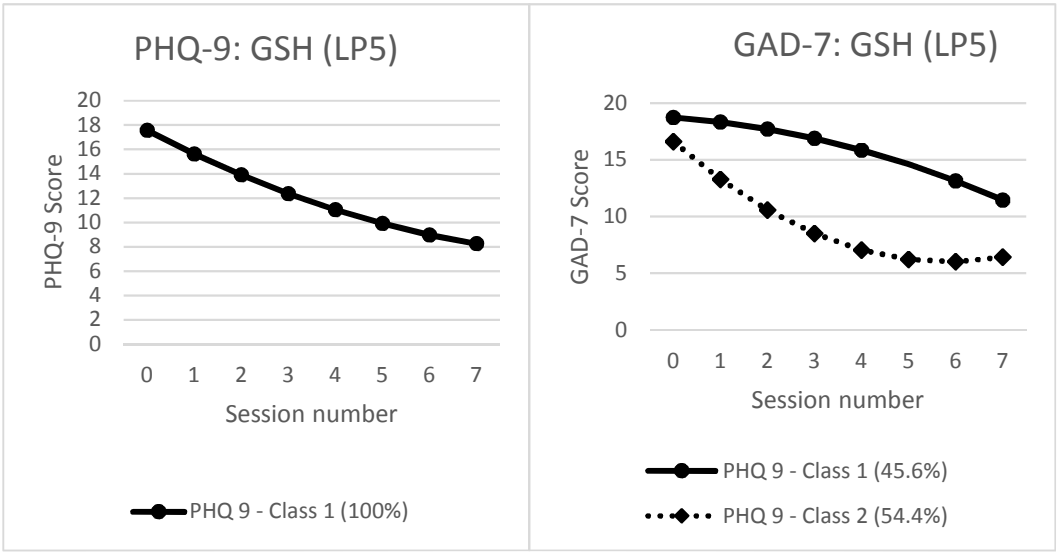

Figure 8.7. Classes identified in LCGA for LP5 (GSH).

The LCGA results for LP5 patients who received GSH are presented in Table E1.7. The LCGA for PHQ found a non-significant VLMR-LRT for the K=2 model (*p*=0.059), suggesting that the K=1 (expected response curve) was a better fit for the model. The LCGA for GAD change identified that the K=2 model was a better fit for the data than the K=3 model, due to a non-significant VLMR-LRT, lower entropy value and limited decrease in the BIC value.

The identified curves are presented in figure 8.7, with the expected response curve for PHQ in the left panel, and the two classes for GAD change presented on the right. The two identified GAD curves indicate one class of patients with a large decrease in symptoms early in treatment that levels off, and a second class with a more gradual and linear reduction in symptoms during treatment.

High Intensity (PHQ = 2 classes, GAD = 4 classes)

The model comparison table for LP5 individuals who received HI interventions is presented in Table E1.8. The VLMR-LRT was statistically significant for the PHQ K=2 model, but not for the K=3 model, suggesting that the K=2 was a better fit for the data. As the BIC and AIC values showed little major decrease between models, the K=2 model was selected for depression change during HI interventions in LP5 patients. The LCGA of GAD change found statistically significant VLMR-LRTs up to the K=4 model, with a non-significant value for K=5 (*p*=0.247), as well as little decrease in BIC values between K=4 and K=5. As a result, the K=4 model was selected for anxiety symptom change,

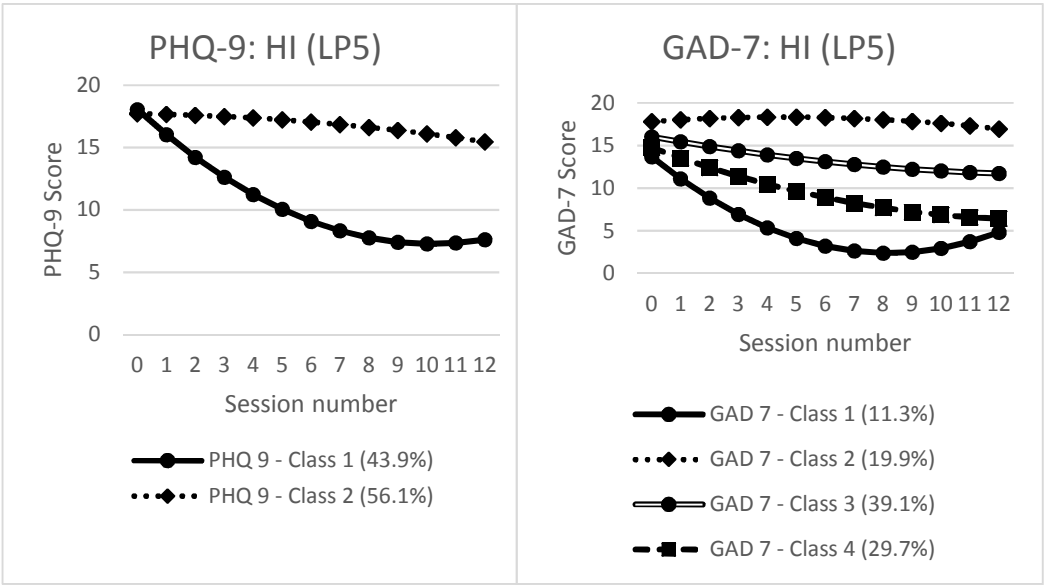

Figure 8.8. Classes identified in LCGA for LP5 (HI).

The form of change for the identified classes are presented in figure 8.8. The left-hand graph displays the two classes of PHQ change, and show one class with a large decrease in depression symptoms during initial treatment sessions that levels out, and a second class who show limited change in symptom scores during treatment. The right-hand graph displays the four GAD LCGA classes, and indicate a very rapid responding group (class 1), and more gradual responding group (class 4), a limited response group (class 3) and a class suggesting limited change in symptoms during treatment (class 2).

LP6

Guided Self-Help (PHQ = 2 classes, GAD = 3 classes)

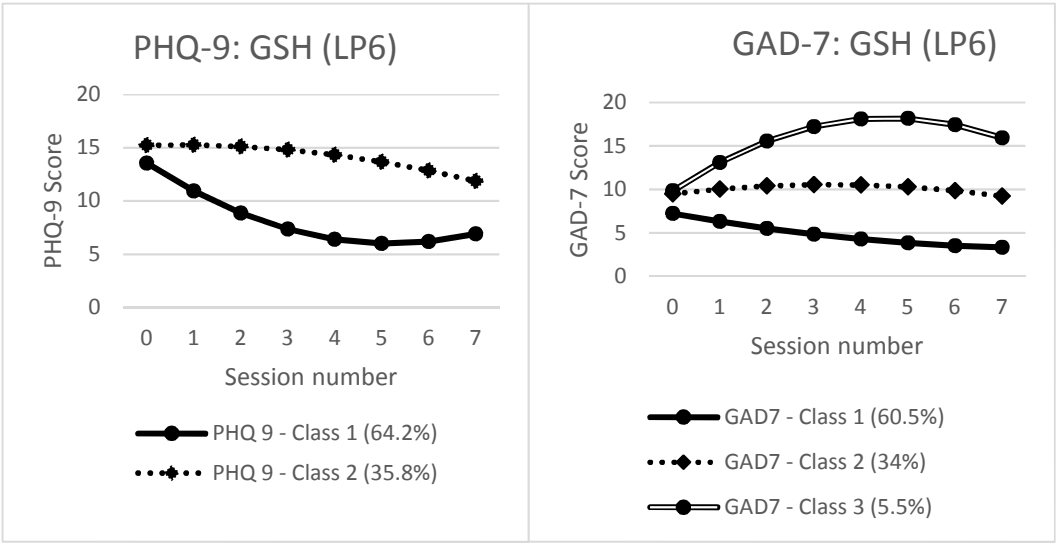

Figure 8.9. Classes identified in LCGA for LP6 (GSH).

The model comparison tables for LP6 patients who received GSH are displayed in Table E1.9. The LCGA for PHQ change indicated that a K=2 model solution was the best fit for the data, whereas the K=3 model was the best fit for GAD change. In both cases the VLMR-LRT became non-significant when the number of classes was increased.

The trajectories of change are presented in figure 8.9. The left-hand panel shows the two PHQ classes identified, class 1 being a group of patients who show a more rapid response to treatment and class 2 suggesting patients who have a more gradual and slight reduction in symptoms. The two classes appear very similar to those for LP4 patients receiving GSH interventions. The three classes of GAD change are displayed in the right panel of figure 8.9. In comparison to the majority of other LCGA analyses, there is no early responding class and instead there is a gradual response class (class 1), and limited response class (class 2) and a deteriorating class of patients (class 3). The expected response curve of GAD change for this LP (presented in chapter 7) suggested a different form of change from all other LPs, therefore it would appear that early rapid change is unlikely for this profile of patients.

High Intensity (PHQ = 2 classes, GAD = 4 classes)

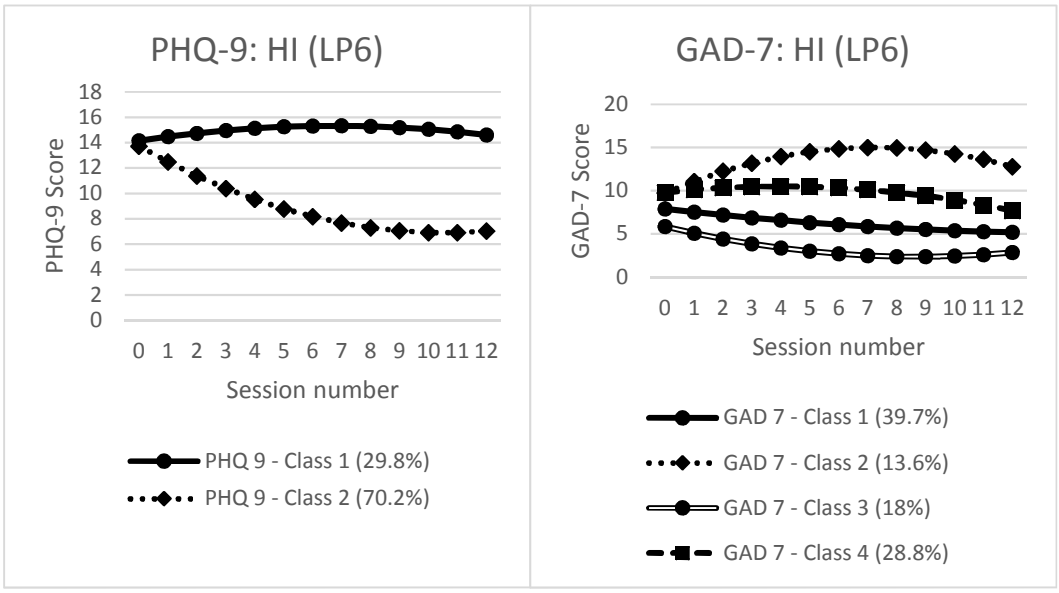

Figure 8.10. Classes identified in LCGA for LP6 (HI).

The results of the LCGA for HI change in PHQ and GAD scores for LP6 patients are shown in Table E1.10. For PHQ change, the K=2 model indicated better fit than K=3, whereas the K=4 model was selected for GAD change based on fit statistics.

The identified classes are presented in the figure 8.10. The two PHQ classes appear to suggest one limited change class (class 1) and a second class who rapidly respond to initial treatment before symptoms level out. The GAD classes show a variety of different

trajectories. Class 3 and class 1 both show responders to treatment whose symptoms level out, and the difference between classes appears to be a slightly higher intercept value for class 1, as well as steeper early decline in symptoms apparent for class 3. Class 2 show clinical deterioration in symptoms in the initial stages of treatment and class 4 show little change until around session 8 where symptoms start to decrease.

LP7

Guided Self-Help (PHQ = 2 classes, GAD = 2 classes)

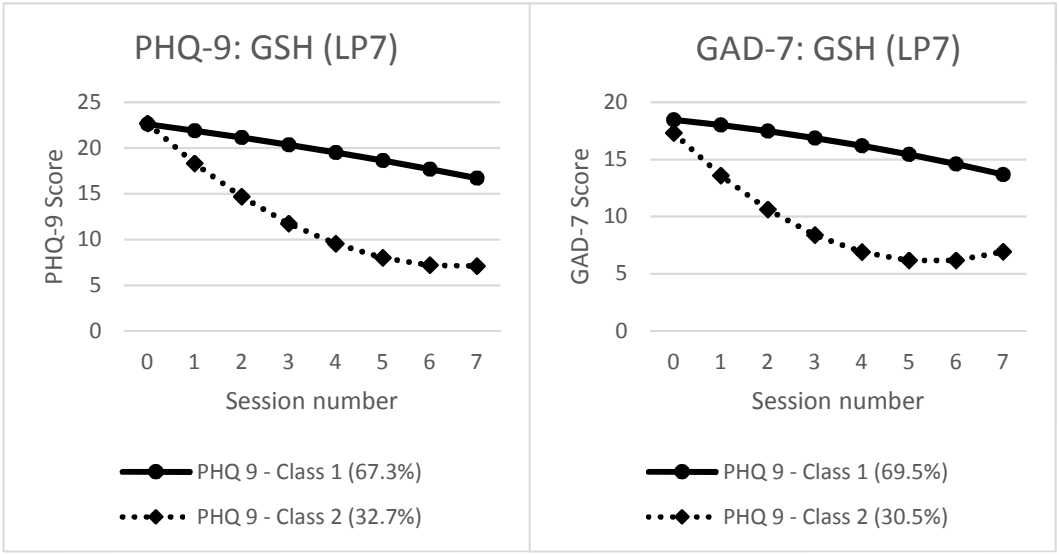

Figure 8.11. Classes identified in LCGA for LP7 (GSH).

The LCGA results for individuals in LP7 who received GSH are presented in Table E1.11 and show that a K=2 model solution was the best for both PHQ and GAD score change during treatment. The graphical representation of these classes, presented in figure 8.11, shows very similar forms of change for each LCGA. Class 1 in both models indicates a group of patients with a gradual, moderate reduction in symptoms over time, whereas class 2 indicate an early responding group whose symptoms level off during the course of treatment.

High Intensity (PHQ = 2 classes, GAD = 3 classes)

The model comparison table for individuals in LP7 who received HI interventions is displayed in Table E1.12. The VLMR-LRT for PHQ change was statistically significant for the K=2 model, but not K=3 (p=0.104). As a result, the K=2 was selected for PHQ change. The VLMR-LRT remained statistically significant for GAD change up to the K=5 model (p=0.257), indicating that K=4 was a better fit for the data. However, the K=4 solution included a class

of patients representing just 3% of the sample, and therefore the K=3 model was selected as a more conservative model.

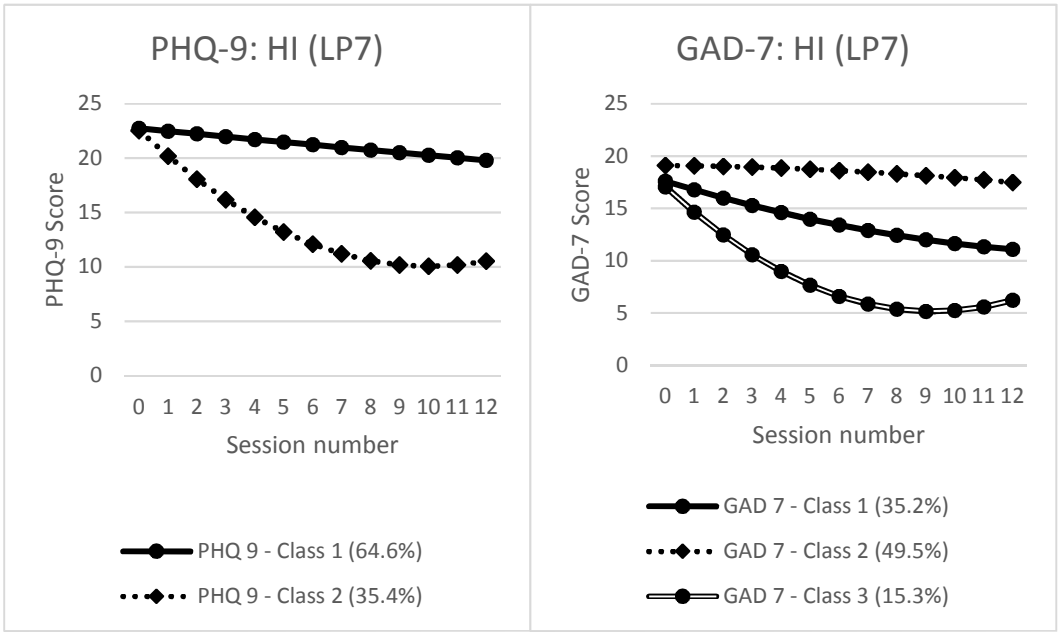

Figure 8.12. Classes identified in LCGA for LP7 (HI).

The trajectories of the identified classes are displayed in figure 8.12. The left-hand panel shows the two PHQ classes, class 1 indicating limited change in depression symptoms over time, and class 2 a more rapid responding group whose symptoms level off during treatment, which is similar to the classes identified in GSH interventions. The GAD classes presented in the right-hand panel show three trajectories of change in anxiety symptoms during HI interventions. The trajectory of class 1 indicates a gradual decrease in anxiety symptoms during treatment, compared to class 2 where limited change is achieved over 12 sessions. Class 3 show a rapid response to initial treatment, indicated by a large decrease in symptoms which then levels off over time.

LP8

Guided Self-Help (PHQ = 3 classes, GAD = 4 classes)

The results of the LCGA for LP8 patients who received GSH are presented in Table E1.13. The analysis of PHQ change found a statistically significant p-value for VLMR-LRTs for K=2 and K=3, but non-significant for K=4 (p=0.138). As there was little reduction in BIC value between K=3 and K=4, as well as a decrease in entropy value, the K=3 model was selected for PHQ change. The results of the GAD analysis indicated statistically significant VLMR-LRT values up to K=5, at which point the BIC value increased, and therefore the K=4 model was selected.

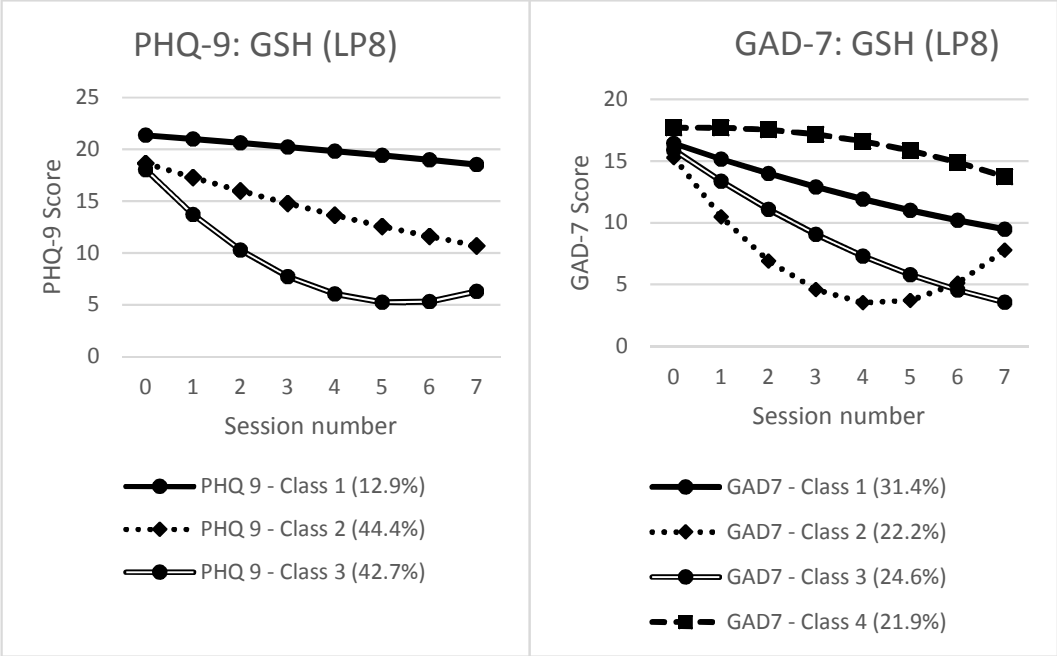

Figure 8.13. Classes identified in LCGA for LP8 (GSH).

The trajectories of the identified classes are presented in figure 8.13. The PHQ classes are shown in the left panel, and the three classes could be described as a limited change group (class 1), a gradual responding group (class 2) and a rapid responding class whose symptoms level out (class 3). The four GAD classes are presented in the right panel, with all classes showing a decrease in anxiety symptoms. Class 1 appear to be a group of patients with a steady gradual and linear decrease in symptoms. The form of change for class 4 appears similar to class 1, however the lack of decrease in symptoms over the first few sessions differentiates class 1 and class 4. The trajectory of class 3 patients is a clear and continuous decrease in anxiety symptoms during treatment, indicating a positive response to treatment. Class 2 however standard out, as the trajectory appears to show a very rapid response to treatment (decrease in symptoms) but after the symptom drop they then appear to increase in later sessions of treatment.

Although some previous curves presented show a small increase in symptoms after levelling out, LP8 GSH GAD class 2 have the most pronounced increase. It is possible that these individuals show a very good response to treatment that is difficult to maintain, potentially due to more co-existing problems such as personality disorders or physical health problems, or due to external factors such as family conflict.

High Intensity (PHQ = 4 classes, GAD = 4 classes)

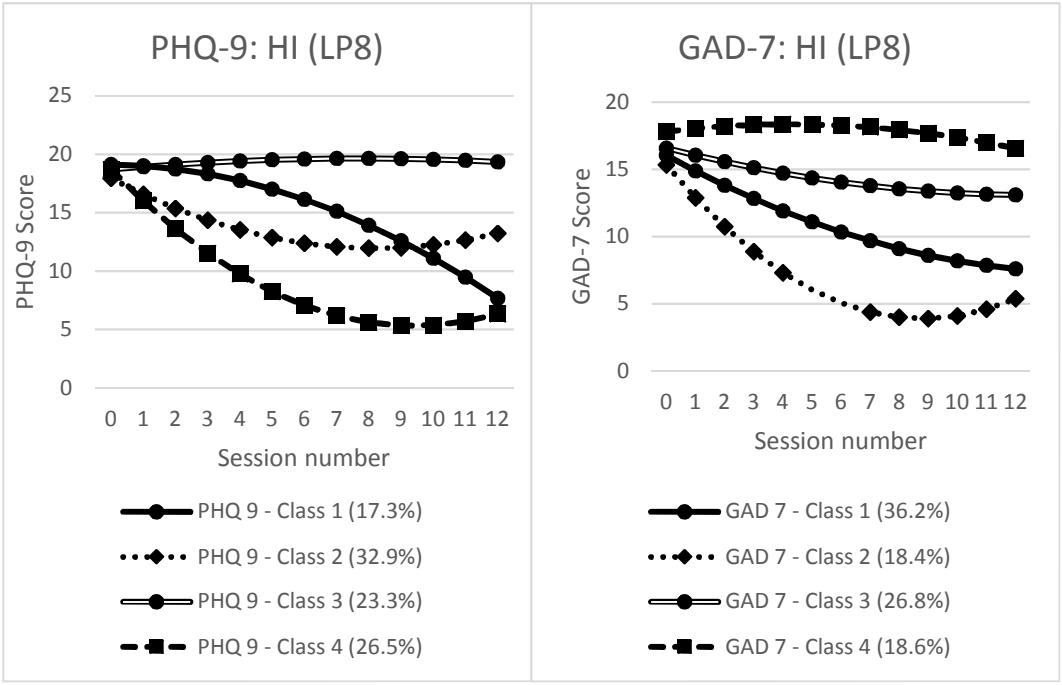

Figure 8.14. Classes identified in LCGA for LP8 (HI).

The model comparison table the LCGA of LP8 patients who received HI interventions is displayed in table E1.14. The VLMR-LRT for PHQ change remained statistically significant from the K=2 model to K=5, with a non-significant p-value at K=6 (*p*=0.114). Although the K=5 model had a slightly lower BIC value & higher entropy value than K=4, it also had a class of patients who made up just 2% of the sample. As a result, the K=4 model was selected for PHQ change. The results of the GAD LCGA showed a statistically significant VLMR-LRT up to the K=5 model (*p*=0.104), with a limited decrease in BIC and entropy values for K=5, and therefore the K=4 model was selected.

The identified classes are presented in figure 8.14, with four curves displayed for PHQ and GAD change each. The left panel shows the four PHQ classes, with class 1 standing out as a class who show limited change in the first few sessions before a large reduction in depression symptoms as treatment sessions increase. Class 3 by comparison show little change overall, but appear to following the trajectory of class 1 until session 3 or 4. Class 2 show reasonable response over the first few weeks which levels out, whereas class 4 are a group of patients who appear to respond very quickly to treatment and this response levels out over later treatment sessions.

The GAD trajectories displayed in the right-hand panel of figure 8.14 show four distinct curves of change in anxiety symptoms in LP8 patients when HI interventions are received. Class 1 suggests a group of responding patients with a gradual and linear change in symptoms during the course of treatment, whereas class 2 show a group of patients with a larger decrease in symptoms during initial sessions that levels out. Both classes together

comprise 55% of the sample. Class 3 indicate a limited reduction in anxiety symptoms during treatment, whereas class 4 suggest a slightly increase in anxiety that levels out, and therefore no response to HI treatment.

Summary

The findings from the series of LCGA performed in this chapter indicate that the number of different trajectories identified, as well as the shape of the trajectory curves, varies between LPs. These could provide a more refined prediction of likely change in treatment based on initial patient characteristics assessed at referral to services. Patients in LP2 for example had three identified trajectories of change in depression symptoms when HI interventions were received, and the results suggest that by session three or four, the clinician and patient may already have an indication of the likely trajectory of symptom scores. This finding supports previous suggestions by researchers (e.g. Lambert, 2013; Szegedi et al., 2009) that it is possible to predict eventual treatment outcome by around the third session of treatment. If a member of LP2 displays a consistent increase in symptoms over the first few sessions then they are likely members of trajectory class 2, which is indicative of a poor outcome from treatment (suggesting deterioration over the first few sessions). Providing this information to clinicians could provide an opportunity for a change in treatment to be considered, and using feedback in this way have been shown to improve patient outcomes (Lambert et al., 2001; Shimowaka et al., 2010). If by the third session there has been very little change in symptoms from baseline assessment then class 3 would appear the most likely trajectory of change, suggesting limited change in symptoms during treatment. Finally, if symptoms decrease consistently in the first few sessions then the trajectory of class 1 would seem to be followed, suggesting that a positive response to psychological treatment is predicted.

For some LCGA comparisons, model fit statistics indicated that the K=2 solution was not a good fit for the data, and instead the K=1 or expected response curve (presented in Chapter 7) was selected. Some examples of this are PHQ change during GSH interventions for LP1 & LP5 patients and PHQ change in LP4 patients when HI was received. The LCGA performed for these sample identified no additional trajectories to explain the patterns of change in these datasets, other than the expected response curve estimated in Chapter 7. These findings may suggest that the majority of patients follow the expected course of symptom reduction, and that the patterns of change in non-responding patients is highly variable, hence LCGA could not identify homogeneous sub-groups with similar responses to treatment. Although these findings suggest that only one trajectory can be identified for these patients, and therefore only one curve can be included in a potential DST for patients of these LPs, this expected response curve can still be of value to the clinician and the patient. In this scenario, a potential DST could flag patients who were not following the

expected response curve, and signal to both the clinician and patient that a potential change in approach may increase the likelihood of positive outcome, similar to the 'not on track' method used as part of other systems (e.g. Lambert et al., 2001).

Although the analyses presented in this chapter typically show trajectories of change differentiating at around the 3rd (at times 1st or 2nd) treatment contact, supporting previous suggestions (e.g. Lambert, 2013; Szegedi et al., 2009), there were some situations where this was not the case and more sessions were needed before the trajectory could be reliably identified. For example, LP8 patients receiving HI interventions two trajectories (class 1 and class 3) do not appear to differentiate until at least sessions 4 or 5, and therefore a clinician following a '3 session rule' may decide to change treatment too early. Providing clinicians with the full information on likely trajectories for patients of each LP could be used to better inform treatment monitoring decisions, and provide treatment for effectively.

The next section of this chapter describes how the results of the LCGA by LP analyses could inform a DST for use in IAPT services to support clinical decision making.

Demonstration of a decision support tool for clinical decisions in IAPT services.

Combining the findings from the latent profile analysis presented in Chapter 5, the variation in outcomes following intensities of psychological interventions in Chapter 6 and the trajectories of change during treatment described in this chapter could be used to build a prototype decision support tool for IAPT services. In the first stage of such a system, an algorithm could be used to identify LP membership for any new patient referred to the services by entering information on the nine included patient characteristics. The DST could then provide information about the relevant LP to the clinician and this information would be used to support a clinical decision about appropriate treatment. This clinical decision will likely be commencing LI treatment, allocating the patient to HI or referring the patient to alternative services, and the DST could only be an aid to clinical judgement rather than replacement for it.

Should the clinician select an LI or HI intervention for the patients, then the next stage of the algorithm could provide the likely trajectories for the relevant LP when receiving the intervention chosen for that patient, which could be used to inform the clinician about the patient's progress in treatment. If the current trajectory of change for the current patient appears more similar to a trajectory indicating a low likelihood of a decrease in symptoms, or an increase (deterioration), then this information could be used to inform a decision to either change treatment modality or to step the patient up to a more intensive treatment. This

section of the chapter presents the stages of a prototype DST that has been developed using the findings presented in this thesis.

Algorithm for identifying latent profiles.

The first stage of the prototype DST was to create an algorithm that identifies the relevant latent profile (LP) for any new patient being assessed. Once the LP has been identified, treatment selection and treatment monitoring decisions can be tailored to members of that LP.

The results of the LPA presented in Chapter 5 were used to generate the first stage of the prototype algorithm, which calculates the probability of profile membership for any new patient referred to an IAPT service after initial assessment, using the nine characteristics included in the IAPT MDS. This algorithm calculates the posterior probabilities of profile membership for any patient once all nine patient characteristics are provided. These posterior probabilities are informed by the results of the LPA, using the equations presented in Chapter 5. The patient information could be entered directly into the DST by the clinician, or the algorithm could be hosted by the service's EPMS. Hosting the algorithm in the system would mean that pre-treatment assessment information could be automatically submitted to the algorithm, once the assessment has been completed and entered into the EPMS.

Once the patient's information from the nine patient characteristics are submitted to the algorithm, profile membership information including the probability of membership to each profile is calculated. The profile to which the patient has the highest probability of profile membership is then assigned to the patient. The algorithm was used in Chapter 6 to provide profile membership for the validation sample, and the front end (input) for this algorithm is presented in figure 8.15 below.

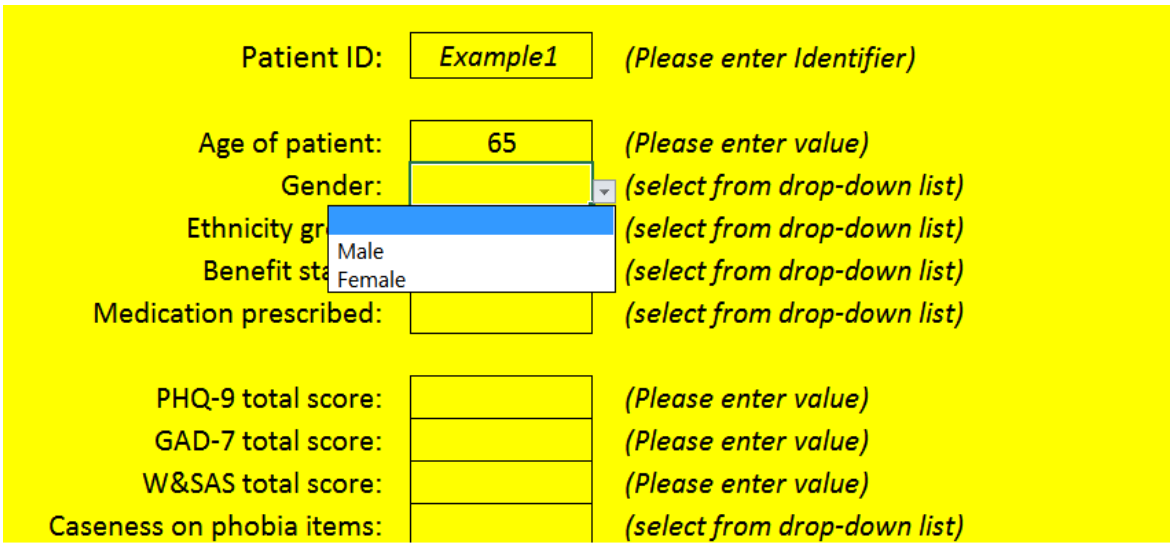

Figure 8.15. Example display: Algorithm input

The prototype algorithm presented here was developed in Microsoft Excel and allows the use of dropdown lists for the five categorical patient characteristics (gender, ethnicity, welfare status, medication and phobia caseness), whereas the four continuous variables (Age, PHQ-9, GAD-7, W&SAS total scores pre-treatment) can be entered as whole numbers.

Aiding treatment selection decisions.

Once the patient characteristics have been entered and the algorithm has identified the patient's latent profile, information about this profile that may be relevant to treatment selection decisions can be made available.

The prototype algorithm could display the likelihood of treatment outcomes, for example providing the probabilities of recovery, reliable change, deterioration and dropout for members of the profile when either LI or HI treatments were received (results presented in Chapter 6). These probabilities could inform an estimate of the expected probabilities, as well as potential utilities, of the different treatment approaches, and it may be that the likelihood of a positive outcome is low and therefore referral to alternative treatment services may more appropriate.

In addition, the DST could also provide a treatment recommendation for the patient based on the likelihood of treatment outcomes. For example, as members of LP7 displayed a low probability of recovery following both LI (11%) and HI (17%) treatment, it might be argued that IAPT delivered treatment was not the most appropriate and instead a referral to another service might increase the likelihood of a positive outcome for the patient. The DST could provide such a recommendation, for example if LP7 was identified:

"Unlikely to benefit from IAPT service, referral to specialist service is recommended."

For members of LP1, who have a high probability of recovery from both LI (75%) and HI (79%), the recommendation might be that LI be considered the initial step of treatment. Information about the likelihood of a positive outcome could also be included in this recommendation to inform the clinician about the likely prognosis, for example:

"Initiate at Step 2. High probability of recovery following treatment, low risk of treatment dropout."

This treatment recommendation would be considered as guidance and used only to supplement clinical judgement, not replace it. There may be important reasons why LI treatment is not appropriate due to characteristics not considered by the algorithm, such as a diagnosis of PTSD, or that the patient has already received LI treatment previously and not benefitted.

These treatment recommendations could however be considered in the treatment selection decision as an aid to clinical judgement, and the IAPT clinician could for example discuss the recommendation during supervision if there was uncertainty. Providing a clear recommendation as well as the supporting evidence for this guidance in the form of probabilities of treatment outcomes could also provide a transparent method of further including the patient in decisions about their care.

Figure 8.16 displays the completed algorithm input as well as the profile membership for an example patient and the treatment recommendation for this profile.

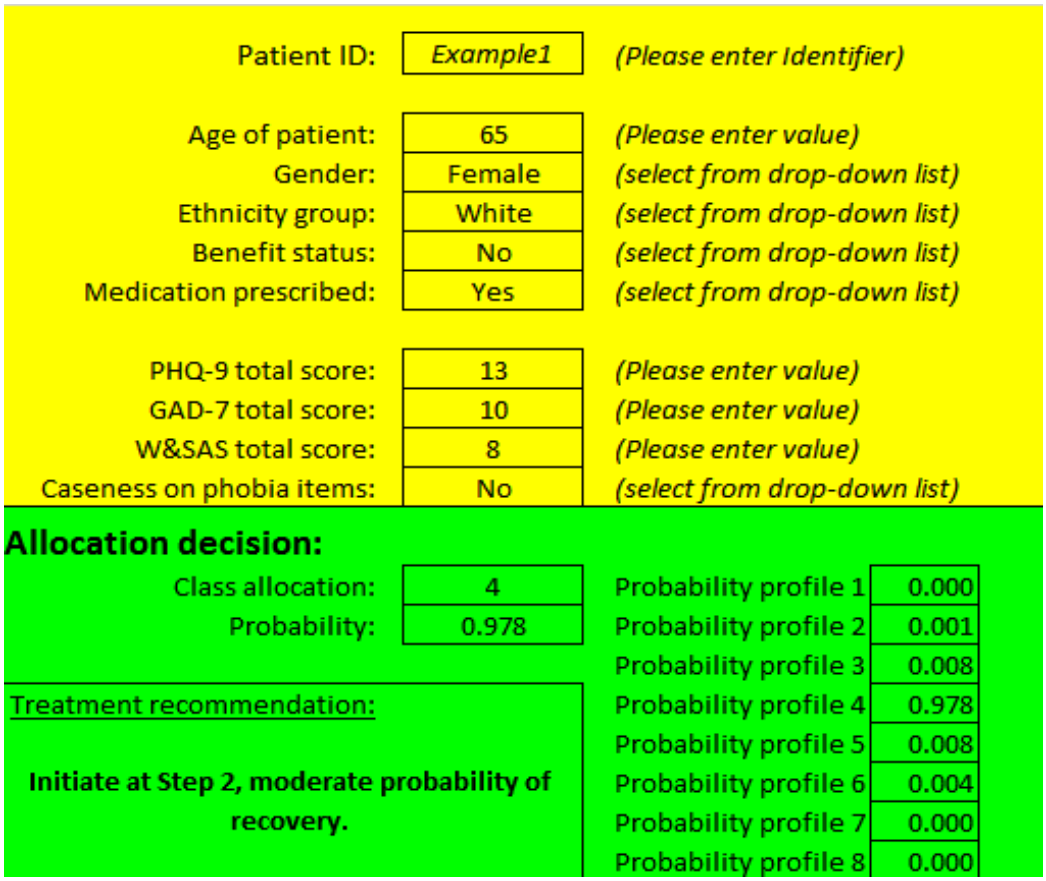

Figure 8.16. Example display: Algorithm treatment recommendation output.

In the example presented, the constellation of patient characteristics has resulted in an allocation to LP4, to which the example patient has a 97.88% probability of membership. The right-hand side of this output shows that the probability of membership to the other seven profiles is low. As members of LP4 have a good likelihood of recovery following LI interventions, and that outcomes are generally better for this treatment compared to HI for this group of patients, the recommendation is to start with step 2 (LI) treatment for members of this profile initially. This is shown in the bottom left corner of figure 8.16.

Aiding treatment monitoring decisions.

The next stage of the prototype algorithm, findings from the LCGA that are presented in this chapter could be incorporated to provide information relevant to treatment monitoring decisions during care. Once the algorithm has identified the patient's profile, the treatment to which the patient has been allocated to (LI or HI) can be selected by the clinician within the DST and the algorithm could display the expected trajectories for the profile, for both PHQ-9 and GAD-7 change.

Figure 8.17 shows the PHQ-9 and GAD-7 trajectories identified for LP4 profiles who received LI interventions. For both symptom scores, two distinct trajectory classes were identified, and the graph also displays the symptom scores for the example patient over the first five sessions. The patient's progress is superimposed and shows how their progress compares to the identified trajectories for the relevant profile. Observing the two graphs below shows that the trajectory of the example patient is much closer to the responding class of patients for both measures, and therefore the clinician could conclude that the current progress is treatment is good and that a change in treatment is not required. If the current patient's trajectory was closer to the non-responding class then the clinician could use this information to consider a potential change in treatment, for example by stepping up, or that referral to alternative treatment services is more appropriate.

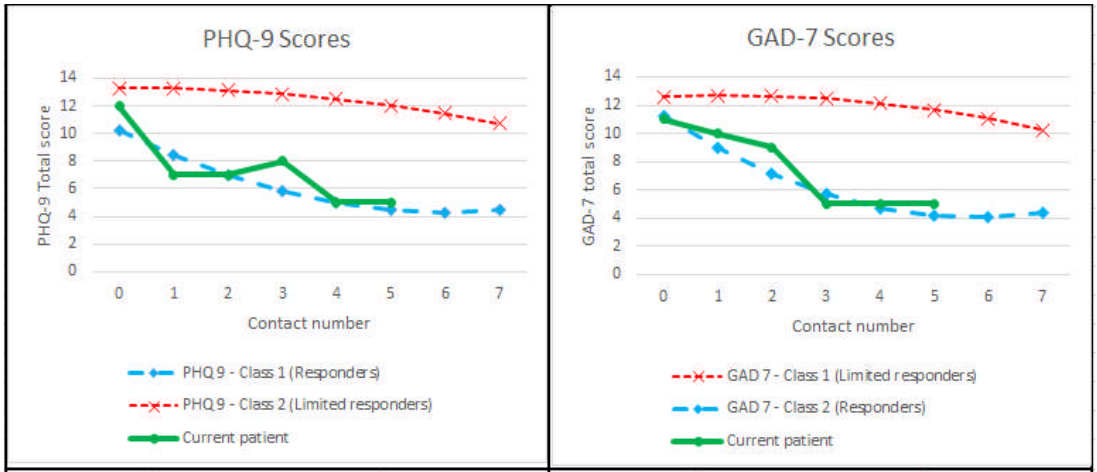

Figure 8.17. Example display: Trajectories of change.

This part of the DST could be built into the EPMS used by the IAPT services to provide feedback to clinicians as to whether treatment is indicative of positive outcome or not, informing decisions needed during treatment monitoring. Whereas other outcome feedback systems (e.g. the outcome assurance system; Lambert, 2013) provide just one expected response curve and suggest that any patient not following this within the first few sessions is not on track, it would be suggested that the prototype DST presented here would instead display all possible trajectories. The value of showing all curves is that it provides

information to the clinician about scenarios where a reduction in symptoms may take longer. For example, the change in PHQ-9 scores for LP8 receiving (figure 8.14) only differs at around sessions 5 for class 1 and class 3, and therefore providing the full range of possible change curves may reduce clinician concern by showing that a positive outcome is still achievable for that profile. The output from the DST could also provide the proportion of the LP that were members of each class, as this would inform the clinician about the relative frequency of the trajectory. As some classes may occur in only 5-10% of the LP, this pre-test probability could be used to aid the clinician estimate the likelihood that the patient will follow specific trajectories.

Discussion

The first aim of this chapter was to identify the distinct trajectories of change in symptoms during treatment for individual latent profiles. Conducting LCGA by these sub-samples of patients who share common characteristics should allow for a more accurate estimate of the rate and nature of change, and therefore provide more valuable information for clinicians to use in making treatment monitoring decisions. The next goal of the chapter was to develop a DST that could be used in IAPT services to support treatment selection and monitoring decisions, using the findings presented in this thesis. A prototype DST algorithm has been developed combining latent profile information presented in Chapters 5 and 6, with the results of the LCGA conducted in this chapter to provide a treatment recommendation as well as information about the likely trajectories of symptom change during treatment. There are currently no other DSTs supporting both these decisions in mental health services, and considering the role IAPT services play in the delivery of psychological interventions in England, there is great potential for improved patient outcomes and service efficiency.

It is of interest that there was a range of different trajectories identified across the LPs, which may not have initially been expected. Although every profile had at least one trajectory class with a similar form of change to the expected response curves presented in Chapter 7, there was significant variation as to the form of the second, third or fourth classes identified. For a number of profiles, at least one trajectory could be described as a non-responding, or limited response to treatment trajectory (e.g. LP5 HI GAD class 2). This split in trajectories appears to replicate the results of Gueorguieva and colleagues (2011), where a responding and non-responding trajectory of patients was observed in a clinical trial dataset of pharmacological treatment for depression.

However, a few profiles included an additional trajectory which more clearly indicated an increase in symptoms, or clinical deterioration. As LP6 showed the highest probability of deterioration in the analysis of outcome presented in chapter 6, it may not be surprising that there was a class of patients showing clinical deterioration for both LI (GAD class 3) and HI (GAD class 2) interventions. It is of interest that these classes were only identified in GAD

trajectories, not PHQ. This suggests that the high probability of overall deterioration in the LP6 samples is likely to be in relation to anxiety symptoms, rather than depression, which may be worth considering both before and during treatment by the therapist. It is possible that these patients have anxiety disorders with symptoms not well captured by the GAD-7, e.g. specific phobias, and the start of treatment may exacerbate these symptoms and cause an increase in worry, therefore higher GAD-7 scores. Providing clinicians with the information that their patient appears to be following a trajectory indicating deterioration immediately, such as the 'not on track' method (Lambert et al., 2001), has the potential to inform a change in treatment and hopefully reduce the likelihood of a negative outcome from IAPT treatments. Whereas LP6 patient appear most a risk of deterioration in anxiety symptoms, the other two deterioration trajectories identified were identified for depression symptom change in HI treatment (LP1 PHQ class 2; LP2 PHQ class 2).

A number of these trajectories may represent typologies of psychotherapy patients that have been described by clinical researchers in the field. For example DeRubeis, Gelfand, German, Fournier, & Forand (2014) have suggested that a number of patients who 'respond' well to treatment will be "spontaneous remitters" for whom a resolution of psychological wellbeing will occur without treatment, as well as "easy" and "pliant" patients for whom a positive outcome is likely when in receipt of good quality therapy. Other trajectories may correspond to 'harder to treat patients' who may often be labelled as "challenging" or "intractable" patients and for whom a positive outcome is difficult to achieve if at all. It is possible that the distribution of these types of patients differs across profiles and may therefore inform the composition and outcomes of the classes observed in the LCGA analyses presented in this chapter. For example, a relatively high proportion of patients in 'early response' type trajectories (e.g. LP1 PHQ-9 GSH class 1 [figure 8.1]; LP1 GAD-7 HI class 1 [figure 8.2]; LP2 PHQ-9 GSH class 1 [figure 8.3]) in some profiles may be explained by these groups of patients containing more "spontaneous remitters", or more pliant patients. Profiles with higher incidences of 'non-response' trajectories (e.g. LP7 PHQ-9 HI class 1 [figure 8.12], LP8 GAD-7 HI class 4 [figure 8.14]) may be made up of more "intractable" patients, for whom psychological treatment is unlikely to result in a positive outcome. The identification of these types of patients, especially within profiles, could increase the clinical utility of a DST in IAPT services.

The different trajectories observed may also be in response to specific components of the therapies delivered. For example, a central component of behavioural activation includes a focus on engaging in activities that patients typically avoid (Veale, 2008), and other treatments for anxiety disorders may include exposure to anxiety provoking situations/items. A consequence of these components of treatment, is that the patient's initial symptoms may increase, or not decrease at the same rate as they would for patients whose treatment is not so focused on avoidance. This may, in part, explain the pattern of responses observed in

patients of the ‘early deterioration’ types classes, as well as those classes with limited early response or a more gradual reduction in symptoms. This type of responder may be present in LP8 PHQ-9 HI class 8 (figure 8.14) where there is no visible response to treatment in the first four sessions of treatment, before a quite dramatic reduction in symptoms occurs after the fourth session. If these classes were to be tested it might be worth exploring how the classes mapped onto clinician’s perceptions of patients using a typology such as that developed by DeRubeis and colleagues (2014).

One limitation of this work, as with the analysis presented in Chapter 7, is that anxiety disorder specific measures (ADSMs) were not available, and there may have been change (or lack of) in other symptoms not included in the GAD-7 that may be of value to inform monitoring decisions. ADSMs are not as routinely used as the GAD-7, and were not available in the dataset used for the current analysis, but it would be of benefit for services to increase their use. It is likely that ADSMs would provide important information that could also be included in the DST developed in this thesis and scores on these measures could be compared to GAD-7 change, therefore future work should explore the use of these measures in predicting response to IAPT delivered interventions.

A further limitation of this analysis is that the algorithm can only provide advice when either GSH or HI treatments are chosen. Although these treatments are used for the majority of patients in the dataset, it may be that other services frequently use other interventions, for example group interventions that were not considered in the current analysis. As discussed in Chapter 7, group-based interventions can vary considerably in purpose, which would make comparisons difficult and therefore analyses should be conducted within group types (e.g. only including trauma-focused groups). In addition, as so few members of LP3 were present in the dataset, this profile was excluded but as a result the DST would not be able to confidently suggest an expected response curve for this profile of patients, and instead would have to assume the trajectory of the profile which seems the most similar. Although the samples for group-based treatments and LP3 patients included insufficient numbers for analysis, the use of data from additional services or the national IAPT data could provide a large dataset with greater representation of these samples, which could further inform the prototype DST.

The prototype algorithm is limited at present to using just the nine patient characteristics included in the LPA, but it is likely that other characteristics related to the patients, as well as external factors, may contribute to treatment outcomes. For example, therapist characteristics may inform both the likely success of treatments and how symptoms change during treatment. As the LP method described in this chapter is used to identify sub-groups of patients with similar characteristics, it is possible that some therapists may have better outcomes for one profile of patients compared to another, because they are able to better identify with them or have more experience working with a specific group of patients. In addition, the prototype DST does not include diagnostic information and therefore there may

be situations were LI treatment is recommended but there is good reason to initiate at HI instead, for example if PTSD or social anxiety disorder is identified. As a result, the DST treatment recommendation will not replace clinical judgement, but instead provide an aid for it.

The DST presented in this chapter is able to provide information about the likely trajectories of change for members of a profile, which would allow a clinician to compare the progress of their current patient against. If there were two likely trajectories then it would likely be straight forward to decide whether the patient was following the 'responding' or 'non-responding' class. However, when there are four trajectories it may be more difficult to tell which trajectory is being followed. Further development of the DST would benefit from the inclusion of potential confidence parameters which can determine how close the current patient's trajectory is to all possible trajectories, and the algorithm could then identify the probability of a poor prognosis based on the expected trajectory. The use of more sophisticated statistical methods could provide the opportunity to further develop the algorithm in this way, and pattern-recognition based machine learning methods could provide the opportunity to identify and signal patients who are at risk of poorer outcomes (Mullainathan & Spiess, 2017).

The prototype DST presented in this chapter could be used to aid clinical decision making in IAPT services, and it would be feasible to build into existing EPMS used by the services. Research has found that DSTs were likely to have most impact when they were available as part of the clinical workflow, providing real time support (Kawamoto et al., 2005) and therefore it would be recommended that this tool were hosted in the background of the management system. However, this prototype algorithm has been developed using data provided by the services but has not considered the clinical impact of using such as tool in routine care. It is possible that clinicians in the services may not embrace the use of an algorithm to support their work, and it would be important to understand how clinicians view such a development before it was incorporated into routine care. In addition, the views of patients are also important to consider, as patients may be uncomfortable knowing that an algorithm was supporting decisions about their treatment. Further work leading on from the findings presented in this thesis should consider interviews with clinicians and patients on their views about using a DST in IAPT services, as this may shape the implementation of the prototype algorithm in routine care.

Summary

The results of these sub-analyses provide an additional component to a prototype DST for use in IAPT services by personalising the prediction of expected change during treatment based on the patient's identified profile. This information can be combined with latent profile outcome information to create a DST with an algorithm identifying latent profiles in the front

end of the system, and the probability of outcomes and trajectories of change provided as system-based outputs. The use of this DST has the potential to aid both treatment selection decisions and treatment monitoring decisions in psychological interventions, and could greatly increase healthcare efficiency. Considering the views of patients and clinicians with regard to the use of DSTs in services, as well as incorporating characteristics not included in the current analyses or the IAPT data may increase the clinical utility of the DST developed in this thesis.

Chapter 9. Review of findings and conclusions.

Review of findings.

The primary aim of this thesis was to explore the development of a personalised treatment approach for common mental health disorders (CMHDs), which could support clinical decision making about the choice of psychological interventions for people presenting to IAPT services. This was achieved through:

* A meta-review of the literature on patient characteristics associated with treatment outcomes in CMHDs.

* The use of regression modelling to explore the role of patient characteristics in predicting treatment outcomes within a dataset of patients attending IAPT services.

* Conducting latent profile analysis to identify statistically distinct subgroups of patients attending the IAPT services, and exploring differential outcomes between these profiles.

* Using latent growth curve and latent class growth analysis to identify the expected form of change in psychological symptoms during treatment, and to explore whether different trajectories existed between latent profiles.

* Combining information from the latent profile analysis and trajectories of change to inform the development of a decision support tool that could be used in clinical practice to recommend an appropriate intensity of psychological intervention, as well as providing information about the likely trajectory of change in psychological symptoms.

* Producing a prototype decision support tool that can be hosted by relevant patient management systems to inform both treatment selection and treatment monitoring decisions in IAPT services.

Personalised medicine and decision support tools

The use of personalised or stratified medicine to provide a more patient-centred approach to healthcare is increasingly being adopted, for example, in the treatment of cancer (Verma & Mukesh, 2012) and cystic fibrosis (Ashley, 2015). This approach aims to tailor interventions delivered to individual patients by incorporating information about patient characteristics, which may make them more or less likely to benefit from particular interventions, into clinical decision making (Dzau & Ginsburg, 2016). For example, advances have been made in the treatment of cancer by identifying individuals with specific genetic mutations which increase the likelihood of response to particular therapies (Rossi et al., 2011; Song, Lee, & Kang, 2011). Personalised medicine is now a key government priority for the English NHS (Keogh, 2015) and has the potential to not only improve patient outcomes but also contribute to a more efficient healthcare system by optimising treatment delivery.

Developments in the personalised medicine initiative have been supported by the growth of large healthcare datasets and the increased power and availability of digital technology, which has revolutionised the way patient information is collected and used (Wilson & Nicholls, 2015). Electronic patient management systems (EPMS) are widely used across healthcare and these systems have the potential to host large volumes of information on patient characteristics that have the potential to inform differential treatment response. A further benefit of large healthcare datasets is that they provide significant sample sizes for complex statistical modelling techniques, which are increasingly used in the development of personalised medicine approaches (Abbasi, 2017).

Whereas much of the personalised medicine initiative has focused on physical health conditions, especially cancer, mental health has currently received little attention, with few examples of stratified approaches (Ozomaro et al., 2013). This is despite 16% of the population estimated to be affected by common mental health disorders (CMHD) such as depression and anxiety (Spiers et al., 2016). One explanation for this may be the lack of convincing evidence for the role of various bio-markers and differential treatment response in CMHDs when compared to physical health conditions. In particular genetic markers, frequently used in areas such as cancer, have only limited evidence for their potential in predicting treatment outcomes in CMHDs, and therefore routine screening is neither justified or available at present (Licinio & Wong, 2011). However, a number of mental health services, in particular IAPT services, have large datasets on basic demographic characteristics, treatment delivery factors and outcomes variables. For this reason, the IAPT services were chosen as the focus for this study.

In the UK healthcare system, psychological interventions are recommended as treatment for CMHDs, and IAPT services are the common means by which psychological interventions are provided by the NHS in England. They deliver either Low Intensity (LI) or High Intensity (HI) evidence based psychological treatments, and over one million patients are referred to

IAPT services nationally each year (NHS Digital, 2016). IAPT services require a treatment selection decision (LI or HI) and also have high quality standardised data on patient characteristics as well as treatment outcomes, and so provide a good setting in which to develop and test a clinical decision support tool.

There are few examples of decision support tools (DSTs) available for use in mental health treatment (Sheehan & Sherman, 2012), which may be linked to the lack of identified bio-markers or large patient datasets (Ozomaro et al., 2013). However, a small number of DSTs have been delivered to support treatment monitoring decisions in mental health. Staged pharmacological treatment algorithms have been developed to inform clinicians which order of treatments to consider, where a lack of response to the first treatment results in the algorithm suggesting the next treatment in the sequence. Examples of these staged treatment algorithms include the Psychopharmacology Algorithm Project Harvard South Shore Program (Stein et al. 2012) and Texas Medication Algorithm Project (Crismon et al., 1999; Rush, Crismon, et al., 2003).

DSTs for treatment monitoring decisions in psychological interventions have adopted alternative approaches, such as the Quality Assurance (QA) system of Lambert and colleagues (Lambert, 2013). This DST does not recommend specific treatments, but instead uses sessional outcome measurement data to display the change in symptoms over time and flags up patients that the system considers ‘not on track’ based on their current trajectory. Identifying patients with a poor potential prognosis in this way provides the opportunity to change approach to treatment, with significantly reduced rates of deterioration observed in studies where information such as this is fed back to clinicians (Hannan et al., 2005; Shimokawa et al., 2010; Slade et al., 2008).

Whereas DSTs for treatment monitoring decisions are available for use in mental health treatment services, methods of supporting treatment selection decisions are currently have shown limited patient benefit when evaluated in treatment services (Watzke et al, 2010). However, a number of recent developments have seen attempts at modelling treatment outcomes using data from clinical trials, especially in the treatment of depression, and future evaluation of these methods in routine is expected in the coming years. For example, the Personalised Advantage Index (PAI; DeRubeis et al, 2014) was developed using an RCT dataset of patients receiving either psychological treatments or pharmacological treatments, and identifies, for individual participants in the dataset, what the likely benefit of one intervention over the other is likely to be. This method has been employed in other trial datasets (Huibers et al., 2015) but has not been evaluated in mental health services at present. Alternative predictive modelling approaches have been employed with other mental health datasets, for example creating a composite score from patient characteristics to inform the benefits of antidepressants compared to psychological treatment for depression (Wallace et al, 2013). However, as these analyses used small samples from more selective

trial datasets, and as these methods require recalculation for each new dataset, there are questions about their applicability to routine datasets such as IAPT.

A very recent paper by Delgadillo and colleagues (2016) used a dataset of n=1347 patients who received treatment at IAPT services in northern England to develop a risk calculation model (the 'Leeds Risk Index', LRI). This model uses patient characteristics included in the IAPT minimum dataset (MDS) as well as additional characteristics collected at these local IAPT services but are not collected nationally (e.g. family history), to group patients into either low, moderate or high risk groups. 'High risk' patients were less likely to achieve positive outcomes following treatment (defined as a reduction in 5 or more points on the PHQ-9 and 4 or more points on the GAD-7), compare to patients in the 'moderate' or 'low' risk groups. Outcomes were better for 'High Risk' patients when HI treatments were received, and the LRI can be used to identify patients for whom HI interventions would be recommended as initial IAPT treatment. Although the LRI has potential to support treatment selection decisions in IAPT services, one drawback of this system is that it uses a number of patient characteristics that are not included in the IAPT MDS, and therefore it could not be used by other national IAPT services unless they were collecting this additional patient information routinely.

The DSTs and modelling approaches discussed in Chapter 2 show potential for aiding either initial treatment selection or treatment monitoring decisions in services such as IAPT, however at present there is not a system that can aid both decisions. The focus of this thesis was to develop such a system, which has potential use for aiding clinical decisions in IAPT services. As these services provide the majority of psychological therapies for CMHDs in England and collect a standardised set of patient characteristics, the development of a DST to personalise treatment in these services could potentially have a significant impact in improving healthcare delivery.

Predictors of treatment outcomes.

A meta-review (a systematic review of previously conducted reviews) was conducted to understand which patient characteristics are associated with outcomes in CMHDs. The review identified a number of patient specific characteristics that had frequently been explored in relation to treatment outcomes, for example age, gender, severity of initial symptoms and comorbid mental health problems, including personality disorders. Some routinely collected demographic characteristics, such as age and gender, were either inconsistently associated with treatment response, or had limited/inconsistent evidence of an association with outcomes in CMHDs.

The level of pre-treatment symptom severity was frequently explored within the previous literature of predictors of treatment outcomes, and findings suggested that lower severity

was typically associated with better response to interventions for CMHDs, especially psychological treatments. The presence of comorbid conditions including personality disorders, having had previous treatment for CMHDs and a longer duration of illness were linked to poorer outcomes in the literature, across both pharmacological and psychological treatments. The review identified a number of patient characteristics which are routinely collected by IAPT services, some of which were consistently associated with outcomes (e.g. severity), some which showed an inconsistent association (e.g. age) and some characteristics which may have potential predictive ability but have been under researched (e.g. ethnicity, functional impairment).

The meta-review identified limited previous research on predictors of treatment dropout, with only one identified review exploring patient characteristics associated with dropout during psychological treatment. As dropout has been estimated to occur in a third of patients attending routine treatment services (Saxon et al., 2017; Wells et al., 2013), and is associated with poor patient outcomes and increased long term costs (Wade & Häring, 2010), considering the likelihood of dropout in treatment selection decisions could improve healthcare efficiency. From the identified systematic reviews, it was suggested that older age and less perceived stigma of mental health treatment were associated with lower dropout rates in a number of reviews, and age, gender, severity and previous treatment were more inconsistently associated with dropout across the literature.

One limitation of the systematic review conducted in Chapter 3 was that there was a large amount of variation between the included reviews, for example between which specific CMHD diagnoses (e.g. depression, anxiety, all CMHDs) and the types of primary studies the reviews included (e.g. cohort studies or only randomised trials). This may explain some of the inconsistencies found between patient characteristics and treatment outcomes. The limited number of reviews of characteristics associated with treatment dropout, especially for psychological treatments, suggests that this is an area in need of further research. As the variation between the included populations and studies may have contributed to the inconsistency in associations, the development of a DST for IAPT services was expected to be better informed by exploring the association of patient characteristics with outcomes in a dataset of patients attending these services. This would also provide an opportunity to explore the potential interactions between patient characteristics, as the meta-review only considered the association between individual patient characteristics and outcomes and ignored the potential impact of co-existing characteristics on treatment outcomes.

Aiding treatment selection decisions

As the aim of this thesis was to develop a DST for use in IAPT services, the next phase of the project was to supplement the findings of the meta-review with an analysis of routine patient data collected at two IAPT services. The use of this dataset was expected to provide

the most relevant information to inform the DST and help develop a personalised treatment approach for these services.

The analyses considered only routinely collected patient characteristics that were collected as part of the IAPT MDS, as using data already collected would limit the burden on clinician and patient time as well as increasing the applicability of the DST to IAPT services nationally. This dataset included data on basic demographics such as age, gender and ethnicity, as well as clinical data on symptom severity and level of functional impairment. The dataset for analysis consisted of n=16636 patients who entered treatment at the IAPT services, of whom n=10963 had both pre and post-treatment severity measures and were considered to be meeting 'caseness' (scoring over the clinical threshold for a CMHD) at assessment.

The first analysis (Chapter 4) used multiple regression methods to explore the association between available patient characteristics and a number of treatment outcomes, including IAPT recovery (scoring below clinical threshold following treatment), reliable change in symptom scores, clinical deterioration (a significant increase in symptom scores post-treatment) and treatment dropout. Latent profile analysis was then performed on the data to identify statistically distinct profiles of patients entering treatment at the services based on the available patient characteristics (Chapter 5), and the variation in outcomes between these profiles was explored in the final chapter in this section (Chapter 6).

Regression analysis

In the first stage of analysis, multiple regression analysis was performed to explore the association between routinely collected patient characteristics pre-treatment and eventual treatment outcomes. Nine patient characteristics were available in the IAPT dataset, which had all shown some association with outcomes in the meta-review conducted in Chapter 3, and included age, gender, pre-treatment severity of depression and anxiety, social impairment, ethnic group, welfare benefit status, presence of significant phobia symptoms and whether psychotropic medication was prescribed.

The results of the regression analysis showed that each of the nine patient characteristics were significantly associated with at least one of the investigated outcomes (e.g. recovery, reliable change, deterioration and dropout), across different intensities of psychological intervention. However, depression and anxiety severity, social impairment and welfare status were most frequently associated with treatment response across sub-analyses by intensity of psychological intervention. Age, depression severity and welfare status were significantly associated with treatment dropout, for all intensities of psychological treatment received.

Consideration of these patient characteristics at initial assessment could be used to inform likely prognosis from IAPT treatment. For example, patients with high levels of anxiety and

depression symptom severity were less likely to be in recovery following IAPT treatment and therefore this information could be used to inform treatment planning, potentially suggesting the need for more intensive treatment. The differences in results between predictors of LI and HI treatment outcomes could be used to inform which intensity might be most appropriate for a patient given their characteristics at presentation to the services. Findings suggested that receiving welfare benefits was linked to higher odds of deterioration in HI, whereas it was not significantly associated with an increase in symptoms during LI treatment. However, being from a non-white ethnic group was associated with deterioration in LI, but not in HI and therefore consideration of demographic characteristics may be important when deciding whether LI or HI treatment is most appropriate.

However, the ability of these results to inform a DST for use in IAPT was limited. The variance explained by the regression models was generally low, and only above the threshold that has been suggested to indicate a clinically meaningful benefit (Uher et al, 2012) for models predicting IAPT recovery. The variance explained by the models predicting reliable change and deterioration was low and was limited to just 1.5 to 3.5% in models predicting treatment dropout across the different intensities of psychological treatment.

Using regression analysis to predict the likelihood of multiple outcomes which may be of interest to a DST was also problematic as different characteristics were associated with different outcomes. Practically this could result in a situation where a constellation of patient characteristics suggested an increased probability of both a positive outcome (e.g. response) and a negative outcome (e.g. dropout) which would be difficult to reconcile in a DST. Instead, alternative statistical methods that can identify sub-groups of patients with similar characteristics were considered.

Latent profile analysis

To address the potential limitations of the regression analyses presented in Chapter 4, a method of identifying sub-groups or clusters of individuals with similar characteristics was considered. The use of clustering methods would allow for the interactions between all patient characteristics to be considered, but it could also replicate potential clinician heuristics where patients are compared to clinician-derived ‘prototypes’ (Garb, 2005). Identifying sub-groups of patients could inform a stratified approach to the selection of LI or HI treatments in IAPT

Latent profile analysis (LPA) (Lazarsfeld & Henry, 1968; Hagenaars & McCutcheon, 2002), an extension of latent class analysis for continuous data, was selected as an appropriate method, due to its use of model fit statistics and more robust classification when compared to alternatives such as K-means clustering (Schreiber & Pekarik, 2014). Latent profile methods have been used to identify sub-groups of patients with CMHDs (e.g. Rosellini &

Brown, 2014; Unick, Snowden & Hastings, 2009) but clinical outcomes between identified profiles/classes have not previously been explored, despite its potential for aiding personalised medicine approaches.

LPA was performed on the dataset of all patients who entered treatment at the IAPT services (n=16638) and included the same nine patient characteristic used in the regression analysis. Eight latent profiles (LPs) were identified in split-sample datasets, suggesting that co-occurrences of patient characteristics in the dataset indicate eight statistically distinct groups of patients entering treatment at the services (Chapter 5).

Examination of the eight profiles suggested that they were made up of quite different groups of people, which could provide important information about the casemix of patients receiving treatment at the services. Latent profile 1 (LP1) were younger than the average age for patients entering treatment (35.5 compared 38 years of age), had lower levels of pre-treatment depression and anxiety severity (PHQ score 5 vs 14; GAD score 5 vs 12) and were less likely to be receiving welfare benefits, reaching caseness for phobia or prescribed medication. In comparison, LP8 were more likely to be from a non-white ethnic group (31% LP8, 17% LP1), more likely to be female (72% vs 66%) and have higher levels of depression (mean = 19) and anxiety (mean = 16) severity pre-treatment. Descriptions of the eight profiles were shown to members of the clinical team at the IAPT services, and the clinicians indicated that they could see how the different profiles mapped onto their own caseloads, which may indicate potential clinical validity of these patient profiles.

In the next stage of analysis (Chapter 6), the probability of recovery, reliable change, recovery or improvement, deterioration and dropout were explored between the profiles, with sub analyses included by intensity of psychological treatment. The proportion of patients within each profile who were in recovery following treatment varied considerably between profiles. Whereas LP1 and LP3 reported recovery in around 75% of cases, only 15% of members of LP7 were in recovery following treatment. The probability of deterioration ranged from 5% for members of LP7 to 20% in LP6, which is significantly higher than the 5-10% deterioration reported in the literature (Boisvert & Faust, 2003; Crawford et al., 2016), and would suggest a group of patients who are at considerably greater risk of negative outcomes following treatment. Treatment dropout ranged from 17 to 40% across profiles. Profile membership was calculated for an additional validation sample, and this dataset was explored to confirm whether the probabilities of treatment outcome by LP were maintained in a separate sample of patients who attended the same services at a later date. The results indicated that the proportion of patients achieving outcomes in each profile was maintained over time, suggesting the profiles were robust in their predictive ability.

Outcomes between LI and HI treatments were also explored within profile to identify whether one intervention intensity was associated with a significantly higher chance of positive outcomes than the other. Results showed that the probability of recovery or reliable change

on both symptom measures was significant higher when HI interventions were received for four of the profiles (LP2, LP6, LP7 and LP8) but not for the other four profiles. Surprisingly for LP3 and LP4, results suggested that the odds of recovery were higher following LI interventions, although this difference was not statistically significant.

Given the differences in outcome likelihood between the profiles, incorporating profile identification into a DST may have potential utility in supporting personalised treatment selection decisions in IAPT services. The identification of LPs would enable a prediction of likely treatment outcomes from both LI and HI treatment, which could then be used to inform clinical decisions about appropriate treatment. For example, if at assessment a patient was identified as being a member of LP1 then, as the probability of a positive outcome (e.g. recovery) was similar between LI and HI treatment for this profile of patients, the less resource intensive intervention might be favoured and suggested as the most appropriate treatment for the patient. Conversely, members of LP7 were significantly more likely to achieve positive outcomes from HI treatment compared to LI, and therefore the more intensive treatment might be recommended as first line treatment.

Aiding treatment monitoring decisions

In the final section of the thesis, methods to support treatment monitoring decisions were explored. As IAPT services have adopted routine outcome measurement, the dataset used in the thesis included depression and anxiety symptom scores measured at each treatment contact, which allowed for the within-treatment change in symptoms to be statistically modelled.

The dataset used for this series of analyses included all patients who received at least three sessions of either guided self-help (LI treatment) or HI interventions. A total of n=3334 patients were included in the analysis of guided self-help symptom change, and n=4394 were included in the HI analysis.

Latent growth curve (LGC) modelling was conducted in the first stage of analysis, and this estimated the mean change in symptoms during treatment within the sample. Growth curves were also estimated for each LP to provide an expected response curve (mean change in symptoms) for each profile of patients (Chapter 7). Latent class growth analysis (LCGA) was then conducted to identify statistically different trajectories of change for all patients receiving either guided self-help and for patients receiving HI treatments. LCGA was then performed for each LP to identify the distinct trajectories that could occur within profiles to provide a more stratified estimate of likely change in symptoms (Chapter 8). The results of this analysis were combined with the findings presented in the previous section of the thesis to develop a prototype DST for use in IAPT services.

Latent growth methods

To inform the potential development of a DST that could aid both treatment selection and treatment monitoring decisions, latent growth curve (LGC) analysis was performed to identify the trajectories or expected response curves for depression and anxiety symptom change during IAPT treatments. Previous researchers have found a greater reduction in symptoms over the first few sessions of psychological treatment, indicated by a steeper curve, which levels out as the number of sessions increases (Cuijpers et al., 2013; Kopta et al., 1994), and this finding was replicated in growth curves estimated for patients receiving the included IAPT treatments.

Latent growth analysis was also performed for each of the eight latent profiles identified in Chapter 5, which estimated the mean change in symptom scores for groups of patients with similar characteristics. A similar form of change (shape of the curve) was identified between most of the profiles, with the major variation between profiles being the intercept or starting value pre-treatment symptom score. The change in anxiety symptoms for LP6 was the only trajectory which had a unique form of change and suggested that anxiety symptoms remain stable for the first few sessions for this profile of patients before reducing as the number of sessions increases. This suggests that LP6 patients may need longer in treatment before a reduction in symptoms is seen, and it could be recommended that clinicians should not be immediately concerned if anxiety does not decrease in the first few sessions for patients identified as LP6.

Latent growth curve analysis models the mean change in symptom scores for a sample, however there could be more potential value in identifying sub-groups of patients with statistically distinct forms of change within the sample. Latent class growth analysis (LCGA) was selected as a method of identifying sub-groups (classes) of individuals with similar trajectories within the full sample, as it was hypothesised that a number of different trajectories of change could exist, for example a group of treatment responders and a group of non-responders (who would so limited reduction in symptoms). By identifying different classes of patients with distinct trajectories, the progress of any new patient receiving treatment at the IAPT services could be compared against the previously identified trajectories, to predict whether a positive outcome is likely considering the patient's current trajectory. Patients following a trajectory that is indicative of a poor outcome, for example a non-responding class of patients, could be identified by the clinician and this might provide an opportunity to decide whether an alternative treatment approach should be considered.

The results of the LCGA for patients who received guided self-help (GSH) and those who received HI treatments identified four distinct classes of change in depression symptoms and five classes of anxiety symptom change. Although the same number of classes were identified, the trajectories themselves were different and suggested that the likely forms of change varied between the two intensities of intervention. Whereas depression symptom

change during GSH identified four classes with distinct intercept values, suggesting that the class could be identified for a patient within the first few sessions of treatment, three of the classes of depression change during HI treatment had a similar intercept value, with the trajectories only becoming distinct towards the fourth or fifth session. Whereas one depression symptom class showed a consistent PHQ-9 score during the course of treatment, indicating no change in symptom severity, a second class showed no initial change in the first three sessions before a dramatic decrease in depression symptoms begins. This finding cautions against the view that prognosis can be identified by the third session/second week of treatment (Gorwood et al., 2013; Lambert, 2013) and instead suggests that, for some sub-groups patients at least, clinicians should wait until the fourth or fifth session before considering a change in approach.

To further clarify this issue, LCGA was then performed by individual latent profile to identify all potential trajectories of change for stratified groups of patients. For some profiles, no additional classes were identified, suggesting that there was only one form of change (the expected response curve), but for other profiles up to four distinct classes were identified. The distribution of patients to these curves varied across profiles, and typically one class included around 40-50% of profile members with the other classes consisting of 10-35% of the sample. Providing this information to clinicians could be used to support a more personalised approach to treatment monitoring decisions, as a patient's within-treatment change could be plotted against the likely known trajectories for members of their profile. This might inform a discussion between the patient and the clinician as to whether a change in treatment appropriate may be beneficial, which could result in more appropriate treatment being made available earlier.

Application in practice

As indicated above, combining the results of the LPA (Chapters 5 and 6) and the LCGA by latent profiles (Chapter 8) was used to develop a DST to support both treatment selection and monitoring decisions, and thereby support the delivery of a personalised medicine approach to the treatment of CMHDs. For any new patient entering an IAPT service, the DST would first use information from the nine routinely collected characteristics described in Chapter 4 to calculate the probability of profile membership to each of the eight latent profiles, using the posterior membership probability equations in Chapter 5. The patient would then be assigned to the profile to which they had the highest probability of membership using an algorithm incorporating the probability calculations, and this algorithm was created for the analyses in Chapter 6 to assign each patient in the validation sample to a latent profile. This algorithm could be hosted by the EPMS used by a service, and as the nine characteristics are already part of the mandatory IAPT MDS, no additional data collection by patients or clinicians would be required.

Details of the patient's latent profile could be provided in real time to the clinician (and potentially to the patient with further development), as soon as the assessment is complete. This profile information could include the probability of different treatment outcomes, including recovery, deterioration and dropout, following both LI and HI interventions, and would be informed by the analyses presented in Chapter 6. The DST could also be developed to use the information about likely treatment outcomes to provide a treatment recommendation (e.g. LI or HI treatment) to aid clinical judgement. This recommendation might be made available to the patient and could inform a discussion between the clinician and patient about treatment choice. However, this system could only inform a judgement, not replace it. There are a number of reasons for this, including the absence of diagnosis information which would inform an automatic allocation to HI treatment for PTSD or social anxiety disorder in IAPT services, due to the limited evidence-base for LI treatments for these conditions. Other patient characteristics such as a high risk of self-harm or having previous unsuccessful LI treatment would likely result in the clinician considering HI as a more appropriate intervention and therefore the algorithm would need further development to incorporate these characteristics.

In the second stage of the algorithm, once the latent profile has been identified and treatment has commenced, information about the current trajectory of the patient, collected through sessional outcome measurement, could be compared to the likely trajectories for patients of the same profile (as identified in Chapter 8). Sessional outcome data is already entered into the EPMS as part of routine practice in IAPT services, and therefore the DST could be built into the system to provide information about the patient's progress in real time. This information could flag up when the patient is following a trajectory that would indicate a poor prognosis, and therefore provide an opportunity for the clinician to change approach if appropriate, or step up to a higher intensity of treatment.

In summary, the proposed DST outlined above has the potential benefit of relatively simple technical implementation as it could be hosted on EPMS with minimal effort, and as the included patient characteristics are already part of the mandatory dataset, there would be no additional staff resource needed. In return, the DST can provide the probability of a range of treatment outcomes, including both positive and negative outcomes, and produce a recommendation about whether LI or HI might be more appropriate for the patient. Once treatment has started, the DST could use information about the likely trajectories of symptom change for the patient's identified profile to inform the clinician whether the patient is displaying a change in symptoms indicative of clinical benefit, or whether a poor prognosis is likely. This personalised approach to treatment could be delivered in real time and inform shared decision making between the clinician and patient, and has the potential to contribute to a more efficient use of healthcare resources.

Limitations

A major limitation of this project is that the datasets used were drawn from two IAPT services only and were therefore limited with regards to available patient characteristics, information about how treatments were chosen, and the treatments available.

The lack of randomisation of patients who received LI and HI treatments meant that the exploration of the potential prescriptive effect of patient characteristics in the IAPT dataset could not be explored in this thesis. Instead, the focus was on identifying patient characteristics which were prognostic within IAPT delivered treatments. There are a number of reasons why certain patients would have been allocated to HI treatments instead of LI. As the analyses presented in this thesis were limited to considering only data routinely collected by the IAPT services this meant, along with the absence of randomisation, it was not possible to explore these factors. T-tests were conducted on the differences in mean symptom severity between members of each profile who received LI or HI treatments, with limited differences found, but there are likely to be further indicators of patient complexity or severity which were not captured in the IAPT data set that may have informed treatment allocation decisions for patients within the same profile to receive different treatments. The use of randomised allocation to LI or HI treatment would have eliminated the risk of confounding by indication but this method was not feasible within the current project.

The IAPT MDS is limited with regard to the number of patient characteristics that are available, and there may be a number of important factors that could have informed the treatment selection decisions that were made, and also be able to inform the likely response to psychological treatment. A number of characteristics were identified by the meta-review presented in Chapter 3, such as relationship status and past treatment history, that were associated with response to treatment, but these characteristics were not available in the IAPT MDS. The inclusion of these characteristics may have further informed the likelihood of treatment outcomes, as well as the structure of the latent profiles.

The absence of reliable diagnostic information in the IAPT datasets is also presents a significant limitation. The allocation of treatment in IAPT services are be driven by diagnosis in some scenarios, as diagnoses such as PTSD and social anxiety disorder where current NICE guidance recommends high intensity treatments as first line interventions (e.g. NICE, 2013). The IAPT data is very limited regarding diagnostic information, with diagnosis or ‘problem description’ missing for over half of IAPT patients in national datasets (HSCIS, 2015). The improved identification of clinical diagnoses in IAPT services would help further inform treatment selection decisions, especially given the national guidance for treatment of specific CMHD diagnoses.

By focusing the current project on IAPT data only, the results of this analysis have limited generalisability to other treatment settings and would require replication in alternative settings to explore their potential to aiding treatment decision in other settings. IAPT services

predominately deliver CBT-based interventions, as these treatments have the largest evidence base, but other treatment modalities have been shown to be effective (NCCMH, 2011). The number of patients in the two IAPT services receiving other modalities such as IPT (n=63, 0.59%) or Couples Therapy (n=37, 0.35%) was very limited, but it would be of interest to explore whether there was a method of personalising which type of treatment modality, rather than just intensity, would be of most benefit to a patient. Alternative datasets from different treatment services, or IAPT services where other modalities are more commonly used could further evolve the personalisation of psychological treatment beyond the scope of this project.

Although the dataset included a relatively large number of patients, the data was derived from only two IAPT services in England, and replication of these analyses in other IAPT services may result in different findings. The results of the LPA, comparison of outcomes and LCGA may not be generalizable to other IAPT services, or other non-IAPT psychological treatment services which may provide services for patients with more complex needs than those typically treated in IAPT services, and therefore replication of these methods in different datasets would be of benefit. It is also to be expected that local IAPT services will vary in relation to local demographics, such as mean age and ethnic backgrounds, and therefore different profile make-ups or even number of identified profiles could exist if alternative datasets were explored. The probability of different outcomes by individual profile could also vary between services, which may relate to the nature or quality of care provided for certain groups of patients, and again further analysis would be of interest to explore this. Although a number of patients treated in IAPT services are prescribed medication (nearly 40% of the included sample), adherence to medication was not routinely collected and so the influence of medication on profile response could not be explored. This may be worth considering in the latent growth modelling in future analysis.

Using data from additional services could also be used to provide a more substantial dataset for some of the sub-analyses that could not be performed in this thesis. For example, the analysis on treatment outcomes for LP3 presented in Chapter 6 included a very limited number of patients, and there were too few LP3 patients that could be included in the latent growth analyses in Chapters 7 and 8. As a result, the prototype DST is limited as to its advice for patients from this profile, and datasets providing more LP3 patients could allow findings for this profile to be interpreted and disseminated with more confidence. In addition, group-based treatments were not evaluated in the latent growth analyses as these interventions could vary with respect to clinical focus (e.g. relapse prevention, bereavement) and therefore included populations. However, acquiring data from more IAPT services could achieve data on a significant number of patients who received different formats of group treatment. Latent growth analysis could be performed by group-based treatment formats (e.g. trauma-focused groups, insomnia support) and results of these analyses might provide the DST with the ability to predict the expected treatment response by different types of

group intervention, and therefore provide highly personalised decision support based on treatment factors. The lower numbers of available patients in specific profiles meant that no validation sample was available for the latent growth analyses, and further analysis would be recommended to explore whether the same classes were identified within profiles from data in other services.

A further limitation of the analyses presented in this thesis is that outcome data was missing for a large number of patients, which meant they were excluded from analyses of treatment outcomes. Of n=16636 patients entering treatment, n=3252 (19.55%) had no further data available other than their initial assessment and were therefore missing any T2 data. For many patients only one treatment session would be required from IAPT, especially for brief self-help interventions or when advice only was provided, as patients may be expected to take the materials away to complete in their own time without the need for follow-up. Although this is in line with treatment guidance, it meant that for many patients receiving these types of intervention the outcomes of treatment were unknown, and it is possible that there may be subgroups of patients for whom this treatment was not effective. It would be recommended that services follow up patients receiving these types of intervention to understand whether treatment had been effective, as this could help personalise treatment delivered by the services,

The lack of follow up data also limits the amount of information we have about the long-term effectiveness of the IAPT-delivered treatments. The current analysis was focused on change within treatment episodes of patients attending IAPT services and predicting outcomes in these episodes only. However, data was not available to understand whether the benefits of treatment were sustained. Some the identified profiles may be associated with differential relapse/recurrence patterns, which could have important clinical implications for the delivery of relapse prevention materials. It is also possible that the benefits of HI treatment over LI that was found in some profiles may not have been sustained long term, which would need incorporating into treatment selection decisions. It would be of value to increase the availability of routine follow up data in services such as IAPT, and especially to explore predictors of relapse/recurrence to inform the likelihood of future service usage.

The use of the GAD-7 as a measure of anxiety symptoms also has potential limitations, as it is limited in its ability to capture all aspects of the range of anxiety disorders treated in IAPT services. Anxiety disorder specific measures (ADSM) are recommended by IAPT services to assessment specific anxiety disorders, but these are not used consistently in all IAPT services and were not available in the dataset used for this thesis. Using ADSM has the potential to further inform the trajectories of change for individual profiles, and it may be of value to explore this in future IAPT datasets where there will be more completion of these additional measures.

In addition, existing clinical decision rules developed by the service which may reflect local priorities, for example the management of waiting times, may have influenced choice of initial treatment, and this may limit the applicability and generalizability of the current DST across settings. For example, IAPT services will have only a limited amount of HI therapists available and therefore may not be able to allocate significant numbers of patients to higher intensity treatments, which may result in some DST recommendations for more intensive treatment being difficult to implement. It is likely that, for a DST to be successfully incorporated into routine clinical practice in IAPT services, it will need to take into account available resource as well as local decision rules to improve potential uptake and applicability.

Future directions

A number of possible future directions for further research and development are set out below, these include:

- Testing the clinical applicability of the algorithm output as a DST.
- Replicating the analysis using data from additional services, including non-IAPT settings.
- Refining the profiles by using additional patient characteristics.
- Expanding the use of the algorithm to involve patient participation in decision making, and increasing access to the algorithm output.

Testing the clinical applicability of the algorithm - initial discussions with clinicians from the IAPT services which provided the data used in this thesis suggested that the profiles presented have reasonable face validity. However this would need to be tested empirically, and such testing would have a minimum of two components i) the refinement of the algorithm and its integration with existing EPMS to develop outputs which provide real time information to support clinical decisions and ii) exploring staff opinion and experience of using the DST in IAPT settings through feasibility studies, as well as formal evaluations (for example, controlled trials) to understand whether using the DST can lead to improved patient and service-level outcomes. Initial discussions with both the service in which this initial development was undertaken and with the providers of the EPMS have been positive. If these feasibility studies were positive, a more formal evaluation of the DST could be considered which may take into account the outcome of other further work described below.

Testing the profile in different settings – as discussed in the limitations section, there would be benefits in replicating this work in alternative IAPT services, as both the analyses

presented in this thesis and a potential feasibility study would have all been conducted within the same services. Following the presentation of the latent profile analysis and related publications, a number of IAPT services have expressed an interest in replicating this work with their data. Initial work with other IAPT services suggests that profile structures and distribution are similar to those identified in this thesis. More substantial development work could focus on the use of the statistical methods used in this thesis in non-IAPT populations with depression and anxiety disorders, including those in receipt of psychological and possibly pharmacological interventions. These methods could also be applied to other mental health conditions, for example schizophrenia, if sufficient patient datasets were available for these populations.

Refining the profiles – There is potential future development work that could focus on collecting and incorporating additional patient characteristics into analyses, such as those patient characteristics identified in the systematic review (Chapter 3), which are not currently included in the IAPT mandatory dataset. In addition to characteristics such as personality disorders and previous treatment, other factors such as relationship quality and self-efficacy (Luszczynska et al., 2009; Mululo et al., 2012) may have potential to further inform prognosis, and future work may wish to further explore these additional characteristics in both IAPT and non-IAPT samples. Including these characteristics within the latent profile analysis could be used to adjust the structure of the current profiles, or identify additional profiles, which may result in more refined and personalised information to aid clinical decision making.

The use of item level information from the PHQ-9 and GAD-7 assessments could also have further predictive validity than using the total scores alone, as has been presented in the current analysis. Item-level data on these measures was not available for the current analysis, but as this information would be provided by the patient during clinical contacts with the services, it should be possible to acquire and record this data in further research. As prediction models have identified specific clusters of symptoms that are associated with treatment outcomes (e.g. Chekroud et al., 2016), it may be that identifying sub-groups of patients with specific clusters of symptoms could be combined with the latent profile information presented in this thesis to create a more accurate and tailored prediction of likely treatment outcome. This has not currently been explored, but has potential to develop a highly personalised approach to treatment in CMHDs.

In addition, there is more patient relevant information available within the profile membership allocation that has not been explored in this thesis, but has the potential to be used to refine the DST recommendations. The latent profile algorithm calculates the probability of membership to each LP for each patient using their patient characteristics, and the profile with the highest probability of membership is allocated to the patient. In many cases this probability will not be 100% and therefore each patient will have also have a probability of membership to additional profiles, which could be defined as their ‘secondary profile(s)’. It is

possible that different members of the same ‘primary LP’ (the profile to which they have the highest probability of membership, e.g. LP2) will have different ‘secondary profiles’ (e.g. LP2/LP1 and LP2/LP3), and there may be differences in the probability of treatment outcomes between members of the same LP who have different ‘secondary profiles’. This analysis has not been explored previously, and there is no mention of using secondary profiles in this way in other datasets, but it has the potential to provide a more detailed stratification of patients attending the IAPT services.

Finally, the growth of computational statistical and modelling techniques provides an opportunity to further explore the role of patient characteristics in the prediction of outcomes in mental health treatment. Although this project has focused on latent variable modelling approaches, alternative statistical approaches such as machine learning (Friedman, 2006) could provide alternative solutions that may further improve the prediction of treatment outcomes. Further analysis using additional and more sophisticated approaches may provide better solutions and are planned as the next stage of analysis in this program of work.

Expanding the use of the algorithm to involve patient participation – the focus on almost all DSTs is on their use by clinicians, and this has been the focus of the work presented in this thesis. However, given the relatively accessible data that the calculation of profile membership is based on, it would be possible to develop outputs for the algorithm that could be used to inform patients of the potential treatment options available to them and the likelihood of their success. If developed in collaboration with patient groups, such outputs could be tested in a similar way to that suggested for the feasibility testing of the clinician focused algorithm, to see if use of the DST has a positive impact on service outcomes and healthcare efficiency.

Conclusions

The aim of this project was to explore the underlying data requirements and analysis to support the development of a decision support tool to aid clinical decision making in psychological treatment for CMHDs, by using routine patient data. A literature search identified a number of patient specific characteristics that have been associated with treatment outcomes, and there was significant overlap between these variables and the IAPT MDS. Exploration of a dataset of patients attending IAPT services found that a number of these characteristics were predictive of response, deterioration and dropout from treatment. Latent profile analysis was performed on the dataset to identify stratified groups of patients with similar presenting characteristics, and analysis found that there was significant variation in outcomes between these profiles. For some profiles there was a significant benefit of HI interventions over LI, which would suggest that these patients should

be allocated to higher intensity treatments as initial treatment rather than receiving an LI treatment which has a lower likelihood of resulting in a positive outcome.

Exploring the trajectories of symptom change for the eight identified profiles using latent class growth analysis showed that there was variation between the groups of patients on the likely change in symptoms during both LI and HI treatments. By combining the results of the latent profile analysis and latent class growth analysis, a prototype decision support tool that could be developed to inform both treatment selection and treatment monitoring decisions is presented. This DST could be incorporated into the existing electronic patient management systems used by the services and would provide real time information to clinicians to aid clinical judgement. A feasibility study evaluating the use of this DST is suggested, and further analyses using datasets from other IAPT services is proposed to replicate the methods and findings presented in this thesis. If feasible, the use of a personalised medicine approach developed as part of this project has the potential to optimise the delivery of psychological treatments in IAPT services.

References

Abbasi, J. (2017). 23andMe, Big Data, and the Genetics of Depression. *JAMA*, *317*(1), 14. https://doi.org/10.1001/jama.2016.14136

Ackerman, D. L., & Greenland, S. (2002). Multivariate meta-analysis of controlled drug studies for obsessive-compulsive disorder. *Journal of Clinical Psychopharmacology*, *22*(3), 309–317. https://doi.org/10.1097/00004714-200206000-00012

Adam, D. (2013). Mental health: On the spectrum. *Nature*, *496*, 416–418. https://doi.org/10.1038/496416a

Al Qasem, A., Smith, F., & Clifford, S. (2011). Adherence to medication among chronic patients in Middle Eastern countries: Review of studies. *Eastern Mediterranean Health Journal*, *17*(4), 356–363. Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L361860871%5Cnhttp://www.emro.who.int/emhj/V17/04/17_4_2011_0356_0363.pdf%5Cnhttp://resolver.ebscohost.com/openurl?sid=EMBASE&issn=10203397&id=doi:&atitle=Adherence+to+medication+amo

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. *Diagnostic and Statistical Manual of Mental Disorders.* https://doi.org/10.1176/appi.books.9780890425596.744053

Andruff, H., Carraro, N., Thompson, A., & Gaudreau, P. (2009). Latent class growth modelling: A tutorial. *Tutorials in Quantitative Methods for Psychology*, *5*(1), 11–24. https://doi.org/10.20982/tqmp.05.1.p011

Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, *77*(4), 693–704. https://doi.org/10.1037/a0016062

Arnow, B. A., Blasey, C., Manber, R., Constantino, M. J., Markowitz, J. C., Klein, D. N., … Rush, A. J. (2007). Dropouts versus completers among chronically depressed outpatients. *Journal of Affective Disorders*, *97*(1–3), 197–202. https://doi.org/10.1016/j.jad.2006.06.017

Ashley, E. (2015). The precision medicine initiative: a new national effort. *JAMA*, *313*(21), 2119–2120. https://doi.org/10.1001/jamaoncol.2014.216.5

Asparouhov, T., & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, *14*.

Ayuso-Mateos, J. L., Nuevo, R., Verdes, E., Naidoo, N., & Chatterji, S. (2010). From depressive symptoms to depressive disorders: The relevance of thresholds. *British Journal of Psychiatry*, *196*(5), 365–371. https://doi.org/10.1192/bjp.bp.109.071191

Bajor, L. A., Ticlea, A. N., & Osser, D. N. (2011). The Psychopharmacology Algorithm Project at the Harvard South Shore Program: An Update on Posttraumatic Stress Disorder. *Harvard Review of Psychiatry*, *19*(5), 240–258. https://doi.org/10.3109/10673229.2011.614483

Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., … Reis, B. Y. (2017). Predicting suicidal behavior from longitudinal electronic health records. *American Journal of Psychiatry*, *174*(2), 154–162. https://doi.org/10.1176/appi.ajp.2016.16010077

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: a comparison of statistical techniques. *Journal of Personality Assessment*, *82*(1), 60–70. https://doi.org/10.1207/s15327752jpa8201_11

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418.

Beard, C. (2011). Cognitive bias modification for anxiety: current evidence and future directions. *Expert Review of Neurotherapeutics*, *11*(2), 299–311. https://doi.org/10.1586/ern.10.194

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck depression inventory-II. *San Antonio, TX: Psychological Corporation*, 1–82.

Bender, D. S. (2005). The therapeutic alliance in the treatment of personality disorders. *Journal of Psychiatric Practice*, *11*(2), 73–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15803042

Berlin, K. S., Parra, G. R., & Williams, N. A. (2014). An introduction to latent variable mixture modeling (Part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology*, *39*(2), 188–203. https://doi.org/10.1093/jpepsy/jst085

Berner, E. S., & Lande, T. J. La. (2007). Overview of Clinical Decision Support Systems. *Decision Support Systems*, *6*, 463–477. https://doi.org/10.1007/978-0-387-38319-4_1

Bhugra, D. (2008). Decision-making in psychiatry: What can we learn? *Acta Psychiatrica Scandinavica*, *118*(1), 1–3. https://doi.org/10.1111/j.1600-0447.2008.01220.x

Blom, M. B. J., Spinhoven, P., Hoffman, T., Jonker, K., Hoencamp, E., Haffmans, P. M. J., & van Dyck, R. (2007). Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *Journal of Affective Disorders*, *104*(1–3), 119–126. https://doi.org/10.1016/j.jad.2007.03.010

Boessen, R., Groenwold, R. H. H., Knol, M. J., Grobbee, D. E., & Roes, K. C. B. (2012). Classifying responders and non-responders; does it help when there is evidence of differentially responding patient groups? *Journal of Psychiatric Research*, *46*(9), 1169–1173. https://doi.org/10.1016/j.jpsychires.2012.05.005

Boisvert, C. M., & Faust, D. (2003). Leading researchers' consensus on psychotherapy research findings: Implications for the teaching and conduct of psychotherapy. *Professional Psychology: Research and Practice*, *34*(5), 508–513. https://doi.org/10.1037/0735-7028.34.5.508

Bostwick, J. M., & Pankratz, V. S. (2000). Affective disorders and suicide risk: A reexamination. *American Journal of Psychiatry*, *157*(12), 1925–1932. https://doi.org/10.1176/appi.ajp.157.12.1925

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., … Liu, E. T.-H. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *The BMJ*, *346*, f540. https://doi.org/10.1136/bmj.f540

Bright, T. J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., … Lobach, D. (2012). Effect of Clinical Decision-Support Systems. *Annals of Internal Medicine*, *157*(1), 29. https://doi.org/10.7326/0003-4819-157-1-201207030-00450

Brixner, D., Biltaji, E., Bress, A., Unni, S., Ye, X., Mamiya, T., … Biskupiak, J. (2016). The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. *Journal of Medical Economics*, *19*(3), 213–228. https://doi.org/10.3111/13696998.2015.1110160

Brown, G. W., Harris, T. O., Kendrick, T., Chatwin, J., Craig, T. K. J., Kelly, V., … Uher, R. (2010). Antidepressants, social adversity and outcome of depression in general practice. *Journal of Affective Disorders*, *121*(3), 239–46. https://doi.org/10.1016/j.jad.2009.06.004

Brown, T. A., Di Nardo, P. A., Lehman, C. L., Campbell, L. A., We, O., Conklin, B., … Barlow, D. (2001). Reliability of DSM-IV Anxiety and Mood Disorders: Implications for the Classification of Emotional Disorders. *Journal of Abnormal Psychology*, *110*(1), 49–58. https://doi.org/10.1037//0021-843X.110.1.49

Bucholz, K. K., Hesselbrock, V. M., Heath, A. C., Kramer, J. R., & Schuckit, M. A. (2000). A latent class analysis of antisocial personality disorder symptom data from a multi-centre family study of alcoholism. *Addiction*, *95*(4), 553–567. https://doi.org/10.1046/j.1360-0443.2000.9545537.x

Bybee, T. S., Lambert, M. J., & Eggett, D. (2007). Curves of expected recovery and their

predictive validity for identifying treatment failure. *Tijdschrift Voor Psychotherapie*, *33*(6), 272–281. https://doi.org/10.1007/BF03062308

Cadarette, S. M., Jaglal, S. B., Kreiger, N., McIsaac, W. J., Darlington, G. A., & Tu, J. V. (2000). Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. *Canadian Medical Association Journal*, *162*(9), 1289–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10813010

Callard, F., Bracken, P., David, A. S., & Sartorius, N. (2013). Has psychiatric diagnosis labelled rather than enabled patients? *BMJ*, *347*, f4312. Retrieved from http://www.bmj.com/content/347/bmj.f4312

Carey, M., Jones, K., Meadows, G., Sanson-Fisher, R., D'Este, C., Inder, K., … Russell, G. (2014). Accuracy of general practitioner unassisted detection of depression. *The Australian and New Zealand Journal of Psychiatry*, *48*(6), 571–8. https://doi.org/10.1177/0004867413520047

Chan, S. W. Y., & Adams, M. (2014). Service Use, Drop-Out Rate and Clinical Outcomes: A Comparison Between High and Low Intensity Treatments in an IAPT Service. *Behavioural and Cognitive Psychotherapy*, *42*(6), 747–759. https://doi.org/10.1017/S1352465813000544

Chapman, G. B., & Sonnenberg, F. A. (2000). *Decision Making in Health Care: Theory, Psychology, and Applications*. Cambridge: Cambridge University Press.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. 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

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

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.1194

Clapp, J. D., Grubaugh, A. L., Allen, J. G., Mahoney, J., Oldham, J. M., Fowler, J. C., … Frueh, B. C. (2013). Modeling trajectory of depressive symptoms among psychiatric inpatients: A latent growth curve approach. *Journal of Clinical Psychiatry*, *74*(5), 492–499. https://doi.org/10.4088/JCP.12m07842

Clarke, K., Mayo-Wilson, E., Kenny, J., & Pilling, S. (2015). Can non-pharmacological interventions prevent relapse in adults who have recovered from depression? A

systematic review and meta-analysis of randomised controlled trials. *Clinical Psychology Review*, *39*, 58–70. https://doi.org/10.1016/j.cpr.2015.04.002

Clifford, P. (1998). M is for outcome: The CORE outcomes initiative. *Journal of Mental Health*, *7*(1), 19–24. https://doi.org/10.1080/09638239818300

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

Connolly Gibbons, M. B., Kurtz, J. E., Thompson, D. L., Mack, R. A., Lee, J. K., Rothbard, A., … Crits-Christoph, P. (2015). The effectiveness of clinician feedback in the treatment of depression in the community mental health system. *Journal of Consulting and Clinical Psychology*, *83*(4), 748–59. https://doi.org/10.1037/a0039302

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). *The British Journal of Psychiatry*, *176*(4), 379–386. https://doi.org/10.1192/bjp.176.4.379

Cox, D. R. (1957). Note on Grouping. *Journal of the American Statistical Association*, *52*(280), 543–547. https://doi.org/10.1080/01621459.1957.10501411

Cox, D. R. (1984). Interaction. *International Statistical Review*, *52*, 1–31. https://doi.org/10.1126/science.ns-9.225.507

Crawford, M. J., Thana, L., Farquharson, L., Palmer, L., Hancock, E., Bassett, P., … Parry, G. D. (2016). Patient experience of negative effects of psychological treatment: results of a national survey. *The British Journal of Psychiatry*, *208*(3), 260–265. https://doi.org/10.1192/bjp.bp.114.162628

Crismon, M. L., Trivedi, M., Pigott, T. A., Rush, A. J., Hirschfeld, R. M. A., Kahn, D. A., … Thase, M. E. (1999). The Texas medication algorithm project: Report of the Texas consensus conference panel on medication treatment of major depressive disorder. *Journal of Clinical Psychiatry*, *60*(3), 142–156.

CSIP. (2007). *IAPT Outline Service Specification*. Retrieved from https://www.uea.ac.uk/documents/246046/11991919/iapt-pathfinder-outline-service-specification.pdf/9fc11891-ecc5-48e4-a974-9bbc8ab0c690

Cuijpers, P. (2014). Personalized treatment for functional outcome in depression. *Medicographia*, *36*, 476–481. Retrieved from http://www.medicographia.com/2015/06/personalized-treatment-for-functional-outcome-in-depression/

Cuijpers, P., & Christensen, H. (2017). Are personalised treatments of adult depression finally within reach. *Epidemiology and Psychiatric Services*, *26*(1), 40–42.

https://doi.org/https://doi.org/10.1017/S204579601600007X

Cuijpers, P., Hollon, S. D., van Straten, A., Bockting, C., Berking, M., & Andersson, G. (2013). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, *3*(4), e002542. https://doi.org/10.1136/bmjopen-2012-002542

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2014). Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *World Psychiatry*, *13*(1), 56–67. https://doi.org/10.1002/wps.20089

Cuijpers, P., van Straten, A., Smit, F., & Andersson, G. (2009). Is psychotherapy for depression equally effective in younger and older adults? A meta-regression analysis. *International Psychogeriatrics*, *21*(1), 16–24. https://doi.org/10.1017/S1041610208008089

Cuijpers, P., Van Straten, A., Warmerdam, L., & Smits, N. (2008). Characteristics of effective psychological treatments of depression: a metaregression analysis. *Psychotherapy Research : Journal of the Society for Psychotherapy Research*, *18*, 225–236. https://doi.org/10.1080/10503300701442027

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

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognitive Development*, *11*(2), 121–136. https://doi.org/10.1080/15248371003699969

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Cuthbert, B. N., & Kozak, M. J. (2013). Constructing constructs for psychopathology: The NIMH research domain criteria. *Journal of Abnormal Psychology*, *122*(3), 928–937. https://doi.org/10.1037/a0034028

Davis, L. L., Frazier, E. C., Williford, R. B., & Newell, J. M. (2006). Long-term pharmacotherapy for post-traumatic stress disorder. *CNS Drugs*, *20*(6), 465–476. https://doi.org/10.2165/00023210-200620060-00003

Davis, M. L., Smits, J. A., & Hofmann, S. G. (2014). Update on the efficacy of pharmacotherapy for social anxiety disorder: a meta-analysis. *Expert Opinion on Pharmacotherapy*, *15*(16), 2281–2291. https://doi.org/10.1517/14656566.2014.955472

Deber, R. B., Kraetschmer, N., & Irvine, J. (2014). What Role Do Patients Wish to Play in Treatment Decision Making? *Archives of Internal Medicine*, *156*(13), 1414–1420. https://doi.org/10.1001/archinte.1996.00440120070006

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

Department of Health. (2008). *Improving Access to Psychological Therapies (IAPT) Commissioning Toolkit*. https://doi.org/10.3399/bjgp09X454043

Department of Health. (2011). *No health without mental health*. London: Department of Health.

Department of Health. (2014). Access to psychological therapies campaign - GOV.UK. Retrieved August 25, 2017, from https://www.gov.uk/government/news/access-to-psychological-therapies-campaign

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE*, *9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others-and some groups of therapists less-about what matters in psychotherapy. *Psychotherapy Research*, *24*(3), 419–428. https://doi.org/10.1080/10503307.2013.838654

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nature Reviews. Neuroscience*, *9*(10), 788–96. https://doi.org/10.1038/nrn2345

Díaz-García, A., González-Robles, A., Fernández-Álvarez, J., García-Palacios, A., Baños, R. M., & Botella, C. (2017). Efficacy of a Transdiagnostic internet-based treatment for emotional disorders with a specific component to address positive affect: Study protocol for a randomized controlled trial. *BMC Psychiatry*, *17*(1), 145. https://doi.org/10.1186/s12888-017-1297-z

Dodd, S., & Berk, M. (2004). Predictors of antidepressant response: A selective review. *International Journal of Psychiatry in Clinical Practice*, *8*(2), 91–100. https://doi.org/http://dx.doi.org/10.1080/13651500410005423

Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., … Gabrieli, J. D. (2013). Predicting Treatment Response in Social Anxiety Disorder From Functional Magnetic Resonance Imaging. *JAMA Psychiatry*, *70*(1), 87–97. https://doi.org/10.1001/2013.jamapsychiatry.5

Dowding, D., & Thompson, C. (2003). Measuring the quality of judgement and decision-making in nursing. *Journal of Advanced Nursing*, *44*(1), 49–57. https://doi.org/10.1046/j.1365-2648.2003.02770.x

Duffy, F. F., Chung, H., Trivedi, M., Rae, D. S., Regier, D. A., & Katzelnick, D. J. (2008). Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatric Services*, *59*(10), 1148–1154. https://doi.org/10.1176/appi.ps.59.10.1148

Duncan, A. E., Neuman, R. J., Kramer, J., Kuperman, S., Hesselbrock, V., Reich, T., & Bucholz, K. K. (2005). Are there subgroups of bulimia nervosa based on comorbid psychiatric disorders? *International Journal of Eating Disorders*, *37*(1), 19–25. https://doi.org/10.1002/eat.20066

Duncan, E., Best, C., & Hagen, S. (2010). Shared decision making interventions for people with mental health conditions. In *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD007297.pub2

Dunlop, B. W., Kelley, M. E., Aponte-Rivera, V., Mletzko-Crowe, T., Kinkead, B., Ritchie, J. C., … Crowe, T. M. (2017). Effects of patient preferences on outcomes in the predictors of remission in depression to individual and combined treatments (PReDICT) Study. *American Journal of Psychiatry*, *174*(6), 546–556. https://doi.org/10.1176/appi.ajp.2016.16050517

Dzau, V. J., & Ginsburg, G. S. (2016). Realizing the Full Potential of Precision Medicine in Health and Health Care. *JAMA*, *316*(16), 1659–1660. https://doi.org/10.1001/jama.2016.14117

Egger, M., Smith, G. D., & Altman, D. G. (2001). *Systematic Reviews in Health Care: Meta-Analysis in Context*. *BMJ Books* (Vol. Second). https://doi.org/10.1093/ije/31.3.697

Eiring, Ø., Landmark, B. F., Aas, E., Salkeld, G., Nylenna, M., & Nytrøen, K. (2015). What matters to patients? A systematic review of preferences for medication-associated outcomes in mental disorders. *BMJ Open*, *5*(4), e007848. https://doi.org/10.1136/bmjopen-2015-007848

Eliacin, J., Salyers, M. P., Kukla, M., & Matthias, M. S. (2015). Factors influencing patients' preferences and perceived involvement in shared decision-making in mental health care. *Journal of Mental Health*, *24*(1), 24–28. https://doi.org/10.3109/09638237.2014.954695

Elstein, A. S. (1999). Heuristics and biases: selected errors in clinical reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, *74*(7), 791–794. https://doi.org/10.1097/00001888-199907000-00012

Elstein, A. S., & Schwartz, A. (2002). Clinical problem solving and diagnostic decision

making: selective review of the cognitive literature. *BMJ*, *324*, 729–732. https://doi.org/10.1136/bmj.324.7339.729

England, E., Nash, V., & Hawthorne, K. (2017). GP training in mental health needs urgent reform. *BMJ*, *356*. Retrieved from http://www.bmj.com/content/356/bmj.j1311

Erekson, D. M., Lambert, M. J., & Eggett, D. L. (2015). The Relationship Between Session Frequency and Psychotherapy Outcome in a Naturalistic Setting. *Journal of Consulting and Clinical Psychology*, *83*(6), 1097–1107. https://doi.org/10.1037/a0039774

Eshghi, A., Haughton, D., Legrand, P., Skaletsky, M. & Woolford, S. (2011). Identifying groups: A comparison of methodologies. *Journal of Data Science*, *9*, 271–291.

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, *1*(3), 70–72.

Evans, K. C., Dougherty, D. D., Pollack, M. H., & Rauch, S. L. (2006). Using neuroimaging to predict treatment response in mood and anxiety disorders. *Annals of Clinical Psychiatry*, *18*(1), 33–42. https://doi.org/10.1080/10401230500464661

Fahey, T., & Van Der Lei, J. (2009). Producing and Using Clinical Prediction Rules. In A. Knottnerus & F. Buntinx (Eds.), *The Evidence Base of Clinical Diagnosis: Theory and methods of diagnostic research: Second Edition* (pp. 213–236). https://doi.org/10.1002/9781444300574.ch11

Farmer, P., & Dyer, J. (2016). *Five Year Forward View for Mental Health*.

Fava, M., Rush, A. J., Alpert, J. E., Balasubramani, G. K., Wisniewski, S. R., Carmin, C. N., … Trivedi, M. H. (2008). Difference in treatment outcome in outpatients with anxious versus nonanxious depression: A STAR*D report. *American Journal of Psychiatry*, *165*(3), 342–351. https://doi.org/10.1176/appi.ajp.2007.06111868

Fear, N. ., Bridges, S., Hatch, S., Hawkins, V., & Wessely, S. (2016). Chapter 4: Posttraumatic stress disorder. In S. Mcmanus, P. Bebbington, R. Jenkins, & T. S. Brugha (Eds.), *Adult psychiatric morbidity in England, 2007 Results of a household survey*. Leeds.

Fekadu, A., Wooderson, S. C., Markopoulo, K., Donaldson, C., Papadopoulos, A., & Cleare, A. J. (2009). What happens to patients with treatment-resistant depression? A systematic review of medium to long term outcome studies. *Journal of Affective Disorders*, *116*(1–2), 4–11. https://doi.org/10.1016/j.jad.2008.10.014

Ferguson, J. M. (2001). SSRI Antidepressant Medications: Adverse Effects and Tolerability. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*(1), 22–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15014625

Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-Analysis of Dropout From Cognitive Behavioral Therapy: Magnitude, Timing, and Moderators. *Journal of Consulting and Clinical Psychology*, *83*(6), 1108–1122. https://doi.org/http://dx.doi.org/10.1037/ccp0000044

Fischhoff, B., Bostrom, A., & Quadrel, M. J. (1993). Risk Perception and Communication. *Annual Review of Public Health*, *14*(1), 183–203. https://doi.org/10.1146/annurev.pu.14.050193.001151

Fisher, P. L., & Durham, R. C. (1999). Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the STAI-T across outcome studies since 1990. *Psychological Medicine*, *29*(6), 1425–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10616949

Fonagy, P. (2010). Psychotherapy research: Do we know what works for whom? *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.bp.110.079657

Forand, N. R., & DeRubeis, R. J. (2013). Pretreatment anxiety predicts patterns of change in cognitive behavioral therapy and medications for depression. *Journal of Consulting and Clinical Psychology*, *81*(5), 774–782. https://doi.org/10.1037/a0032985

Fossati, a, Maffei, C., Battaglia, M., Bagnato, M., Donati, D., Donini, M., … Novella, L. (2001). Latent class analysis of DSM-IV schizotypal personality disorder criteria in psychiatric patients. *Schizophrenia Bulletin*, *27*(1), 59–71.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*, *303*(1), 47–53. https://doi.org/10.1001/jama.2009.1943

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of Response to Medication and Cognitive Therapy in the Treatment of Moderate to Severe Depression. *Journal of Consulting and Clinical Psychology*, *77*(4), 775–787. https://doi.org/10.1037/a0015401

Friedman, J. H. (2006). Recent advances in predictive (machine) learning. *Journal of Classification*, *23*(2), 175–197. https://doi.org/10.1007/s00357-006-0012-4

Fu, C. H. Y., Williams, S. C. R., Cleare, A. J., Brammer, M. J., Walsh, N. D., Kim, J., … Bullmore, E. T. (2004). Attenuation of the neural response to sad faces in major depression by antidepressant treatment: a prospective, event-related functional magnetic resonance imaging study. *Archives of General Psychiatry*, *61*(9), 877–89. https://doi.org/10.1001/archpsyc.61.9.877

Gandhi, T. K., Kachalia, A., Thomas, E. J., Puopolo, A. L., Yoon, C., Brennan, T. A., & Studdert, D. M. (2006). Missed and delayed diagnoses in the ambulatory setting: A

study of closed malpractice claims. *Annals of Internal Medicine*, *145*(7), 488–496. https://doi.org/145/7/488 [pii]

Garb, H. N. (2005). Clinical Judgment and Decision Making. *Annual Review of Clinical Psychology*, *1*(1), 67–89. https://doi.org/10.1146/annurev.clinpsy.1.102803.143810

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., … Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *JAMA*, *293*(10), 1223–38. https://doi.org/10.1001/jama.293.10.1223

Gava, I., Barbui, C., Aguglia, E., Carlino, D., Churchill, R., De Vanna, M., & McGuire, H. F. (2007). Psychological treatments versus treatment as usual for obsessive compulsive disorder (OCD). *Cochrane Database of Systematic Reviews (Online)*, *1*(2), CD005333. https://doi.org/10.1002/14651858.CD005333.pub2

Geiser, C. (2013). *Data analysis with Mplus*. (L. Class/profile;, Ed.). New York: Guilford.

Geissbuhler, A., & Miller, R. A. (2000). Computer-Assisted Clinical Decision Support. In G. B. Chapman & F. A. Sonnenberg (Eds.), *Decision making in health care: Theory, psychology, and applications.* Cambridge: Cambridge University Press.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*(1), 1–11. https://doi.org/10.1037/cou0000131

Goodman, L. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215–231. https://doi.org/10.2307/2334349

Gorwood, P., Bayle, F., Vaiva, G., Courtet, P., Corruble, E., & Llorca, P. M. (2013). Is it worth assessing progress as early as week 2 to adapt antidepressive treatment strategy? Results from a study on agomelatine and a global meta-analysis. *European Psychiatry*, *28*(6), 362–371. https://doi.org/10.1016/j.eurpsy.2012.11.004

Grammer, G. G., Kuhle, A. R., Clark, C. C., Dretsch, M. N., Williams, K. A., & Cole, J. T. (2015). Severity of Depression Predicts Remission Rates Using Transcranial Magnetic Stimulation. *Frontiers in Psychiatry*, *6*, 114. https://doi.org/10.3389/fpsyt.2015.00114

Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W., … Kaplan, K. (2004). Prevalence and Co-occurrence of Substance Use Disorders and IndependentMood and Anxiety Disorders: Results From the National Epidemiologic Survey on Alcohol and RelatedConditions. *Archives of General Psychiatry*, *61*(8), 807–816. https://doi.org/10.1001/archpsyc.61.8.807

Graur, S., & Siegle, G. (2013). Pupillary motility: Bringing neuroscience to the psychiatry

clinic of the future. *Current Neurology and Neuroscience Reports*, *13*(8), 365. https://doi.org/10.1007/s11910-013-0365-0

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

Gueorguieva, R., Mallinckrodt, C., & Krystal, J. H. (2011). Trajectories of depression severity in clinical trials of duloxetine: insights into antidepressant and placebo responses. *Archives of General Psychiatry*, *68*(12), 1227–37. https://doi.org/10.1001/archgenpsychiatry.2011.132

Guze, S. B., & Robins, E. (1970). Suicide and primary affective disorders. *British Journal of Psychiatry*, *117*(539), 437–438. https://doi.org/10.1192/bjp.117.539.437

Gyani, A., Shafran, R., Layard, R., & Clark, D. (2011). Enhancing Recovery Rates in IAPT Services: lessons from analysis of the Year One data. *LSE Research Online*, 84. https://doi.org/10.1016/j.brat.2013.06.004

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

Habert, J., Katzman, M. A., Oluboka, O. J., McIntyre, R. S., McIntosh, D., MacQueen, G. M., … Kennedy, S. H. (2016). Functional Recovery in Major Depressive Disorder. *The Primary Care Companion For CNS Disorders*, *18*(5). https://doi.org/10.4088/PCC.15r01926

Haby, M. M., Donnelly, M., Corry, J., & Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: a meta-regression of factors that may predict outcome. *The Australian and New Zealand Journal of Psychiatry*, *40*(1), 9–19. https://doi.org/10.1111/j.1440-1614.2006.01736.x

Hagenaars, J.A., & McCutcheon, A. . (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, *61*(2), 155–163. https://doi.org/10.1002/jclp.20108

Harder, T., Remschmidt, C., Haller, S., Eckmanns, T., & Wichmann, O. (2016). Use of existing systematic reviews for evidence assessments in infectious disease prevention: a comparative case study. *Systematic Reviews*, *5*(1), 171. https://doi.org/10.1186/s13643-016-0347-9

Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, *122*(3), 184–91. https://doi.org/10.1111/j.1600-0447.2009.01519.x

Hargraves, I., & Montori, V. M. (2014). Decision aids, empowerment, and shared decision making. *BMJ*, *349*, g5811. https://doi.org/10.1136/bmj.g5811

Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist client feedback and clinical support tools. *Psychotherapy Research*, *17*(4), 379–392. https://doi.org/Doi 10.1080/10503300600702331

Hatcher, S. (2008). The STAR*D trial: the 300 lb gorilla is in the room, but does it block all the light? *Evidence-Based Mental Health*, *11*(4), 97–9. https://doi.org/10.1136/ebmh.11.4.97

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

Henderson, C. E., Rowe, C. L., Dakof, G. A., Hawes, S. W., & Liddle, H. A. (2009). Parenting Practices as Mediators of Treatment Effects in an Early-Intervention Trial of Multidimensional Family Therapy. *The American Journal of Drug and Alcohol Abuse*, *35*(4), 220–226. https://doi.org/10.1080/00952990903005890

Henzen, A., Moeglin, C., Giannakopoulos, P., & Sentissi, O. (2016). Determinants of dropout in a community-based mental health crisis centre. *BMC Psychiatry*, *16*, 111. https://doi.org/10.1186/s12888-016-0819-4

Hillier, T. A., Cauley, J. A., Rizzo, J. H., Pedula, K. L., Ensrud, K. E., Bauer, D. C., … Cummings, S. R. (2011). WHO absolute fracture risk models (FRAX): do clinical risk factors improve fracture prediction in older women without osteoporosis? *Journal of Bone and Mineral Research*, *26*(8), 1774–82. https://doi.org/10.1002/jbmr.372

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*, *335*. Retrieved from http://www.bmj.com/content/early/2006/12/31/bmj.39261.471806.55

Hirschfeld, R. M. A. (2001). The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Primary Care Companion to The Journal of Clinical Psychiatry*, *3*(6), 244–254. https://doi.org/10.4088/PCC.v03n0609

Hoffmann, T. C., Montori, V. M., & Del Mar, C. (2014). The connection between evidence-based medicine and shared decision making. *Journal of the American Medical*

Association, *312*(13), 1295–1296. https://doi.org/10.1001/jama.2014.10186

Hollon, S. D., DeRubeis, R. J., Fawcett, J., Amsterdam, J. D., Shelton, R. C., Zajecka, J., … Gallop, R. (2014). Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. *JAMA Psychiatry*, *71*(10), 1157–64. https://doi.org/10.1001/jamapsychiatry.2014.1054

HSCIC. (2014). *Psychological Therapies, England: Annual Report on the use of Improving Access to Psychological Therapies services – 2012/13*.

HSCIC. (2015). *Psychological Therapies, Annual Report on the use of IAPT services: England 2014-15*.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Hu, X. H., Bull, S. A., Hunkeler, E. M., Ming, E., Lee, J. Y., Fireman, B., & Markson, L. E. (2004). Incidence and duration of side effects and those rated as bothersome with selective serotonin reuptake inhibitor treatment for depression: Patient report versus physician estimate. *Journal of Clinical Psychiatry*, *65*(7), 959–965. https://doi.org/10.4088/JCP.v65n0712

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

Hunt, D. L., Haynes, R. B., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes - A systematic review. *JAMA: The Journal of the American Medical Association*, *280*(15), 1339. https://doi.org/10.1001/jama.280.15.1339

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

IAPT. (2011). *The IAPT data handbook*. Retrieved from http://www.iapt.nhs.uk/silo/files/iapt-data-handbook-appendicies-v2.pdf

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., … Uher, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research*, *78*, 94–102.

https://doi.org/10.1016/j.jpsychires.2016.03.016

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2006). Beyond efficacy: The STAR*D trial. *American Journal of Psychiatry*, *163*(1), 5–7. https://doi.org/10.1176/appi.ajp.163.1.5

Ioannidis, J. P. ., & Lau, J. (2000). Evidence-Based Medicine: A Quantitative Approach to Decision making. In G. . Chapman & F. A. Sonnenberg (Eds.), *Decision making in health care: theory, psychology, and applications*. Cambridge: Cambridge University Press.

Ipser, J. C., & Stein, D. J. (2012). Evidence-based pharmacotherapy of post-traumatic stress disorder (PTSD). *The International Journal of Neuropsychopharmacology*, *15*(6), 825–840. https://doi.org/10.1017/S1461145711001209

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

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jefferies-Sewell, K., Sharma, S., Gale, T. M., Hawley, C. J., Georgiou, G. J., & Laws, K. R. (2015). To admit or not to admit? The effect of framing on risk assessment decision making in psychiatrists. *Journal of Mental Health*, *24*(1), 20–23. https://doi.org/10.3109/09638237.2014.951477

Jia, P., Zhang, L., Chen, J., Zhao, P., Zhang, M., & Burns, G. (2016). The Effects of Clinical Decision Support Systems on Medication Safety: An Overview. *PLOS ONE*, *11*(12), e0167683. https://doi.org/10.1371/journal.pone.0167683

Johansson, P., & Høglend, P. (2007). Identifying mechanisms of change in psychotherapy: Mediators of treatment outcome. *Clinical Psychology and Psychotherapy*, *14*(1), 1–9. https://doi.org/10.1002/cpp.514

Johnston, B. J. (2013). The Role of Patient Experience and its Influence on Adherence to Antidepressant Treatment. *Journal of Psychosocial Nursing and Mental Health Services*, *51*(12), 29–37. https://doi.org/10.3928/02793695-20130930-04

Joosten, E. A. G., DeFuentes-Merillas, L., De Weert, G. H., Sensky, T., Van Der Staak, C. P. F., & De Jong, C. A. J. (2008). Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. *Psychotherapy*

and Psychosomatics, *77*(4), 219–226. https://doi.org/10.1159/000126073

Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kawamoto, K., Houlihan, C. a, Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal*, *5*(12), 1409–1416. https://doi.org/10.1136/bmj.38398.500764.8F

Keeffe, B., Subramanian, U., Tierney, W. M., Udris, E., Willems, J., McDonell, M., & Fihn, S. D. (2005). Provider response to computer-based care suggestions for chronic heart failure. *Medical Care*, *43*(5), 461–465. https://doi.org/00005650-200505000-00006 [pii]

Kennedy, R. D., Quinn, J. E., Mullan, P. B., Johnston, P. G., & Harkin, D. P. (2004). The Role of BRCA1 in the Cellular Response to Chemotherapy. *JNCI Journal of the National Cancer Institute*, *96*(22), 1659–1668. https://doi.org/10.1093/jnci/djh312

Keogh, B. (2015). *Personalised Medicine Strategy*. London: NHS England.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., McGonagle, K. a, Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., … Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*(1), 8–19. https://doi.org/10.1001/archpsyc.1994.03950010008002

Khoury, M. J., & Galea, S. (2016). Will Precision Medicine Improve Population Health? *JAMA*, *316*(13), 1357–1358. https://doi.org/10.1001/jama.2016.12260

Kienle, G. S., & Kiene, H. (2011). Clinical judgement and the medical profession. *Journal of Evaluation in Clinical Practice*, *17*(4), 621–7. https://doi.org/10.1111/j.1365-2753.2010.01560.x

King, M., Nazareth, I., Levy, G., Walker, C., Morris, R., Weich, S., … Torres-Gonzalez, F. (2008). Prevalence of common mental disorders in general practice attendees across Europe. *The British Journal of Psychiatry : The Journal of Mental Science*, *192*(5), 362–7. https://doi.org/10.1192/bjp.bp.107.039966

King, M., Walker, C., Levy, G., Bottomley, C., Royston, P., Weich, S., … Nazareth, I. (2008). Development and Validation of an International Risk Prediction Algorithm for Episodes of Major Depression in General Practice Attendees. *Archives of General Psychiatry*, *65*(12), 1368. https://doi.org/10.1001/archpsyc.65.12.1368

Kircanski, K., LeMoult, J., Ordaz, S., & Gotlib, I. H. (2017). Investigating the nature of co-occurring depression and anxiety: Comparing diagnostic and dimensional research approaches. *Journal of Affective Disorders*, *216*, 123–135. https://doi.org/10.1016/j.jad.2016.08.006

Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *British Journal of Psychiatry*, *195*(1), 15–22. https://doi.org/10.1192/bjp.bp.108.053967

Knopp, J., Knowles, S., Bee, P., Lovell, K., & Bower, P. (2013). A systematic review of predictors and moderators of response to psychological therapies in OCD: Do we have enough empirical evidence to target treatment? *Clinical Psychology Review*, *33*(8), 1067–1081. https://doi.org/10.1016/j.cpr.2013.08.008

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: building a safer health system*. *Annales francaises d'anesthesie et de reanimation* (Vol. 21). https://doi.org/10.1017/S095026880100509X

Konrad, J., Loos, S., Neumann, P., Zentner, N., Mayer, B., Slade, M., … Puschner, B. (2015). Content and implementation of clinical decisions in the routine care of people with severe mental illness. *Journal of Mental Health*, *24*(1), 15–19. https://doi.org/10.3109/09638237.2014.951478

Kool, S., Schoevers, R., De Maat, S., Van, R., Molenaar, P., Vink, A., & Dekker, J. (2005). Efficacy of pharmacotherapy in depressed patients with and without personality disorders: A systematic review and meta-analysis. *Journal of Affective Disorders*, *88*(3), 269–278. https://doi.org/10.1016/j.jad.2005.05.017

Kopta, S., Howard, K. I., Lowry, J. L., Beutler, L. E., Jane Manford, M., Hooper, T., … Hill, J. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*(5), 1009–1016. https://doi.org/10.1037/0022-006X.62.5.1009

Kraemer, H. C. (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: A parametric approach. *Statistics in Medicine*, *32*(11), 1964–1973. https://doi.org/10.1002/sim.5734

Krantz, A., Korn, R., & Menninger, M. (2009). Rethinking museum visitors: using k-means cluster analysis to explore a museum's audience. *Curator: The Museum Journal*, *52*(4), 363–374. https://doi.org/10.1111/j.2151-6952.2009.tb00358.x

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief

depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11556941

Kumbhani, D. J., Wells, B. J., Lincoff, a M., Jain, A., Arrigain, S., Yu, C., … Kattan, M. W. (2013). Predictive models for short- and long-term adverse outcomes following discharge in a contemporary population with acute coronary syndromes. *American Journal of Cardiovascular Disease*, *3*(1), 39–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23467552

Kyriacou, D. N., & Lewis, R. J. (2016). Confounding by indication in clinical research. *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2016.16435

Lack, C. W. (2012). Obsessive-compulsive disorder: Evidence-based treatments and future directions for research. *World Journal of Psychiatry*, *2*(6), 86–90. https://doi.org/10.5498/wjp.v2.i6.86

Lambert, M. J. (1983). Introduction to assessment of psychotherapy outcome: historial perspective and current issues. In *The Assessment of Psychotherapy Outcome* (pp. 3–32). New York: Wiley.

Lambert, M. J. (2001). Psychotherapy outcome and quality improvement: introduction to the special section on patient-focused research. *Journal of Consulting and Clinical Psychology*, *69*(2), 147–9. https://doi.org/10.1037/0022-006X.69.2.147

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy*, *50*(1), 42–51. https://doi.org/10.1037/a0030682

Lambert, M. J., & Shimokawa, K. (2011). Collecting Client Feedback. In *Psychotherapy Relationships That Work: Evidence-Based Responsiveness*. https://doi.org/10.1093/acprof:oso/9780199737208.003.0010

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The Effects of Providing Therapists With Feedback on Patient Progress During Psychotherapy: Are Outcomes Enhanced? *Psychotherapy Research*, *11*(1), 49–68. https://doi.org/10.1080/713663852

Landis, R., Edwards, B., & Corina, J. (2009). Correlated residuals among items in the estimation of measurement models. In V. R.J. & LanceC.E. (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences.* (pp. 195–214). https://doi.org/10.4324/9780203867266

Lang, J. M., Ford, J. D., & Fitzgerald, M. M. (2010). An algorithm for determining use of trauma-focused cognitive–behavioral therapy. *Psychotherapy: Theory, Research, Practice, Training*, *47*(4), 554–569. https://doi.org/10.1037/a0021184

Lazarsfield. P.F., & Henry, N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.

Lee, M.-S., Flammer, A. J., Lerman, L. O., & Lerman, A. (2012). Personalized medicine in cardiovascular diseases. *Korean Circulation Journal*, *42*(9), 583–91. https://doi.org/10.4070/kcj.2012.42.9.583

Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., … Leibing, E. (2013). Psychodynamic Therapy and Cognitive-Behavioral Therapy in Social Anxiety Disorder: A Multicenter Randomized Controlled Trial. *American Journal of Psychiatry*, *170*(7), 759–767. https://doi.org/10.1176/appi.ajp.2013.12081125

Levy, K. N., Ellison, W. D., Scott, L. N., & Bernecker, S. L. (2011). Attachment style. *Journal of Clinical Psychology*, *67*(2), 193–203. https://doi.org/10.1002/jclp.20756

Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. (2000). Clinical implications of "subthreshold" depressive symptoms. *Journal of Abnormal Psychology*, *109*(2), 345–351. https://doi.org/10.1037/0021-843X.109.2.345

Lewis, C. C., Simons, A. D., & Kim, H. K. (2012). The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. *Journal of Consulting and Clinical Psychology*, *80*(4), 525–534. https://doi.org/10.1037/a0029131

Licinio, J., & Wong, M.-L. (2011). Pharmacogenomics of antidepressant treatment effects. *Dialogues in Clinical Neuroscience*, *13*(1), 63–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21485747%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3181965

Lieblich, S. M., Castle, D. J., Pantelis, C., Hopwood, M., Young, A. H., & Everall, I. P. (2015). High heterogeneity and low reliability in the diagnosis of major depression will impair the development of new drugs. *British Journal of Psychiatry Open*, *1*(2), e5–e7. Retrieved from http://bjpo.rcpsych.org/content/1/2/e5

Lingam, R., & Scott, J. (2002). Treatment non-adherence in affective disorders. *Acta Psychiatr Scand*, *105*, 164–172. https://doi.org/10.1034/j.1600-0447.2002.1r084.x

Lo, Y., Mendell, N.R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *8*, 767–778.

Luszczynska, A., Benight, C. C., & Cieslak, R. (2009). Self-Efficacy and Health-Related Outcomes of Collective Trauma. *European Psychologist*, *14*(1), 51–62. https://doi.org/10.1027/1016-9040.14.1.51

Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. a, Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and

cognitive-behavioural therapy for depression. *The British Journal of Psychiatry : The Journal of Mental Science*, *190*, 496–502. https://doi.org/10.1192/bjp.bp.106.024729

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., … Iveson, S. (2005). Predicting change for individual psychotherapy clients based on their nearest neighbors. *Journal of Consulting and Clinical Psychology*, *73*, 904–913.

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., … Tholen, S. (2006). Empirically and clinically useful decision making in psychotherapy: differential predictions with treatment response models. *Psychological Assessment*, *18*(2), 133–141. https://doi.org/10.1037/1040-3590.18.2.133

Lutz, W., Stulz, N., Martinovich, Z., Leon, S., & Saunders, S. M. (2009). Methodological background of decision rules and feedback tools for outcomes management in psychotherapy. *Psychotherapy Research*, *19*(4–5), 502–510. https://doi.org/10.1080/10503300802688486

Lutz, W., Zimmermann, D., Müller, V. N. L. S., Deisenhofer, A.-K., & Rubel, J. A. (2017). Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment outcome in outpatient psychotherapy: study protocol. *BMC Psychiatry*, *17*(1), 306. https://doi.org/10.1186/s12888-017-1464-2

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Retrieved from http://projecteuclid.org/euclid.bsmsp/1200512992

Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, *20*(1), 37–44. https://doi.org/ISSN: 1614-1881

Malpass, A., Shaw, A., Sharp, D., Walter, F., Feder, G., Ridd, M., & Kessler, D. (2009). “Medication career” or “Moral career”? The two sides of managing antidepressants: A meta-ethnography of patients’ experience of antidepressants. *Social Science and Medicine*, *68*(1), 154–168. https://doi.org/10.1016/j.socscimed.2008.09.068

Mancini, M., Perna, G., Rossi, A., & Petralia, A. (2010). Use of duloxetine in patients with an anxiety disorder, or with comorbid anxiety and major depressive disorder: A review of the literature. *Expert Opinion on Pharmacotherapy*, *11*(7), 1167–1181. https://doi.org/http://dx.doi.org/10.1517/14656561003747441

Mansell, W., Harvey, A., Watkins, E., & Shafran, R. (2009). Conceptual Foundations of the Transdiagnostic Approach to CBT. *Journal of Cognitive Psychotherapy*, *23*(1), 6–19. https://doi.org/10.1891/0889-8391.23.1.6

Marks, I. (1998). Overcoming obstacles to routine outcome measurement: The nuts and

bolts of implementing clinical audit. *British Journal of Psychiatry*, *173*, 281–286. Retrieved from 10.1192/bjp.173.4.281%5Cnhttp://ezproxy.lib.ucf.edu/login?URL=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1998-12640-002&site=ehost-live

Mattick, R. P., Peters, L., & Clarke, J. C. (1989). Exposure and cognitive restructuring for social phobia: A controlled study. *Behavior Therapy*, *20*(1), 3–23. https://doi.org/10.1016/S0005-7894(89)80115-7

Mauskopf, J. A., Simon, G. E., Kalsekar, A., Nimsch, C., Dunayevich, E., & Cameron, A. (2009). Nonresponse, partial response, and failure to achieve remission: humanistic and cost burden in major depressive disorder. *Depression and Anxiety*, *26*(1), 83–97. https://doi.org/10.1002/da.20505

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, *1*(5), 368–376. https://doi.org/10.1016/S2215-0366(14)70329-3

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, *16*(1), 163–200. https://doi.org/10.1007/BF01065485

McCarthy, M. (2013). Patient participation in decision making may raise cost of care, study shows (News). *BMJ*, *346*. https://doi.org/http://dx.doi.org/10.1136/bmj.f3597

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). Paying the price: the cost of mental health care in England to 2026. *The British Journal of Psychiatry : The Journal of Mental Science*, *184*, 386–92. https://doi.org/10.1192/bjp.184.5.386

McGinn, T., Moore, C., & Ho, W. (2002). Practice corner: Using clinical prediction rule. *Evidence Based Medicine*, *7*, 132–134. Retrieved from http://ebm.bmj.com/content/7/5/132

McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., … Mayberg, H. S. (2013). Toward a Neuroimaging Treatment Selection Biomarker for Major Depressive Disorder. *JAMA Psychiatry*, *70*(8), 821. https://doi.org/10.1001/jamapsychiatry.2013.143

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

McManus, S., Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2009). *Adult psychiatric morbidity in England, 2007 Results of a household survey*. *The Health and*

Social Care Information Centre, Social Care Statistics. Retrieved from http://discovery.ucl.ac.uk/164862/

McPherson, S., Cairns, P., Carlyle, J., Shapiro, D. A., Richardson, P., & Taylor, D. (2005). The effectiveness of psychological treatments for treatment-resistant depression: A systematic review. *Acta Psychiatrica Scandinavica*, *111*(5), 331–340. https://doi.org/10.1111/j.1600-0447.2004.00498.x

Mehrotra, A., Reid, R. O., Adams, J. L., Friedberg, M. W., McGlynn, E. A., & Hussey, P. S. (2012). Physicians with the least experience have higher cost profiles than do physicians with the most experience. *Health Affairs*, *31*(11), 2453–2463. https://doi.org/10.1377/hlthaff.2011.0252

Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*, *4*, 47–55. https://doi.org/10.2147/RMHP.S12985

Mental Health Policy Group. (2006). *The Depression Report, A New Deal for Depression and Anxiety Disorders*. *Health Policy*. Retrieved from http://eprints.lse.ac.uk/818/

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*, *374*(9690), 609–19. https://doi.org/10.1016/S0140-6736(09)60879-5

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group. (2009). Preferred Reporting / tems for Systematic Reviews and Meta-Analysis: The PRISMA Statement. *PLoSMed*. https://doi.org/10.1371/journal.pmed1000097

Morina, N., Wicherts, J. M., Lobbrecht, J., & Priebe, S. (2014). Remission from post-traumatic stress disorder in adults: A systematic review and meta-analysis of long term outcome studies. *Clinical Psychology Review*, *34*(3), 249–255. https://doi.org/10.1016/j.cpr.2014.03.002

Moynihan, R. (2013). The future of medicine lies in truly shared decision making. *BMJ*, *346*, f2789–f2789. https://doi.org/10.1136/bmj.f2789

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J., Coryell, W., … Maser, J. D. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *American Journal of Psychiatry*, *156*(7), 1000–1006. https://doi.org/10.1176/ajp.156.7.1000

Mulder, R. T. (2002). Personality pathology and treatment outcome in major depression: A review. *American Journal of Psychiatry*, *159*(3), 359–371. https://doi.org/10.1176/appi.ajp.159.3.359

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach.

Journal of Economic Perspectives, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Mululo, S. C. C., de Menezes, G. B., Vigne, P., & Fontenelle, L. F. (2012). A review on predictors of treatment outcome in social anxiety disorder. *Official Journal of the Brazilian Psychiatric Association*, *34*(1), 92–100. https://doi.org/10.1590/S1516-44462012000100016

Munafò, M. R., Zammit, S., & Flint, J. (2014). Practitioner review: A critical perspective on gene-environment interaction models--what impact should they have on clinical perceptions and practice? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(10), 1092–101. https://doi.org/10.1111/jcpp.12261

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

Murray, C. J. L., & Lopez, A. D. (1997). Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*, *349*, 1498–1504. https://doi.org/10.1016/S0140-6736(96)07492-2

Musliner, K. L., Munk-Olsen, T., Laursen, T. M., Eaton, W. W., Zandi, P. P., & Mortensen, P. B. (2016). Heterogeneity in 10-Year Course Trajectories of Moderate to Severe Major Depressive Disorder. *JAMA Psychiatry*, *73*(4), 346. https://doi.org/10.1001/jamapsychiatry.2015.3365

Muthén, B. O. (2001). Second-generation structural equation modelling with combination of categorical and continuous latent variables: New opportunities for latent class / latent growth modelling. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington DC: American Psychological Association.

Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., … Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, *3*(4), 459–75. https://doi.org/10.1093/biostatistics/3.4.459

Muthén, L. K., & Muthén, B. O. (2008). *Growth Modeling With Latent Variables Using Mplus: Introductory and Intermediate Growth Models*. Retrieved from http://statistics.ats.ucla.edu/stat/mplus/seminars/gm/Topic3_FINAL.pdf

Muthén, L., & Muthén, B. (2012). Mplus Version 7 user's guide. *Los Angeles, CA: Muthén & Muthén*.

Najt, P., Fusar-Poli, P., & Brambilla, P. (2011). Co-occurring mental and substance abuse disorders: A review on the potential predictors and clinical outcomes. *Psychiatry Research*, *186*(2–3), 159–164. https://doi.org/10.1016/j.psychres.2010.07.042

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable

course of illness and treatment outcome in depression: A meta-analysis. *American Journal of Psychiatry*, *169*(2), 141–151. https://doi.org/10.1176/appi.ajp.2011.11020335

Naudet, F., Maria, A. S., & Falissard, B. (2011). Antidepressant response in major depressive disorder: a meta-regression comparison of randomized controlled trials and observational studies. *PLoS One*, *6*, e20811–e20811. https://doi.org/10.1371/journal.pone.0020811

NCCC. (2009). *Advanced Breast Cancer*. *Advanced Breast Cancer: Diagnosis and Treatment*. National Collaborating Centre for Cancer (UK). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21901868

Nelson, J. C., Delucchi, K. L., & Schneider, L. S. (2013). Moderators of outcome in late-life depression: A patient-level meta-analysis. *American Journal of Psychiatry*, *170*(6), 651–659. https://doi.org/10.1176/appi.ajp.2012.12070927

Newton-Howes, G., Tyrer, P., & Johnson, T. (2006). Personality disorder and the outcome of depression: meta-analysis of published studies. *The British Journal of Psychiatry*, *188*, 13–20. https://doi.org/10.1192/bjp.188.1.13

NHS Digital. (2016). *Psychological Therapies, Annual Report on the use of IAPT services: England 2015-16*.

NICE. (2008). Stroke and transient ischaemic attack in over 16s: diagnosis and initial management. *Nice Guideline*, (July). Retrieved from http://www.nice.org.uk/guidance/CG68

NICE. (2009). *Depression in Adults: Recognition and Management*. *NICE Clinical Guideline*. Retrieved from guidance.nice.org.uk/cg90

NICE. (2011a). *Commissioning stepped care for people with common mental health disorders*. London: National Institute for Health and Clinical Excellence.

NICE. (2011b). *Common mental health disorders: The NICE guideline on identification and pathways to care*. *The British Psychological Society and The Royal College of Psychiatrists*. https://doi.org/clinical guideline CG123.2011

NICE. (2011c). *Common mental health problems: identification and pathways to care*. NICE. Retrieved from https://www.nice.org.uk/guidance/cg123/evidence

NICE. (2013). *Social Anxiety Disorder: Recognition, Assessment and Treatment*. *NICE Clinical Guideline 159*. Retrieved from http://www.nice.org.uk/guidance/cg159/resources/guidance-social-anxiety-disorder-recognition-assessment-and-treatment-pdf

Nunes, E. V, & Levin, F. R. (2004). Treatment of depression in patients with alcohol or other

drug dependence: a meta-analysis. *JAMA*, *291*(15), 1887–1896. https://doi.org/10.1001/jama.291.15.1887

Oestergaard, S., & Møldrup, C. (2009). Application of pharmacogenomics to clinical problems in depression. *Personalized Medicine*, *6*(5), 501–515. https://doi.org/10.2217/pme.09.32

Oldham, M., Kellett, S., Miles, E., & Sheeran, P. (2012). Interventions to increase attendance at psychotherapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *80*(5), 928–939. https://doi.org/10.1037/a0029630

Ormerod, S., McDowell, S. E., Coleman, J. J., & Ferner, R. E. (2008). Ethnic differences in the risks of adverse reactions to drugs used in the treatment of psychoses and depression: A systematic review and meta-analysis. *Drug Safety*, *31*(7), 597–607. https://doi.org/10.2165/00002018-200831070-00005

Ozomaro, U., Wahlestedt, C., Nemeroff, C. B., Pulford, J., Miles, W., Sheridan, J., … Stefansson, K. (2013). Personalized medicine in psychiatry: problems and promises. *BMC Medicine*, *11*(1), 132. https://doi.org/10.1186/1741-7015-11-132

Paige, L., & Mansell, W. (2013). To attend or not attend? A critical review of the factors impacting on initial appointment attendance from an approach–avoidance perspective. *Journal of Mental Health*, *22*(1), 72–82. https://doi.org/10.3109/09638237.2012.705924

Paksarian, D., Cui, L., Angst, J., Ajdacic-Gross, V., Rössler, W., & Merikangas, K. R. (2016). Latent Trajectories of Common Mental Health Disorder Risk Across 3 Decades of Adulthood in a Population-Based Cohort. *JAMA Psychiatry*, *73*(10), 1023. https://doi.org/10.1001/jamapsychiatry.2016.1921

Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B., & Munizza, C. (2004). Combined pharmacotherapy and psychological treatment for depression: a systematic review. *Archives of General Psychiatry*, *61*(7), 714–719. https://doi.org/10.1001/archpsyc.61.7.714

Papakostas, G. I., & Fava, M. (2008). Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. *Dialogues in Clinical Neuroscience*, *10*(4), 439–451.

Papakostas, G. I., & Fava, M. (2010). *Pharmacotherapy for depression and treatment-resistant depression*. Hackensack, NJ: Word Scientific.

Papakostas, G. I., Petersen, T., Mischoulon, D., Hughes, M. E., Spector, A. R., Alpert, J. E., … Nierenberg, A. A. (2003). Functioning and interpersonal relationships as predictors of response in treatment-resistant depression. *Comprehensive Psychiatry*, *44*(1), 44–50. https://doi.org/10.1053/comp.2003.50012

Parker, G., Wilhelm, K., Mitchell, P., & Gladstone, G. (2000). Predictors of 1-year outcome in depression. *Australian and New Zealand Journal of Psychiatry*, *34*(1), 56–64. https://doi.org/10.1046/j.1440-1614.2000.00698.x

Pencina, M. J., & Peterson, E. D. (2016). Moving From Clinical Trials to Precision Medicine: The Role for Predictive Modeling. *JAMA*, *315*(16), 1713–4. https://doi.org/10.1001/jama.2016.4839

Perna, G., & Nemeroff, C. B. (2017). Personalized Medicine in Psychiatry: Back to the Future. *Personalized Medicine in Psychiatry*, *1–2*, 1. https://doi.org/10.1016/j.pmip.2017.01.001

Peterson, R.A. & Reiss, S. (1992). *Anxiety Sensitivity Index manual* (2nd ed.). San Antonio: Psychological Corporation.

Pies, R. (2007). How “objective” are psychiatric diagnoses?: (guess again). *Psychiatry (Edgmont)*, *4*(10), 18–22.

Pirmohamed, M. (2014). Personalized Pharmacogenomics: Predicting Efficacy and Adverse Drug Reactions. *Annual Review of Genomics and Human Genetics*, *15*(1), 349–370. https://doi.org/10.1146/annurev-genom-090413-025419

Pompili, M., Serafini, G., Del Casale, A., Rigucci, S., Innamorati, M., Girardi, P., … Lester, D. (2009). Improving adherence in mood disorders: the struggle against relapse, recurrence and suicide risk. *Expert Review of Neurotherapeutics*, *9*(7), 985–1004. https://doi.org/10.1586/ern.09.62

Pompili, M., Venturini, P., Palermo, M., Stefani, H., Seretti, M. E., Lamis, D. A., … Girardi, P. (2013). Mood disorders medications: Predictors of nonadherence—Review of the current literature. *Expert Review of Neurotherapeutics*, *13*(7), 809–825. https://doi.org/http://dx.doi.org/10.1586/14737175.2013.811976

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Britten, N. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: Final report*. *ESRC Methods Programme*. https://doi.org/10.13140/2.1.1018.4643

Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., & Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review*, *30*(6), 635–641. https://doi.org/10.1016/j.cpr.2010.04.007

Proudfoot, J. (2013). The future is in our hands: The role of mobile phones in the prevention and management of mental disorders. *Australian and New Zealand Journal of Psychiatry*, *47*(2), 111–113. https://doi.org/10.1007/s10488-012-0424-x.\nhttp://dx.doi.org/10.1177/0004867412471441

Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., &

Lafortune, L. (2013). Cost of Improving Access to Psychological Therapies (IAPT) programme: An analysis of cost of session, treatment and recovery in selected Primary Care Trusts in the East of England region. *Behaviour Research and Therapy*, *51*(1), 37–45. https://doi.org/10.1016/j.brat.2012.10.001

Rait, G., Walters, K., Griffin, M., Buszewicz, M., Petersen, I., & Nazareth, I. (2009). Recent trends in the incidence of recorded depression in primary care. *The British Journal of Psychiatry : The Journal of Mental Science*, *195*(6), 520–4. https://doi.org/10.1192/bjp.bp.108.058636

Ramchandani, P., & Stein, A. (2003). The impact of parental psychiatric disorder on children: avoiding stimga, improving care. *British Medical Journal*, *327*, 242–243. https://doi.org/2003;327:242–3

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*(8), 741–756. https://doi.org/10.1016/S0005-7967(97)00022-3

Reneses, B., Muñoz, E., & López-Ibor, J. J. (2009). Factors predicting drop-out in community mental health centres. *World Psychiatry*, *8*(3), 173–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19812755

Richards, D., & Suckling, R. (2009). Improving access to psychological therapies: Phase IV prospective cohort study. *British Journal of Clinical Psychology*, *48*(4), 377–396. https://doi.org/10.1348/014466509X405178

Richards, T., Coulter, A., & Wicks, P. (2015). Time to deliver patient centred care. *BMJ*, *350*, h530–h530. https://doi.org/10.1136/bmj.h530

Richardson, W. S. (2007). We should overcome the barriers to evidence-based clinical diagnosis! *Journal of Clinical Epidemiology*, *60*(3), 217–227.

Rivero-Santana, A., Perestelo-Perez, L., Perez-Ramos, J., Serrano-Aguilar, P., & De las Cuevas, C. (2013). Sociodemographic and clinical predictors of compliance with antidepressants for depressive disorders: Systematic review of observational studies. *Patient Preference and Adherence*, *7*, 151–169. https://doi.org/10.2147/PPA.S39382

Roiser, J. P., Elliott, R., & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology*, *37*(1), 117–36. https://doi.org/10.1038/npp.2011.183

Rosellini, A. J., & Brown, T. A. (2014). Initial interpretation and evaluation of a profile-based classification system for the anxiety and mood disorders: Incremental validity compared to DSM-IV categories. *Psychological Assessment*, *26*(4), 1212–1224. https://doi.org/10.1037/pas0000023

Roshanaei-Moghaddam, B., Pauly, M. C., Atkins, D. C., Baldwin, S. A., Stein, M. B., & Roy-Byrne, P. (2011). Relative effects of CBT and pharmacotherapy in depression versus anxiety: Is medication somewhat better for depression, and CBT somewhat better for anxiety? *Depression and Anxiety*, *28*(7), 560–567. https://doi.org/10.1002/da.20829

Rossi, D., Rasi, S., Di Rocco, A., Fabbri, A., Forconi, F., Gloghini, A., … Gaidano, G. (2011). The host genetic background of DNA repair mechanisms is an independent predictor of survival in diffuse large B-cell lymphoma. *Blood*, *117*(8), 2405–2413. https://doi.org/10.1182/blood-2010-07-296244

Rozental, A., Magnusson, K., Boettcher, J., Andersson, G., & Carlbring, P. (2017). For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy. *Journal of Consulting and Clinical Psychology*, *85*(2), 160–177. https://doi.org/10.1037/ccp0000158

Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, *27*(2), 478–488. https://doi.org/10.1037/pas0000060

Rush, A. J., Crismon, M. L., Kashner, T. M., Toprac, M. G., Carmody, T. J., Trivedi, M. H., … Altshuler, K. Z. (2003). Texas Medication Algorithm Project, phase 3 (TMAP-3): Rationale and study design. *Journal of Clinical Psychiatry*, *64*(4), 357–369. https://doi.org/10.4088/JCP.v64n0403

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., … 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., Stewart, J. W., Nierenberg, A. A., Fava, M., Kurian, B. T., … Wisniewski, S. R. (2011). Combining Medications to Enhance Depression Outcomes (CO-MED): Acute and long-term outcomes of a single-blind randomized study. *American Journal of Psychiatry*, *168*(7), 689–701. https://doi.org/10.1176/appi.ajp.2011.10111645

Sackett, D. L., & Rosenberg, W. M. C. (1995). On the need for evidence-based medicine. *Health Economics*, *4*(4), 249–254. https://doi.org/10.1002/hec.4730040401

San, O.M., Huynh, V. Nakamori, Y. (2004). An alternative extension of the k-means algorithm for clustering categorical data. *International Journal of Applied Mathematics and Computer Science*, *14*(2), 241–247.

Sánchez-Meca, J., Rosa-Alcázar, A. I., Marín-Martínez, F., & Gómez-Conesa, A. (2010).

Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review*, *30*(1), 37–50. https://doi.org/10.1016/j.cpr.2009.08.011

Sanderson, W. C., Wetzler, S., Beck, a T., & Betz, F. (1992). Prevalence of personality disorders in patients with major depression and dysthymia. *Psychiatry Research*, *42*(1), 93–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9718237

Sansone, R. A., & Sansone, L. A. (2012). Antidepressant adherence: Are patients taking their medications? *Innovations in Clinical Neuroscience*, *9*(5–6), 41–46.

Saverno, K. R., Hines, L. E., Warholak, T. L., Grizzle, A. J., Babits, L., Clark, C., … Malone, D. C. (2011). Ability of pharmacy clinical decision-support software to alert users about clinically important drug—drug interactions. *Journal of the American Medical Informatics Association*, *18*(1), 32–37. https://doi.org/10.1136/jamia.2010.007609

Saxon, D., Barkham, M., Foster, A., & Parry, G. (2017). The Contribution of Therapist Effects to Patient Dropout and Deterioration in the Psychological Therapies. *Clinical Psychology and Psychotherapy*, *24*(3), 575–588. https://doi.org/10.1002/cpp.2028

Schalkwijk, S., Undurraga, J., Tondo, L., & Baldessarini, R. J. (2014). Declining efficacy in controlled trials of antidepressants: effects of placebo dropout. *The International Journal of Neuropsychopharmacology*, *17*(8), 1343–52. https://doi.org/10.1017/S1461145714000224

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74. https://doi.org/10.1002/0470010940

Schoemaker, P. J. H. (1982). The Expected Utility Model: Its Variants , Purposes , Evidence and Limitations. *Journal of Economic Literature*, *20*(2), 529–563. https://doi.org/10.2307/2724488

Schreiber, J. B., & Pekarik, A. J. (2014). Technical Note: Using Latent Class Analysis versus K-means or Hierarchical Clustering to Understand Museum Visitors. *Curator: The Museum Journal*, *57*(1), 45–59. https://doi.org/10.1111/cura.12050

Schwaederle, M., Zhao, M., Lee, J. J., Lazar, V., Leyland-Jones, B., Schilsky, R. L., … Kurzrock, R. (2016). Association of Biomarker-Based Treatment Strategies With Response Rates and Progression-Free Survival in Refractory Malignant Neoplasms. *JAMA Oncology*, *2*(11), 1452. https://doi.org/10.1001/jamaoncol.2016.2129

Schwartz, A., & Elstein, A. S. (2009). Clinical Problem Solving and Diagnostic Decision Making: A Selective Review of the Cognitive Research Literature. In A. Knottnerus & F. Buntinx (Eds.), *The Evidence Base of Clinical Diagnosis: Theory and methods of*

diagnostic research: Second Edition (pp. 237–255). https://doi.org/10.1002/9781444300574.ch12

Serretti, A., Chiesa, A., Calati, R., Perna, G., Bellodi, L., & De Ronchi, D. (2009). Common genetic, clinical, demographic and psychosocial predictors of response to pharmacotherapy in mood and anxiety disorders. *International Clinical Psychopharmacology*, *24*(1), 1–18. https://doi.org/10.1097/YIC.0b013e32831db2d7

Serretti, A., Gibiino, S., & Drago, A. (2011). Specificity profile of paroxetine in major depressive disorder: Meta-regression of double-blind, randomized clinical trials. *Journal of Affective Disorders*, *132*(1–2), 14–25. https://doi.org/10.1016/j.jad.2010.08.018

Shea, M. T., Pilkonis, P. A., Beckham, E., Collins, J. F., Elkin, I., Sotsky, S. M., & Docherty, J. P. (1990). Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, *147*(6), 711–718. https://doi.org/10.1176/ajp.147.6.711

Sheehan, J., & Sherman, K. A. (2012). Computerised decision aids: A systematic review of their effectiveness in facilitating high-quality decision-making in various health-related contexts. *Patient Education and Counseling*, *88*(1), 69–86. https://doi.org/10.1016/j.pec.2011.11.006

Shelton, R. C., & Brown, L. L. (2001). Mechanisms of action in the treatment of anxiety. *Journal of Clinical Psychiatry*, *62 Suppl 1*, 10–15. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11430613

Sherer, M., & E.Maddux, J. (1982). The self-efficacy scale-construction and validation. *Psychological Reports*, *51*, 663–671. https://doi.org/10.2466/pr0.1982.51.2.663

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, *78*(3), 298–311. https://doi.org/10.1037/a0019247

Siegle, G. J., Carter, C. S., & Thase, M. E. (2006). Use of fMRI to predict recovery from unipolar depression with cognitive behavior therapy. *American Journal of Psychiatry*, *163*(4), 735–738. https://doi.org/10.1176/appi.ajp.163.4.735

Silveira, H., Moraes, H., Oliveira, N., Coutinho, E. S. F., Laks, J., & Deslandes, A. (2013). Physical exercise and clinically depressed patients: A systematic review and meta-analysis. *Neuropsychobiology*, *67*(2), 61–68. https://doi.org/10.1159/000345160

Simon, G. E., & Perlis, R. H. (2010). Personalized medicine for depression: Can we match patients with treatments? *American Journal of Psychiatry*, *167*(12), 1445–1455.

https://doi.org/10.1176/appi.ajp.2010.09111680

Simon, G. E., Rutter, C. M., Stewart, C., Pabiniak, C., & Wehnes, L. (2012). Response to past depression treatments is not accurately recalled: Comparison of structured recall and patient health questionnaire scores in medical records. *Journal of Clinical Psychiatry*, *73*(12), 1503–1508. https://doi.org/10.4088/JCP.12m07883

Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy*, *15*(5), 287–303. https://doi.org/10.1002/cpp.594

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the effects heuristic for behavioral economics. *Journal of Socio-Economics*, *31*(4), 329–342. https://doi.org/10.1016/S1053-5357(02)00174-9

Snyder, D. K., Castellani, A. M., & Whisman, M. A. (2006). Current Status and Future Directions in Couple Therapy. *Annual Review of Psychology*, *57*(1), 317–344. https://doi.org/10.1146/annurev.psych.56.091103.070154

Solomon, D. A., Leon, A. C., Coryell, W., Mueller, T. I., Posternak, M., Endicott, J., & Keller, M. B. (2008). Predicting recovery from episodes of major depression. *Journal of Affective Disorders*, *107*(1–3), 285–291. https://doi.org/10.1016/j.jad.2007.09.001

Song, M., Lee, K.-M., & Kang, D. (2011). Breast cancer prevention based on gene-environment interaction. *Molecular Carcinogenesis*, *50*(4), 280–290. https://doi.org/10.1002/mc.20639

Sousa, T. V, Viveiros, V., Chai, M. V, Vicente, F. L., Jesus, G., Carnot, M. J., … Ferreira, P. L. (2015). Reliability and validity of the Portuguese version of the Generalized Anxiety Disorder (GAD-7) scale. *Health and Quality of Life Outcomes*, *13*, 50. https://doi.org/10.1186/s12955-015-0244-2

Spatz, E. S., Krumholz, H. M., & Moulton, B. W. (2016). The new era of informed consent: Getting to a reasonable-patient standard through shared decision making. *JAMA*, *315*(19), 2063–2064. https://doi.org/10.1001/jama.2016.3070

Spence, D. (2013). Are antidepressants overprescribed? Yes. *BMJ*, *346*, f191–f191. https://doi.org/10.1136/bmj.f191

Spiers, N., Qassem, T., Bebbington, P., McManus, S., King, M., Jenkins, R., … Brugha, T. S. (2016). Prevalence and treatment of common mental disorders in the English national population, 1993-2007. *The British Journal of Psychiatry*, *209*(2), 150–156. https://doi.org/10.1192/bjp.bp.115.174979

Spinhoven, P., Batelaan, N., Rhebergen, D., van Balkom, A., Schoevers, R., & Penninx, B.

W. (2016). Prediction of 6-yr symptom course trajectories of anxiety disorders by diagnostic, clinical and psychological variables. *Journal of Anxiety Disorders*, *44*, 92–101. https://doi.org/10.1016/j.janxdis.2016.10.011

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

Stansfield, S., Clark, C., Bebbington, P., King, M., Jenkins, R., & Hinchliffe, S. (2016). Chapter 2: Common mental disorders. In S. McManus, P. Bebbington, R. Jenkins, & T. S. Brugha (Eds.), *Adult psychiatric morbidity in England, 2007 Results of a household survey*. Leeds: NHS Digital.

StataCorp LP. (2011). Stata Statistical Software: Release 12. *2011*. https://doi.org/10.2307/2234838

Steger, M. F., & Kashdan, T. B. (2009). Depression and Everyday Social Activity, Belonging, and Well-Being. *Journal of Counseling Psychology*, *56*, 289-300--. https://doi.org/10.1037/a0015416

Stein, D. J., Koen, N., Fineberg, N., Fontenelle, L. F., Matsunaga, H., Osser, D., & Simpson, H. B. (2012). A 2012 evidence-based algorithm for the pharmacotherapy for obsessive-compulsive disorder. *Current Psychiatry Reports*, *14*(3), 211–219. https://doi.org/10.1007/s11920-012-0268-9

Strunk, D. R., Cooper, A. A., Ryan, E. T., DeRubeis, R. J., & Hollon, S. D. (2012). The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intersession symptom gains. *Journal of Consulting and Clinical Psychology*, *80*(5), 730–738. https://doi.org/10.1037/a0029281

Stulz, N., Lutz, W., Kopta, S. M., Minami, T., & Saunders, S. M. (2013). Dose–effect relationship in routine outpatient psychotherapy: Does treatment duration matter? *Journal of Counseling Psychology*, *60*(4), 593–600. https://doi.org/10.1037/a0033589

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, *75*(6), 864–874. https://doi.org/10.1037/0022-006X.75.6.864

Sullivan, L. E., Fiellin, D. A., & O'Connor, P. G. (2005). The prevalence and impact of alcohol problems in major depression: A systematic review. *The American Journal of Medicine*, *118*(4), 330–341. https://doi.org/10.1016/j.amjmed.2005.01.007

Surguladze, S., Brammer, M. J., Keedwell, P., Giampietro, V., Young, A. W., Travis, M. J., … Phillips, M. L. (2005). A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder. *Biological Psychiatry*, *57*(3), 201–209. https://doi.org/10.1016/j.biopsych.2004.10.028

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(4), 547–559. https://doi.org/10.1037/a0028226

Szegedi, A., Jansen, W. T., Van Willigenburg, A. P. P., Van Der Meulen, E., Stassen, H. H., & Thase, M. E. (2009). Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: A meta-analysis including 6562 patients. *Journal of Clinical Psychiatry*, *70*(3), 344–353. https://doi.org/10.4088/JCP.07m03780

Tedeschini, E., Levkovitz, Y., Iovieno, N., Ameral, V. E., Nelson, J. C., & Papakostas, G. I. (2011). Efficacy of antidepressants for late-life depression: a meta-analysis and meta-regression of placebo-controlled randomized trials. *The Journal of Clinical Psychiatry*, *72*(12), 1660–8. https://doi.org/10.4088/JCP.10r06531

Thiel, N., Hertenstein, E., Nissen, C., Herbst, N., Külz, A. K., & Voderholzer, U. (2013). The effect of personality disorders on treatment outcomes in patients with obsessive-compulsive disorders. *Journal of Personality Disorders*, *27*(6), 697–715. https://doi.org/10.1521/pedi_2013_27_104

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., … 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

Trivedi, R. B., Nieuwsma, J. A., & Williams, J. W. (2011). Examination of the utility of psychotherapy for patients with treatment resistant depression: A systematic review. *Journal of General Internal Medicine*, *26*(6), 643–650. https://doi.org/10.1007/s11606-010-1608-2

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Uher, R., Farmer, A., Henigsberg, N., Rietschel, M., Mors, O., Maier, W., … Aitchison, K. J. (2009). Adverse reactions to antidepressants. *British Journal of Psychiatry*, *195*(3), 202–210. https://doi.org/10.1192/bjp.bp.108.061960

Uher, R., Huezo-Diaz, P., Perroud, N., Smith, R., Rietschel, M., Mors, O., … Craig, I. (2009). Genetic predictors of response to antidepressants in the GENDEP project. *The Pharmacogenomics Journal*, *9*(4), 225–233. https://doi.org/10.1038/tpj.2009.12

Uher, R., Tansey, K. E., Malki, K., & Perlis, R. H. (2012). Biomarkers predicting treatment outcome in depression: what is clinically significant? *Pharmacogenomics*, *13*(2), 233–40. https://doi.org/10.2217/pgs.11.161

Unick, G. J., Snowden, L., & Hastings, J. (2009). Heterogeneity in comorbidity between major depressive disorder and generalized anxiety disorder and its clinical consequences. *The Journal of Nervous and Mental Disease*, *197*(4), 215–224. https://doi.org/10.1097/NMD.0b013e31819d954f

Van, H. L., Schoevers, R. A., & Dekker, J. (2008). Predicting the outcome of antidepressants and psychotherapy for depression: a qualitative, systematic review. *Harvard Review of Psychiatry*, *16*(4), 225–234. https://doi.org/10.1080/10673220802277938

van Hees, M. L. J. M., Rotter, T., Ellermann, T., & Evers, S. M. A. A. (2013). The effectiveness of individual interpersonal psychotherapy as a treatment for major depressive disorder in adult outpatients: a systematic review. *BMC Psychiatry*, *13*, 22. https://doi.org/10.1186/1471-244X-13-22

Van Schaik, D. J. F., Klijn, A. F. J., Van Hout, H. P. J., Van Marwijk, H. W. J., Beekman, A. T. F., De Haan, M., & Van Dyck, R. (2004). Patients' preferences in the treatment of depressive disorder in primary care. *General Hospital Psychiatry*, *26*(3), 184–189. https://doi.org/10.1016/j.genhosppsych.2003.12.001

Veale, D. (2008). Behavioural activation for depression. *Advances in Psychiatric Treatment*, *14*(1), 29–36. https://doi.org/10.1192/apt.bp.107.004051

Verma, M., & Mukesh. (2012). Personalized Medicine and Cancer. *Journal of Personalized Medicine*, *2*(4), 1–14. https://doi.org/10.3390/jpm2010001

Vittengl, J. R., Jarrett, R. B., Weitz, E., Hollon, S. D., Twisk, J., Cristea, I., … Cuijpers, P. (2016). Divergent Outcomes in Cognitive-Behavioral Therapy and Pharmacotherapy for Adult Depression. *American Journal of Psychiatry*, *173*(5), 481–490. https://doi.org/10.1176/appi.ajp.2015.15040492

Wade, A. G., & Häring, J. (2010). A review of the costs associated with depression and treatment noncompliance: the potential benefits of online support. *International Clinical Psychopharmacology*, *25*(5), 288–296. https://doi.org/10.1097/YIC.0b013e328339fbcf

Wade, T. D., Crosby, R. D., & Martin, N. G. (2006). Use of latent profile analysis to identify eating disorder phenotypes in an adult Australian twin cohort. *Archives of General Psychiatry*, *63*(12), 1377–1384. https://doi.org/10.1001/archpsyc.63.12.1377

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An Investigation of Self-Assessment Bias in Mental Health Providers. *Psychological Reports*, *110*(2), 639–644. https://doi.org/10.2466/02.07.17.PR0.110.2.639-644

Wallace, M. L., Frank, E., & Kraemer, H. C. (2013). A Novel Approach for Developing and Interpreting Treatment Moderator Profiles in Randomized Clinical Trials. *JAMA Psychiatry*, *70*(11), 1241–1247. https://doi.org/10.1001/jamapsychiatry.2013.1960

Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, *77*, 129–137. https://doi.org/10.1016/j.brat.2015.12.005

Walsh, T., Barr, P. J., Thompson, R., Ozanne, E., O'Neill, C., & Elwyn, G. (2014). Undetermined impact of patient decision support interventions on healthcare costs and savings: systematic review. *British M*, *348*, g188. https://doi.org/10.1136/bmj.g188

Wang, J. L., Manuel, D., Williams, J., Schmitz, N., Gilmour, H., Patten, S., … Birney, A. (2013). Development and validation of prediction algorithms for major depressive episode in the general population. *Journal of Affective Disorders*, *151*(1), 39–45. https://doi.org/10.1016/j.jad.2013.05.045

Wang, M., Wright, J., Buswell, R., & Brownlee, A. (2013). A comparison of approaches to stepwise regression for global sensitivity analysis used with evolutionary optimization. In *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*.

Warden, D., Rush, a J., Carmody, T. J., Kashner, T. M., Biggs, M. M., Crismon, M. L., & Trivedi, M. H. (2009). Predictors of attrition during one year of depression treatment: a roadmap to personalized intervention. *Journal of Psychiatric Practice*, *15*(2), 113–24. https://doi.org/10.1097/01.pra.0000348364.88676.83

Watzke, B., Rüddel, H., Jürgensen, R., Koch, U., Kriston, L., Grothgar, B., & Schulz, H. (2010). Effectiveness of systematic treatment selection for psychodynamic and cognitive-behavioural therapy: Randomised controlled trial in routine mental healthcare. *British Journal of Psychiatry*, *197*(2), 96–105. https://doi.org/10.1192/bjp.bp.109.072835

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*(2), 200–211. https://doi.org/10.1037/a0018912

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

Wells, J. E., Browne, M. O., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Angermeyer, M. C., … Kessler, R. C. (2013). Drop out from out-patient mental healthcare in the World Health Organization's World Mental Health Survey initiative. *The British Journal of Psychiatry*, *202*(1), 42–49. https://doi.org/http://dx.doi.org/10.1192/bjp.bp.112.113134

WHO. (1992). *International Statistical Classification of Diseases and Related Health Problems. 10th Revision*. *Occupational Health* (Vol. 41).

https://doi.org/http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf

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*. https://doi.org/10.1038/tp.2016.61

Williams, R., Farquharson, L., Palmer, L., Bassett, P., Clarke, J., Clark, D. M., & Crawford, M. J. (2016). Patient preference in psychological treatment and associations with self-reported outcome: national cross-sectional survey in England and Wales. *BMC Psychiatry*, *16*(1), 4. https://doi.org/10.1186/s12888-015-0702-8

Wills, C. E., & Holmes-Rovner, M. (2006). Integrating Decision Making and Mental Health Interventions Research: Research Directions. *Clinical Psychology : A Publication of the Division of Clinical Psychology of the American Psychological Association*, *13*(1), 9–25. https://doi.org/10.1111/j.1468-2850.2006.00002.x

Wilson, B. J., & Nicholls, S. G. (2015). The human genome project, and recent advances in personalized genomics. *Risk Management and Healthcare Policy*, *8*, 9–20. https://doi.org/10.2147/RMHP.S58728

Wittchen, H. U. (2002). Generalized anxiety disorder: Prevalence, burden, and cost to society. *Depression and Anxiety*, *16*(4), 162–171. https://doi.org/10.1002/da.10065

Yamagishi, K. (1997). When a 12.86 % Mortality is More Dangerous than 24.14 %: Implications for Risk Communication. *Applied Cognitive Psychology*, *11*, 495–506.

Yee, M., Javitt, C., & Miller, A. (2015). Replacing *DSM* categorical analyses with dimensional analyses in psychiatry research: The Research Domain Criteria initiative. *JAMA Psychiatry*, *72*(12), 1159.

Yonkers, K. A., & Brawman-Mintzer, O. (2002). The pharmacologic treatment of depression: is gender a critical factor? *The Journal of Clinical Psychiatry*, *63*(7), 610–5. https://doi.org/10.4088/JCP.v63n0714

Zimmerman, M., Mattia, J. I., & Posternak, M. A. (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry*, *159*(3), 469–473. https://doi.org/10.1176/appi.ajp.159.3.469

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, *162*(10), 1911–1918. https://doi.org/10.1176/appi.ajp.162.10.1911

Zimmerman, M., Walsh, E., Chelminski, I., & Dalrymple, K. (2017). Has the symptom

severity inclusion requirement narrowed the definition of major depressive disorder in antidepressant efficacy trials? *Journal of Affective Disorders*, *211*, 60–64. https://doi.org/10.1016/j.jad.2017.01.008

Appendices

Appendix A: Search strings for the systematic review of predictors to treatment outcome.

Databases: EMBASE, MEDLINE, PsycINFO

#	searches
1	anxiety/ use emez or body dysmorphic disorders/ use emez or compulsive behaviour/ use emez or obsessive behaviour/ use emez or panic/ use emez or stress, psychological/ use emez or exp anxiety disorders/ use emez or exp depression/ use emez or exp depressive disorder/ use emez
2	anxiety/ use mesz or exp depression/ use mesz or body dysmorphic disorder/ use mesz or critical incident stress/ use mesz or psychotrauma/ use mesz or exp anxiety disorder/ use mesz
3	anxiety/ use psyh or body dysmorphic disorder/ use psyh or compulsions/ use psyh or emotional trauma/ use psyh or exp anxiety disorders/ use psyh or fear/ use psyh or obsessions/ use psyh or panic/ use psyh or panic attack/ use psyh or psychological stress/ use psyh or traumatic neurosis/ use psyh
4	(agoraphobi$ or anxiet$ or anxious$ or body dysmorphi$ or claustrophobi$ or clean response$ or compulsi$ or depress$ or dysmorphophobi$ or obsession$ or obsessive$ or ocd or panic$ or phobi$ or posttraum$ or post trauma$ or ptsd or recur$ thought$ or seasonal affective disorder$ or ((acute or chronic$ or extreme or incessant$ or intense$ or persistent$ or serious$ or sever$) adj2 (apprehens$ or doom or fear$ or terror$)) or railway spine or (rape adj2 trauma$) or reexperienc$ or re experienc$ or torture syndrome or (trauma$ adj (neuros$ or stress$)) or (trauma$ and (avoidance or grief or horror or death$ or nightmare$ or night mare$ or emotion$)) or (combat adj (disorder$ or fatigue or neuros$ or syndrome$)) or war neuros$ or concentration camp syndrome or ((extreme or trauma$) adj stress) or flash back$ or flashback$ or hypervigil$ or psych$ stress or psych$ trauma$ or psychotrauma$ or acute stress or asd or desnos or post traumatic$ or posttraumatic$ or ptsd or stress disorder$).ti,ab.
5	or/1-4
6	"prediction and forecasting"/ use emez or disease course/ use emez or prediction/ use emez or predictor variable/ use emez or prognosis/ use emez or decision making/ use emez or decision support system/ use emez or algorithm/ use emez
7	"predictive value of tests"/ use mesz or exp prognosis/ use mesz or decision making/ use mesz or decision support systems, clincal/ use mesz or algorithms/ use mesz
8	disease course/ use psyh or prediction/ use psyh or prognosis/ use psyh or decision making/ use psyh or decision support systems/ use psyh or algorithm/ use psyh
9	(predict$ or prognos$ or decision making).ti,ab.
10	(expected treatment response or identification of treatment failure or differential prediction$ or (patient$ adj2 deteriorat$)). ti, ab, kw.
11	or/6-10
12	risk/ use emez or recurrence risk/ use emez or risk assessment/ use emez or risk factor$.sh. use emez or risk reduction/ use emez
13	risk/ use mesz or risk assessment/ use mesz or risk factor$.sh. use mesz
14	at risk populations/ use psyh or risk/ use psyh or risk assessment/ use psyh or risk factor$.sh. use psyh
15	risk$.ti,ab.
16	or/12-15

17 drug response/ use emez or hospital readmission/ use emez or patient compliance/ use emez or recurrent disease/ use emez or recurrence risk/ use emez or relapse/ use emez or remission/ use emez or remission induction/ use emez or exp treatment failure/ use emez or treatment outcome/ use emez or treatment refusal/ or treatment response/ use emez or treatment withdrawal/ use emez or patient satisfaction/ use emez

18 medication adherence/ use mesz or patient compliance/ use mesz or patient dropouts/ use mesz or patient readmission/ use mesz or recurrence/ use mesz or remission/ use mesz or treatment failure/ use mesz or treatment outcome/ use mesz or treatment refusal/ use mesz or patient satisfaction/ use mesz

19 psychiatric hospital readmission/ use psyh or treatment dropouts/ or treatment refusal/ use psyh or psychotherapeutic resistance/ use psyh or "relapse (disorders)"/ use psyh or treatment compliance/ use psyh or treatment outcomes/ use psyh or client satisfaction/ use psyh

20 (adher$ or complian$ or nonadher$ or nonrespon$ or recur$ or readmi$ or re admi$ or rehospital$ or re hospital$ or relaps$ or remission$ or respond$ or response$ or attrition or drop out$ or select$ or allocat$ or prefer$ or declin$ or uptak$ or ((antidepres$ or interven$ or therap$ or treatment) adj3 (refus$ or success$))).ti,ab.

21 (expected treatment response or identification of treatment failure or (patient$ adj2 deteriorat$)). ti, ab, kw.

22 or/17-21

23 ((risk$ adj2 (adher$ or complian$ or nonadher$ or nonrespon$ or recur$ or readmi$ or re admi$ or rehospital$ or re hospital$ or relaps$ or remission$ or respond$ or response$ or ((antidepres$ or interven$ or therap$ or treatment) adj3 refus$))) or (risk$ adj4 (predict$ or prognos$))).ti,ab.

24 systematic review/ use emez or meta analysis/ use emez

25 meta analysis.sh,pt. use mesz or "review literature as topic"/ use mesz

26 literature review/ use psyh or meta analysis/ use psyh

27 (exp bibliographic database/ use emez or (((electronic or computer$ or online) adj database$) or bids or cochrane or embase or index medicus or isi citation or medline or psyclit or psychlit or scisearch or science citation or (web adj2 science)).ti,ab.) and (review$.ti,ab,sh,pt. or systematic$.ti,ab.)

28 (exp databases, bibliographic/ use mesz or (((electronic or computer$ or online) adj database$) or bids or cochrane or embase or index medicus or isi citation or medline or psyclit or psychlit or scisearch or science citation or (web adj2 science)).ti,ab.) and (review$.ti,ab,sh,pt. or systematic$.ti,ab.)

29 (computer searching/ use psyh or (((electronic or computer$ or online) adj database$) or bids or cochrane or embase or index medicus or isi citation or medline or psyclit or psychlit or scisearch or science citation or (web adj2 science)).ti,ab.) and (review$.ti,ab,pt. or systematic$.ti,ab.)

30 ((evidence or quantitative$ or systematic$) adj2 (overview or review)).ti,ab.

31 (metaanal$ or meta anal$ or metasynthes$ or meta synthes$).ti,ab.

32 ((pool$ or combined or combining) adj2 (data or trials or studies or results)).ti,ab.

33 or/24-32

34 (5 and (or/11,16) and 22 and 33) or (and/5,23,33)

35 remove duplicates from 34

36 35 not (comment$ or dissertation$ or editorial$ or letter$).pt.

Appendix B: NICE methodology checklist for systematic reviews and meta-analyses.

Study identification *Include author, title, reference, year of publication*			
Guideline topic:	**Review question no:**		
Checklist completed by:			
SCREENING QUESTIONS			
In a well-conducted, relevant systematic review:	***Circle one option for each question***		
The review addresses an appropriate and clearly focused question that is relevant to the guideline review question	Yes	No	Unclear
The review collects the type of studies you consider relevant to the guideline review question	Yes	No	Unclear
The literature search is sufficiently rigorous to identify all the relevant studies	Yes	No	Unclear
Study quality is assessed and reported	Yes	No	Unclear
An adequate description of the methodology used is included, and the methods used are appropriate to the question	Yes	No	Unclear

Appendix C: Number of specific treatments delivered by the services.

Intervention	Number of patients	Percentage of included patients
CBT recovery self-help book	682	6.38%
CBT anxiety self-help book	470	4.40%
Books on prescription	558	5.22%
Self-help (other)	5502	51.45%
Cognitive behavioural therapy (CBT)	4242	39.67%
Interpersonal therapy	63	0.59%
Couples	37	0.35%
Counseling	453	4.24%
Computerised CBT (CCBT)	327	3.06%
Group psycho-education	576	5.39%
Behavioural activation	1490	13.93%
Medication advice	174	1.63%
Structure exercise	276	2.58%
Signposting	2565	23.99%
information only	2586	24.18%
Custom therapy	166	1.56%

Appendix D: Graphical representation of the latent profiles

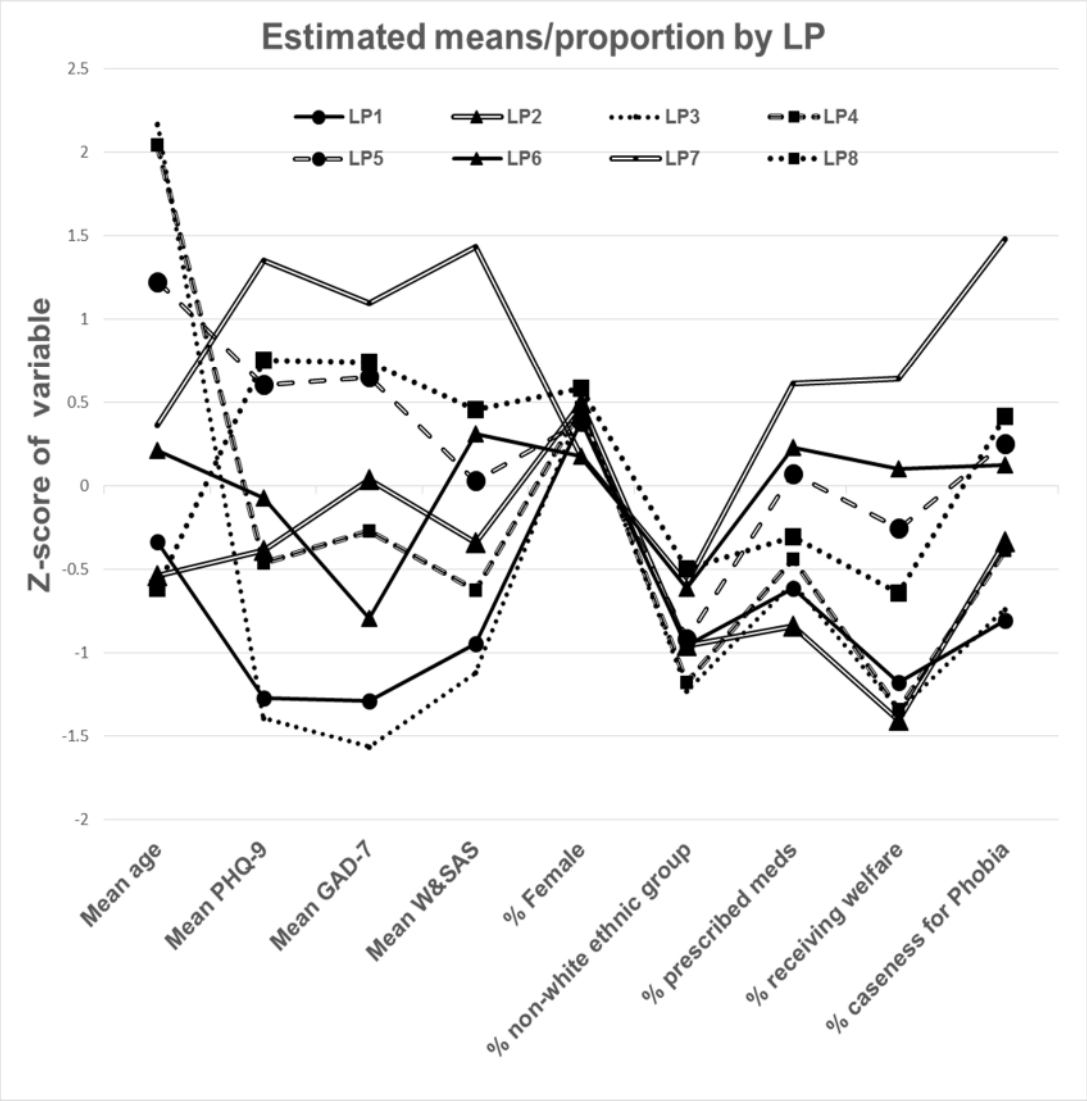

Figure C1. Graphical representation of the distribution of patient variables between latent profiles.

The distribution of the patient variables for each LP is presented in figure A1. Due to the different scaling of the continuous and dichotomous items included in the latent profile analysis, it was not possible to directly compare the mean scores and proportions of each variable between LPs. Instead, all mean scores and proportions for each item were standardised and z-scores were used to present the distribution between mean scores and proportions for each profile.

Appendix E: Latent growth curve model comparison, including patients with three or more sessions of additional interventions.

Guided Self-help (n=3465)

	PHQ-9				GAD			
	CFI	TFI	RMSEA	SRMR	CFI	TFI	RMSEA	SRMR
Linear	0.907	0.916	0.08	0.58	0.878	0.89	0.083	0.172
Quadratic	0.979	0.978	0.04	0.033	0.977	0.976	0.039	0.047
Quadratic, correlated residuals	0.994	0.991	0.026	0.024	0.99	0.986	0.03	0.04

High intensity (n=4903)

	PHQ-9				GAD			
	CFI	TFI	RMSEA	SRMR	CFI	TFI	RMSEA	SRMR
Linear	0.913	0.913	0.071	0.097	0.902	0.911	0.072	0.116
Quadratic	0.961	0.961	0.049	0.04	0.959	0.964	0.048	0.045
Quadratic, correlated residuals	0.987	0.986	0.031	0.028	0.986	0.985	0.03	0.034

Appendix F: Latent Class Growth Analysis by Latent profiles: model comparison tables.

LP1.

Table E1.1. Model fit statistics LP1 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	3114	3177	3108	0.134	0.893	49/51
k = 1	3122	3192	3116	n/a	n/a	n/a
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	3088	3151	3082	0.002	0.714	34/66
k = 3	3068	3143	3061	0.13	0.905	37/60/3

Table E1.2. Model fit statistics LP1 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	5679	5775	5671	0.042	0.951	94/6
k = 3	5672	5780	5663	0.082	0.949	6/92/1
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	5697	5790	5689	0.02	0.815	69/31
k = 3	5655	5760	5646	0.733	0.792	34/10/56

LP2.

Table E1.3. Model fit statistics LP2 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	27982	28092	28022	<0.001	0.734	72/28
k = 3	27841	27971	27888	<0.001	0.675	54/11/36
k = 4	27805	27955	27860	0.521	0.579	43/23/9/26
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	28008	28118	28048	<0.001	0.731	26/74
k = 3	27913	28044	27961	0.001	0.646	32/12/55
k = 4	27880	28030	27935	0.033	0.619	16/25/47/12
k = 5	27856	28027	27919	0.431	0.65	18/6/45/25/8

Table E1.4. Model fit statistics LP2 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	47108	47272	47167	<0.001	0.67	70/30
k = 3	47004	47188	47070	0.003	0.693	57/5/38
k = 4	46958	47161	47031	0.504	0.626	29/50/4/18
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	47293	47452	47350	<0.001	0.814	63/37
k = 3	46832	47010	46896	0.037	0.757	42/41/18
k = 4	46612	46811	46683	<0.001	0.741	11/23/47/39
k = 5	46539	46757	46617	0.077	0.717	13/15/38/11/23

LP4.

Table E1.5. Model fit statistics LP4 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	4361	4430	4360	0.007	0.811	78/22
k = 3	4333	4414	4332	0.115	0.733	11/35/54
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	4272	4340	4271	0.007	0.786	22/78
k = 3	4257	4338	4256	0.288	0.751	68/20/12

Table E1.6. Model fit statistics LP4 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	7725	7830	7726	0.293	0.682	29/71
k = 1	7703	7812	7704	n/a	n/a	n/a
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	7785	7888	7768	<0.001	0.818	68/32
k = 3	7735	7850	7736	0.04	0.746	14/47/40
k = 4	7722	7850	7723	0.812	0.729	7/25/49/18

LP5.

Table E1.7. Model fit statistics LP5 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	10254	10339	10270	0.059	0.702	46/54
k = 1	10211	10304	10228	n/a	n/a	n/a
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	9808	9893	9824	<0.001	0.75	51/49
k = 3	9764	9866	9783	0.16	0.678	37/24/39

Table E1.8. Model fit statistics LP5 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	20587	20723	20618	<0.001	0.634	44/56
k = 3	20512	20665	20547	0.1	0.678	52/26/22
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	19854	19985	19884	<0.001	0.774	49/51
k = 3	19598	19746	19632	0.041	0.773	50/26/24
k = 4	19491	19656	19529	<0.001	0.754	11/20/39/30
k = 5	19466	19647	19508	0.247	0.724	19/36/24/12/8

LP6.

Table E1.9. Model fit statistics LP6 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	8253	8336	8267	<0.001	0.686	64/36
k = 3	8231	8330	8247	0.325	0.7	7/40/53
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	7889	7973	7903	<0.001	0.762	68/32
k = 3	7844	7942	7860	0.015	0.798	61/34/6
k = 4	7814	7928	7833	0.091	0.725	38/5/25/31

Table E1.10. Model fit statistics LP6 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	21635	21773	21668	<0.001	0.686	30/70
k = 3	21588	21743	21626	0.175	0.564	32/21/47
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	20767	20901	20800	<0.001	0.815	40/60
k = 3	20576	20727	20613	<0.001	0.767	16/40/44
k = 4	20508	20676	20549	0.004	0.729	40/14/18/29
k = 5	20485	20669	20529	0.164	0.717	24/7/38/18/14

LP7.

Table E1.11. Model fit statistics LP7 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	8299	8382	8312	<0.001	0.742	67/33
k = 3	8253	8351	8269	0.196	0.66	46/29/25
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	7940	8024	7954	<0.001	0.819	70/30
k = 3	7910	8009	7926	0.381	0.751	14/60/27

Table E1.12. Model fit statistics LP7 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	39562	39722	39617	<0.001	0.771	65/35
k = 3	39302	39480	39363	0.104	0.785	8/55/36
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	37749	37903	37801	<0.001	0.846	35/65
k = 3	37148	37322	37208	<0.001	0.818	35/50/15
k = 4	37000	37193	37066	0.028	0.838	3/15/34/48
k = 5	36897	37110	36970	0.257	0.818	10/45/11/3/31

LP8.

Table E1.13. Model fit statistics LP8 GSH

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	23629	23735	23665	<0.001	0.69	54/46
k = 3	23502	23627	23545	<0.001	0.695	13/44/43
k = 4	23478	23622	23526	0.138	0.62	37/11/32/20
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	22672	22777	22707	<0.001	0.712	49/51
k = 3	22563	22688	22605	0.005	0.669	29/32/39
k = 4	22531	22675	22580	0.012	0.637	21/22/25/22
k = 5	22521	22684	22576	0.653	0.633	12/11/32/24/21

Table E1.14. Model fit statistics LP8 HI

PHQ						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	51149	51315	51210	<0.001	0.655	50/50
k = 3	50949	51135	51018	<0.001	0.619	29/46/25
k = 4	50846	51052	50922	0.012	0.583	17/33/23/27
k = 5	50806	51033	50890	0.001	0.642	23/35/2/14/26
k = 6	50787	51034	50878	0.114	0.585	2/26/25/14/17/16
GAD						
k model	AIC	BIC	Adj-BIC	VLMR-LRT (p=)	Entropy	% individuals per class
k = 2	49475	49637	49535	<0.001	0.784	46/54
k = 3	48936	49118	49003	<0.001	0.755	31/45/24
k = 4	48809	49010	48883	0.005	0.696	36/18/27/19
k = 5	48702	48924	48784	0.104	0.667	19/25/17/16/23