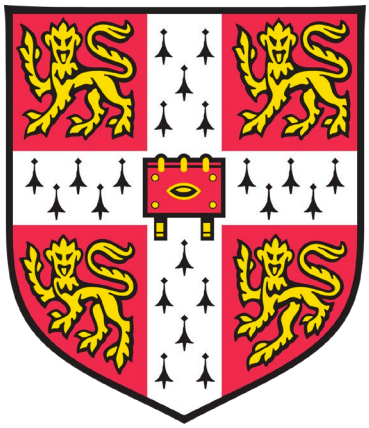
# Predicting trajectories of symptom change during and following treatment in adolescents with Unipolar Major Depression.



Siân Emma Davies

Jesus College,

University of Cambridge

24th April 2019

This thesis is submitted for the degree of Doctor of Philosophy

**Declaration**

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

**Predicting trajectories of symptom change during and following treatment in adolescents with Unipolar Major Depression.**

Siân Emma Davies

**Summary/Abstract**

**Objective**: Definitions of treatment response used in randomised controlled trials for unipolar major depression are non-standardised and arbitrary. Proportion of non-responders has been estimated as ranging from 20%-40% across such trials. I aimed to classify depressed adolescents recruited to the UK IMPACT trial into different trajectories of depression symptom response using a longitudinal data-driven approach: growth mixture modelling (GMM) and investigate potential predictors of trajectory classes in this cohort.

**Method**: 465 depressed adolescents received manualised psychological therapies in the IMPACT trial. GMM was used to plot the trajectories of self-reported depressive symptoms measured at 6 nominal time points over 86 weeks from randomisation, and categorise patients into their most likely trajectory class. Chapters 2-4 investigated the prognostic value of a number of variables. Chapter 2 investigated a battery of demographic and clinical variables including subclinical psychotic symptoms. Chapter 3 focused on a subsample of patients: 109 of the 465 with structural magnetic resonance imaging (MRI) data. FreeSurfer was used to extract cortical thickness (CT) and surface area (SA) measures from 4 regions of interest (ROI; rostral anterior cingulate, dorsolateral prefrontal cortex, orbitofrontal cortex, and insular cortex). Chapter 4 focused on another subsample of patients: 166 of the 465 with salivary basal cortisol data at both waking and evening. Logistic regressions were used in Chapters 2-4 to investigate whether these variables were associated with increased likelihood of membership to a certain GMM class.

**Results**: A piecewise GMM categorised patients into two classes with initially similar and subsequently distinct trajectories. Both groups had a significant decline in depressive symptoms over the first 18 weeks. Eighty-four per cent of patients were classed as "continued-improvers" through reporting an improvement in symptoms over the full duration of the study. A further class of 15.9% of patients were termed "halted-improvers" who had higher depression scores at baseline, faster recovery but no further improvement after 18 weeks. This data-driven method of classification showed only moderate agreement with a priori classification methods, and suggested misclassification rate could be as great as 31%. Co-morbid psychiatric disorders at baseline moderately increased the liability of being

a member of the halted-improvers class (OR = 1.40, CI 1.00-1.96). No other clinical, neurological or cortisol variable reached statistical significance for predicting trajectory class. **Conclusion**: A fast reduction in depressive symptoms in the first few weeks of treatment may not indicate a good prognosis. Further, halted-improvement may not be apparent until after 18 weeks of treatment. Capitalizing on repeated symptom assessments with longitudinal data driven modelling may improve the precision of revealing patient groups with differential responses to treatment. Further work should seek to validate these trajectories in a separate sample of adolescents.

**Contents**

**Acknowledgements**

The IMPACT study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project number 06/05/01). MR-IMPACT was funded by the Medical Research Council (grant: G0802226) and IMPACT-GH was funded by the Evelyn Trust. My doctoral studentship was awarded by the Neuroscience in Psychiatry (NSPN.Org) Consortium itself funded by a strategic award from the Wellcome Trust (095844/Z/11/Z) awarded to Professor Ian Goodyer and Professor Peter Fonagy. I would like to thank the principle investigators, project managers and all research assistants who collected and curated the IMPACT data over the three research sites, and am especially grateful to all the patients who gave up so much of their time for this research, without whom this work would not be possible.

I would also like to thank my supervisors Dr. Paul Wilkinson, and Professor Ian Goodyer for their guidance throughout these 3 years, Dr. Anne-Laura van Harmelen and Dr. John Suckling for their neuroimaging advice and Dr. Rogier Kievit, Dr. Lizanne Schweren, Sharon Neufeld and Eleonore van Sprang for their statistical expertise.

On a personal note, I would like to thank all my friends who have supported me, knowingly or unknowingly during this work. Maggie Westwater, Dan Davies, Jessica Fritz, Johanna Finnemann, Ciara O'Donnell, Jenny Scott, Christina Maurice, Junaid Bhatti, Rebecca Broome, Ellie Tew, Sarah Davidson, Jodie Chalmers, Alexa Netty, Laura Collins, Patrick Moreno, Emma Long, Sarah Sauchelli-Toran and Tarja de Silva: your brilliant minds and kind hearts have helped me through my time at Cambridge in many ways, and filled my days with laughter, words of encouragement, cake and camaraderie.

Finally, words cannot express my gratitude to my wonderful mum, brother and sister-in-law. Their unwavering belief in my ability, unconditional support, positivity and wise offerings of perspective has meant the world to me.

There is much more to this thesis than the words on these pages. There is a multitude of people holding up these words, as they did me throughout my PhD. I cannot thank them enough, and I dedicate this work to them all.

**Abstract**

**Objective**: Definitions of treatment response used in randomised controlled trials for unipolar major depression are non-standardised and arbitrary. Proportion of non-responders has been estimated as ranging from 20%-40% across such trials. I aimed to classify depressed adolescents recruited to the UK IMPACT trial into different trajectories of depression symptom response using a longitudinal data-driven approach: growth mixture modelling (GMM) and investigate potential predictors of trajectory classes in this cohort.

**Method**: 465 depressed adolescents received manualised psychological therapies in the IMPACT trial. GMM was used to plot the trajectories of self-reported depressive symptoms measured at 6 nominal time points over 86 weeks from randomisation, and categorise patients into their most likely trajectory class. Chapters 2-4 investigated the prognostic value of a number of variables. Chapter 2 investigated a battery of demographic and clinical variables including subclinical psychotic symptoms. Chapter 3 focused on a subsample of patients: 109 of the 465 with structural magnetic resonance imaging (MRI) data. FreeSurfer was used to extract cortical thickness (CT) and surface area (SA) measures from 4 regions of interest (ROI; rostral anterior cingulate, dorsolateral prefrontal cortex, orbitofrontal cortex, and insular cortex). Chapter 4 focused on another subsample of patients: 166 of the 465 with salivary basal cortisol data at both waking and evening. Logistic regressions were used in Chapters 2-4 to investigate whether these variables were associated with increased likelihood of membership to a certain GMM class.

**Results**: A piecewise GMM categorised patients into two classes with initially similar and subsequently distinct trajectories. Both groups had a significant decline in depressive symptoms over the first 18 weeks. Eighty-four per cent of patients were classed as "continued-improvers" through reporting an improvement in symptoms over the full duration of the study. A further class of 15.9% of patients were termed "halted-improvers" who had higher depression scores at baseline, faster recovery but no further improvement after 18 weeks. This data-driven method of classification showed only moderate agreement with a priori classification methods, and suggested misclassification rate could be as great as 31%. Co-morbid psychiatric disorders at baseline moderately increased the liability of being a member of the halted-improvers class (OR = 1.40, CI 1.00-1.96). No other clinical, neurological or cortisol variable reached statistical significance for predicting trajectory class.

**Conclusion**: A fast reduction in depressive symptoms in the first few weeks of treatment may not indicate a good prognosis. Further, halted-improvement may not be apparent until

after 18 weeks of treatment. Capitalizing on repeated symptom assessments with longitudinal data driven modelling may improve the precision of revealing patient groups with differential responses to treatment. Further work should seek to validate these trajectories in a separate sample of adolescents.

**Depression and treatment response in adolescence: where do we currently stand?**

*Importance*

Major depressive disorder (MDD) is one of the most common mental health disorders in the world (Alonso et al., 2004; Kessler et al., 2003; Wittchen et al., 2011). Lifetime prevalence rates have been reported between 14 and 16% of the population in Europe (Alonso et al., 2004) and the United States (Kessler et al., 2003) respectively. These may even be an underestimate: for example within the European Union (EU), it has been shown that from 5 years of age around 38% of the population suffer a mental health disorder each year (Wittchen et al., 2011). Of these, 7% are attributable to major depression (Wittchen et al., 2011). Furthermore, the UK Biobank study reported that the prevalence rates of probable recurrent major depression were 12.2% (Smith et al., 2013). Such high prevalence rates indicate a significant disease burden (Wittchen et al., 2011). Indeed Murray and colleagues (2013) highlighted that within the UK, the major cause of years lived with disability were mental health disorders, and depression has been reported responsible for the largest amount of non-fatal disease burden in the world (Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004). Depression is currently the fourth leading cause of disease burden worldwide (Üstün et al., 2004), and only expected to increase in rank by 2030 (Mathers & Loncar, 2006). Consequently, this poses a substantial economic cost to society. Depression has been ranked one of the top 5 most costly brain disorders in the UK (Fineberg et al., 2013), and in the year 2000, the UK spent £370 million on treatment of adult depression, with the total cost of the illness estimated at over £9 billion (Thomas & Morris, 2003a).

Beyond contributing to such global and economic burden, the suffering experienced by the individual patient is substantial (Cuijpers, 2017). Major depression significantly reduces a patient's quality of life (Pyne, Patterson, Kaplan, Gillin, Koch, & Grant, 1997; Saarni et al., 2007), and has an impact on future work, social function, and personal relationships (Weissman et al., 1999). Patients with depression have also been shown to exhibit higher incidences of physical illness (Farmer et al., 2008). Despite such systemic implications of major depression, Wittchen and colleagues (2011) reported that there was no evidence that the treatment of mental health disorders has improved since 2005, and statistical modelling has suggested that 60% of the current burden of mental health disorders is due to the inadequacies of current treatment (Andrews, Issakidis, Sanderson, Corry, & Lapsley, 2004).

Adequate treatment of MDD has been reported as low as 42% (Kessler et al., 2003), with the percentage of patients failing to experience a significant therapeutic response and classed as resistant to treatment ranging from 20% to 55% (Goodyer et al., 2008; Thomas et al., 2013).

Why does treatment of major depression show such mediocre response rates? It is not insignificant that a quarter of adults suffering depression report that their illness began during adolescence (Zisook et al., 2007), and the mean age of onset is reportedly around 15 years of age (Lewinsohn, Clarke, Seeley, & Rohde, 1994). Rates of new onset of depression at this age bracket increase from 3% to 7% (Merikangas, Nakamura, & Kessler, 2009), such that adolescence denotes the time of highest incidence risk rate for the emergence of major depression over the life course (Avenevoli, Knight, Kessler, & Merikangas, 2008). Onset in this second decade is a risk factor for subsequent relapse and recurrence in adulthood (Fombonne, Wostear, Cooper, Harrington, & Rutter, 2001; Lewinsohn, Clarke, Seeley, & Rohde, 1994) and impairment in adult life (Zisook et al., 2007). Moreover, the adult literature suggests that in clinical samples, factors relating to chronicity of illness often associate with treatment resistance, recurrence and relapse including duration and number of episodes (Mueller et al., 1999), early age of onset (Lewinsohn, Rohde, Seeley, & Fischer, 1993) and first episode versus chronic depression (Cuijpers, Andersson, Donker, & van Straten, 2011). Consequently, by the time patients enter treatment as adults, their condition may be too chronic to show favourable responses to current treatment.

To optimise outcome for patients, it is imperative that we advance our understanding of depression in adolescence, and investigate the efficacy of current treatment at this delicate stage of neural, psychological and social development. However, the number of treatment trials for adolescent depression is markedly less than those for adult depression (Cipriani et al., 2018). Current guidelines for adolescent depression have developed from adapting treatments that have shown success first in adult populations (Weersing, Jeffreys, Do, Schwartz, & Bolano, 2017). As such, I shall begin with a discussion of the current state of the treatment of adult depression.

*Treatment of adult depression*

The recommended guidelines for the treatment of adult depression consist of either pharmacological or psychological treatment, or a combination of both (NICE, 2015, NICE,

2016). Pharmacological treatment is typically second-generation antidepressants, such as selective serotonin reuptake inhibitors (SSRIs). SSRIs are more commonly prescribed as a first-line treatment, often because of resource availability rather than patient preference or treatment efficacy (Cipriani et al., 2018). However, a significant proportion of patients voice a preference for non-pharmacological options where possible (Gartlehner et al., 2017), due to an awareness of risks and side-effects associated with antidepressants (Gartlehner et al., 2008) and established evidence base for psychological therapies (NICE, 2015, NICE, 2016).

Measuring the effectiveness of the treatment of depression seems simple: do patients get better? Do they no longer suffer from depression? However, these questions require clinical researchers to measure the change in a patient's depressive condition objectively. Typically, this is achieved through the documentation of a change in the presence, and/or severity of symptoms over a given time period (Nierenberg & DeCecco, 2001). Treatment efficacy results can therefore be viewed as how much support there is for the claim that a given treatment significantly reduces a patient's symptoms, more so than they would without the treatment.

The efficacies of both pharmacological and psychological therapies in adult depression have been extensively investigated (Cipriani et al., 2018; Cuijpers, 2017). For instance, a recent network meta-analysis showed that a wide range of second-generation antidepressants demonstrate significant efficacy rates above placebo in treating MDD (Cipriani et al., 2018). The authors noted that while differences between antidepressants were much smaller than between drug versus placebo, a number of SSRIs (including escitalopram, paroxetine and sertraline) had higher response and lower dropout rates than other antidepressants. This helps inform clinical practice of more favourable first-line treatment options for patients. Prior to this review, the field has been associated with significant publication bias, with the percentage of studies showing a positive outcome decreasing by approximately 40% when unpublished data were considered (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). Cipriani and colleagues (2018) however, included a large amount of unpublished work in their analysis of adult depression, making their efficacy findings quite robust.

In terms of psychological therapies, Cuijpers (2017) reviewed a series of meta-analyses on a variety of psychological treatment modalities in adult depression. He concluded that all were

equally effective at treatment MDD, with very few differences between treatments (Cuijpers, 2017). Psychological therapies were also associated with improvement in a number of quality-of-life measures, including social support (Park, Cuijpers, van Straten, & Reynolds, 2014) and dysfunctional thinking (Cristea et al., 2015). Moreover, a number of meta-analyses have found that the difference in efficacy between antidepressant medications and psychotherapies is of little clinical relevance (Cuijpers, Straten, Oppen, & Andersson, 2008; Cuijpers, Sijbrandij, et al., 2013). Combined treatment however, generally out-performs monotherapies (Cuijpers, 2017; Cuijpers, Sijbrandij, et al., 2014; Cuijpers, van Straten, Warmerdam, & Andersson, 2009a), in both the acute-phase of treatment and maintenance phase (Karyotaki, Smit, Holdt Henningsen, et al., 2016). Taken together, the literature to date provides some reassurance that currently prescribed treatments for adult depression show a degree of effectiveness, in that they appear to significantly improve the rate of patient's symptom reduction over time. Moreover, extreme deterioration following treatment of either modality is uncommon; occurring in only 1% of cases (Vittengl et al., 2016).

*Problems with research in the treatment of adult depression*

While short-term treatment outcomes appear promising, our understanding of longer-term effects is less clear. There is some evidence of the effects of psychological therapies lasting 6 months or more, and more likely if "maintenance" sessions are offered beyond the acute treatment phase (Karyotaki, Smit, de Beurs, et al., 2016). However, the long-term effects of antidepressant medication are uncertain (Cipriani et al., 2018). One meta-analysis directly compared efficacy rates between cognitive-behavioural therapy (CBT) and pharmacotherapies for 6-18 months post-treatment (Moncrieff, 2018). The authors found that despite comparable efficacies during acute-treatment, pharmacotherapy required continuation to maintain efficacies matching CBT. Conversely, CBT was more likely to have continued effects post-treatment even if terminated at the acute-phase, with reduced risk of relapse for patients compared with the termination of pharmacotherapies (Moncrieff, 2018).

Some authors have also questioned the reliability of the effect-sizes reported in meta-analytic data, particularly for anti-depressant trials (Cuijpers, 2017; Moncrieff, 2018). The effectiveness reported by Cipriani and colleagues (2018) must be considered in a short-term context, and may be an overestimation for longer-term follow-ups. Moreover, side-effects of

the drugs themselves could allow for patients to become un-blinded to their treatment condition in placebo trials, and consequently result in inflated effect-sizes (Moncrieff, 2018). Psychotherapies have also faced similar questions of effect-size reliability due to the influence of study quality and publication bias (Cuijpers, 2017; Cuijpers et al., 2011).

Perhaps more critical to this current thesis is the argument of relevance. Patients receiving treatment may indeed show a greater reduction in severity above placebo, but if the size of reduction is not meaningful, then treatments can be efficacious without really having any impact on disease burden. There is some debate over how large a statistical effect is necessary to translate to clinically observable improvement in symptom reduction and patient recovery. Moncrieff (2018) argued that the mean difference reported in Cipriani's (2018) review equated to an improvement of only 2 points on the Hamilton Rating Scale for Depression (HRSD); a minimum difference of 8 has been suggested necessary for a patient-perceived improvement (Moncrieff & Kirsch, 2015). Indeed, one the largest longitudinal studies investigating the long-term effects of antidepressant medications found that remission, defined as <7 on the HRSD, was only achieved by 28% of patients (Trivedi et al., 2006). Moreover, despite further multiply-staged treatments for unremitted patients (Vittengl et al., 2016), remission became increasingly unlikely for patients requiring more treatment steps, and relapse rates also increased (Rush et al., 2006; Pigott, Leventhal, Alter, & Boren, 2010). This is perhaps unsurprising given that the evidence for efficacy differences between second-generation antidepressants is small (Cipriani et al., 2018), and that no augmentation option offered in this study was psychologically-based (Rush et al., 2006). While treatments may show better outcomes compared with placebo, the exact extent of their ability to provide a clinically meaningful reduction in symptoms remains uncertain.
A number of studies have investigated the effectiveness of psychotherapies as a prevention (rather than intervention) method, in patients with subclinical depressive symptoms (van Zoonen et al., 2014). The meta-analytic results of such studies are encouraging, showing that preventive interventions were associated with an average 21% reduction in the number of new cases of clinical depression over long-term follow-ups (typically 6-12 months), compared with control conditions (van Zoonen et al., 2014). Catching and treating depressive symptomatology early, before the onset of clinical levels of severity, may therefore be the optimal time at which our current treatments can be most effective

(Cuijpers, 2017). Consequently, it is perhaps more imperative for us to investigate treatment effects when the condition typically begins to manifest: adolescence.

*Treatment of adolescent depression*

Clinicians have found that appropriate treatment during adolescence can result in rapid improvements in depressive symptoms and functioning in first-episode patients (Goodyer & Wilkinson, 2018). As mentioned above, the field of research for the treatment of adolescent is small but continuously growing (Weersing et al., 2017). Current guidelines for the treatment of adolescent depression, similar to adults, typically involve psychotherapy, alone or in combination with SSRIs (NICE, 2015). While developed from adapting adult treatments (Weersing et al., 2017), two of the largest randomised controlled trials of antidepressants and psychotherapies conducted in depressed adolescents supported these current advised practices (Brent et al., 2008; March et al., 2004). They concluded that combination treatment (SSRIs plus CBT) were superior to either medication or psychotherapy alone (Brent et al., 2008; March et al., 2004). It is important to note however that despite promising support for combination treatment, more research is required. The effect-size reported for combination treatment above SSRIs alone in those trials were not large differences, and were only true for some outcome measures (Brent et al., 2008; March et al., 2004). Moreover, the use of antidepressant medication in adolescents has had some debate. For instance, meta-analytic results of medication trials have suggested that only fluoxetine showed a significant difference to placebo (Cipriani et al., 2016). Other antidepressants have shown much higher intolerance rates and a greater number of adverse events compared with fluoxetine (Cipriani et al., 2016). SSRIs have also been reported to carry a higher risk of suicidality (Goodyer & Wilkinson, 2018) and aggressive behaviour (Sharma, Guski, Freund, & Gøtzsche, 2016) in adolescents. Consequently, the risk-benefit ratio for antidepressant medication in adolescents means that their prescription should be considered with caution (Cipriani et al., 2016).

Conversely, the evidence base for psychological treatments in adolescent depression is largely supportive. Firstly, the study upon which this thesis draws its data (outlined later), showed that 77% of patients randomised to one of three psychological treatments, achieved remission at long-term follow-up (86 weeks) (Goodyer et al., 2017). This was a rise from 48% at 12 weeks, which demonstrates improving effects following psychotherapies in this

population. While this study lacked a control group for adequate comparison, two recent reviews (one incorporating a network meta-analysis) have strongly supported the efficacy of psychological treatments in adolescents (Weersing et al., 2017; Zhou et al., 2015). From the evidence of 42 randomised controlled trials of CBT and interpersonal psychotherapy (IPT), most studies found that CBT or IPT outperformed control groups across studies. Moreover, no study found either treatment faired worse than active controls (Weersing et al., 2017). Although the authors acknowledged the small database for IPT studies, they concluded that both treatments were well established interventions for adolescent populations and equally efficacious (Weersing et al., 2017). In addition, CBT has been found to have substantial effects on specifically reducing suicidality (March et al., 2004), which is not without clinical significance.

*Problems with research in the treatment of adolescent depression*

As the literature stands, current specialist treatments on the whole appear to reduce symptoms of depression in adolescents above routine care, but a number of discrepancies still exist. Firstly, as with the adult literature, the long-term effects of combination treatments are not clear (Goodyer et al., 2007). Goodyer and colleagues (2007) investigated the effects of SSRI alone or in combination with CBT at 28, rather than the 12 week period of previous studies. They concluded that there was no beneficial effect of combined SSRI plus CBT treatment over SSRIs plus routine specialist care in adolescent depression. However, that may be because of higher-than-expected effectiveness of the control intervention rather than lack of effectiveness of the additional CBT. Moreover, while the Treatment for Adolescent Depression Study (TADS) found a superiority effect for combination treatment, CBT alone failed to show significantly better short-term responses than pill-placebo in their trial (March et al., 2004).

Similar to the observations of the adult literature (Cipriani et al., 2018; Cuijpers, 2017), the effect size of treatment in adolescent depression has also demonstrated a decline over time (Weisz, McCarty, & Valeri, 2006). The overall reported effect sizes in adolescent depression are moderate at best (March et al., 2004; Weisz et al., 2017, 2006) and compared with other mental health conditions, are the smallest in the literature (Weisz et al., 2017). The effect sizes also do not seem to improve over increased duration of follow-up (Weisz et al., 2006). Relapse and recurrence rates remain undesirably high, occurring in at least 25% of patients

(Kennard et al., 2009; B. Vitiello et al., 2011). Consequently, some have suggested that we have reached a ceiling with the effectiveness of our current strategies in adolescence (Weersing et al., 2017; Weisz et al., 2006).

Recent reviews of the field commented on the trend of declining effect size, suggesting that factors relating to study quality and population criteria may have contributed to these changes (Goodyer & Wilkinson, 2018; Weersing et al., 2017). Interestingly, studies that have contributed to raising caution over efficacy reports in adolescent depression have been those that have employed the use of active control conditions (Brent et al., 1997; Goodyer et al., 2017) and recruitment of more clinically severe samples, with significant comorbidities (Goodyer et al., 2007; Goodyer et al., 2017; March et al., 2004; Shirk, Deprince, Crisostomo, & Labus, 2013). Studies of more complex cohorts were more likely to report null effects (Goodyer et al., 2007; Shirk et al., 2013), and a recent multilevel meta-analysis actually failed to show a significant benefit of psychotherapies in cases of multiple comorbidities (Weisz et al., 2017). Taken together, the results suggest that while current treatments may benefit some patients (Goodyer & Wilkinson, 2018; Weersing et al., 2017), the extent to which they help reduce symptoms, particularly in those more complex cases (which in turn, are those cases more likely to be seen in real-world clinics), is still unclear.

Many questions regarding current treatments modalities for adult and adolescent depression are still outstanding. Cuijpers (2017) advocated that given the equivalency of treatments available, our efforts should not focus on designing novel treatments, because it is unlikely that these will substantially differ. We need to understand why these treatments are not more effective for patients and what specific characteristics of both the patient, their depressive condition and current treatment modalities are not matching up to give an optimal response. Our focus needs to be on targeting these questions (Cuijpers, 2017). This more precise approach to treatment choice is more likely to guide research in understanding what treatment works best for which depressed patient.

It is possible that the cause of modest effect sizes in the treatment of adolescent depression is clinical heterogeneity. In group comparisons, the inclusion of patients for whom treatment fails completely may lower the overall effect size of the study, thus masking the effect size in those for which treatment is extremely effective. This is not insignificant, considering that at

least 20% of adolescents show no response to available treatments (Goodyer et al., 2008; March et al., 2004). This is why some authors have advocated that to understand treatment effectiveness fully, we must investigate individual differences in response (Goodyer & Wilkinson, 2018). Who is the treatment particularly effective for, and why? Identifying these characteristics that could predict such response in adolescent depression would be beneficial for advancing our treatment of this condition, and reducing the overall disease burden.

However, before these questions can be addressed, we must ask ourselves what does treatment response look like? This question is of critical importance. The misclassification of patients as responsive can have substantial influence on the outcomes of clinical trials, affecting prescribing guidelines and subsequently, impacting the treatment any individual patient receives. Despite this, the current literature is a long way from agreeing upon a gold standard definition (Berlim & Turecki, 2007). Is a simple pre-defined reduction in symptoms sufficient to classify an individual as responsive, or does such a definition omit important information of clinical relevance? Without clarity in the definition of an unfavourable outcome, investigations of patients defined as such will be meaningless (Thibodeau et al., 2015).

*Thesis Outline*

This doctoral work aims to focus in on questions relating to individual differences in symptom change and predictors of such change. My first objective however, will be to take a step back and evaluate whether our current methods for measuring symptom change over time are the most effective for accurately differentiating patients who show unfavourable responses. In doing so I shall investigate alternative empirical methods previously used in depressed adults, to define subgroups of patients, based on their trajectory of symptom change over time. I will do this in a group that have been receiving psychological treatment (Goodyer et al., 2011), and thus are expected to observe a downward trajectory of symptom change, at least in the majority of patients. The evaluation of this approach to defining symptom change in adolescent depression will be the focus of my first chapter of work.

From the foundation set out in Chapter 1, Chapters 2-4 will provide novel perspectives on questions relating to prognosis in adolescent depression. I shall utilise the categories defined

in Chapter 1 to investigate predictors of symptom change trajectories. Chapter 2 will specifically focus on a battery of demographic and clinical predictors of these trajectories, with a specific a priori interest in subclinical psychotic symptoms.

Empirical approaches are yet to investigate biological predictors of trajectory classes. Thus, Chapters 3 and 4 will provide a preliminary investigation of these questions, by using specific subsamples of the Improving Mood with Psychoanalytic and Cognitive Therapies (IMPACT) cohort. Chapter 3 will utilise the Magnetic Resonance-IMPACT (MR-IMPACT) cohort; a sample of 109 adolescents for whom structural magnetic resonance imaging (MRI) were collected (Hagan et al., 2013). This chapter will focus on two neurological measures of cortical thickness and cortical surface area of 4 specific cortical regions that have demonstrated associations with treatment response in depression by a priori methods. Chapter 4 will utilise the IMPACT-Genes and Hormones (IMPACT-GH) cohort: a sample of 166 adolescents for whom salivary cortisol samples were collected. This chapter will focus on peak morning and evening cortisol, which have been shown to differentially relate to treatment response.

Finally, Chapter 5 will discuss the theoretical implications of the findings, and provide a reflective discussion of the methodology used in this thesis and its potential utility in advancing this field.

## Chapter 1: Measuring change.

*Why do we need to measure change?*

It is common practice in medical research for the proportions of responders and non-responders to be considered a measure of efficacy and effectiveness for current and novel treatments (Thibodeau et al., 2015). It is an easily understandable statistic for patients and clinicians. Moreover, classifying individuals into response types could offer great economic and public health advantages through personalized treatment strategies (Gueorguieva, Mallinckrodt, & Krystal, 2011; Thibodeau et al., 2015). For example, rapid responders could receive a short course of treatment, and resistant subtypes could be offered more aggressive treatment at outset (Thibodeau et al., 2015). This could help reduce relapse and recurrence rates and also the cost and strain that mental health services currently face (Thomas & Morris, 2003). Furthermore, it would be highly useful to know which variables influence this probability of response for all patients ('predictors') and which variables influence the response to treatment differently in patients receiving different treatments ('moderators'). All of these questions require a measurement of symptom change in depressed patients.

However being depressed or well is not a binary entity. Depressive symptoms actually lie on a continuous spectrum, and, in common with a lot of diseases (e.g. hypertension (Chobanian et al., 2003) and diabetes mellitus (American Diabetes Association, 2003)), an arbitrary cut-point has been chosen for diagnosis (Malhi & Byrow, 2016). Therefore to define 'response', researchers must firstly use an accurate measure of depression severity; and secondly operationally quantify the symptom change in depression that would be considered a reasonable 'response' to treatment. At present, both these tasks in depression research are executed inconsistently. The variation that exists across studies in choice of assessment tool and definition of response makes it impossible to achieve a coherent consensus that can guide research and clinical practice (Berlim & Turecki, 2007).

The main focus of my first piece of work will be addressing the latter of these tasks: operationally defining what is a reasonable response to treatment. Analyses of treatment efficacy and effectiveness, or predictors and moderators of change, are all dependent on the choice in the definition of response. As eloquently stated by Thibodeau and colleagues

(2015), *"…the utility of classifying nonresponders is only proportional to the validity of how a nonresponder is defined"* (p.214). This definition has the potential to be an influential factor in the outcome of clinical trials, and consequently affect protocols in medical practice. It is therefore imperative for the definition to reflect patient experience accurately.

*Defining a clinical response to treatment: a continuous approach*

Some efficacy studies report mean change in depression severity, rather than reporting percentage of patients meeting pre-defined categories of responders and non-responders (Emslie et al., 2002; Greeson et al., 2015; Wang et al., 2015). Continuous measures have the advantage of retaining more information regarding absolute change, as the data are not reduced into a small number of groups (Jureidini et al., 2004; Joanna Moncrieff & Kirsch, 2005). This increases the power of the study to detect subtle changes that may be present over time (Altman & Royston, 2006; Royston, Altman, & Sauerbrei, 2006). A study of adolescent depression by Emslie and colleagues, (2002) nicely illustrates this concept: in their placebo-controlled trial their primary efficacy measure was categorical (response; defined as at least a 30% reduction in symptoms), and showed no significant difference between treatment and placebo. However, the authors found that on a continuous scale, patients treated with fluoxetine experienced a significantly greater reduction in depression symptoms than those treated with placebo, and this allowed the authors to report that fluoxetine is efficacious in adolescents with depression.

Many studies investigating predictors of response on a continuous scale use correlations or regressions to uncover associations between improved symptoms and patient characteristics. These studies have illustrated that a constellation of patient characteristics can predict outcomes to a certain treatment including; baseline severity (Tedlow et al., 1998), length of current episode (Perugi, Medda, Zanello, Toni, & Cassano, 2012), gender (Walker & Druss, 2015), comorbidity (Bock, Bukh, Vinberg, Gether, & Kessing, 2010), structural deficits of the brain (Li et al., 2010), social support (Walker & Druss, 2015) and cortisol levels (Zobel, Yassouridis, Frieboes, & Holsboer, 1999), all with varying effect sizes and power. However, while informative, there are two substantial issues with this approach. Firstly, subtle effects and moderators may be academically interesting, but may not be clinically useful. For instance, the effect sizes of the change reported by Emslie and colleagues (2002) were small to moderate at best, and were not sufficient for patients to

experience an improvement that felt meaningful to them. Secondly, the decision to administer treatment is fundamentally binary. It is delivered in a step-wise manner, often starting with one treatment (psychological or medication), then switching treatment or adding augmentation therapies if response is inadequate (NICE, 2015, NICE, 2016). Unless predictors of response can inform the categorical decision-making of clinical practice, they will struggle to influence patient care directly (Uher et al., 2010).

*Defining a clinical response to treatment: a priori approaches*

For research to inform clinical practice, it has become necessary to adhere to a categorical framework (Uher et al., 2010). Consequently, standard reporting of clinical outcomes tend to dichotomize individuals into responders and non-responders (Uher et al., 2010). However, dichotomizing patient response is complicated, because two key questions need to be asked. Firstly, what is a reasonable response to treatment? Secondly, over what timescale does improvement need to come? I shall discuss each in turn below.

At a clinical level, there is a universal agreement that the ideal response to treatment is sustained remission of depression (Nierenberg & DeCecco, 2001). For this reason, Berlim and Turecki (2007) argued that "remission" should be used as a "gold standard" definition for responders. However, defining remission itself is complicated. If remission is taken to mean "free from diagnosis" then theoretically patients could still retain up to 4 symptoms that meet clinical threshold, without meeting a threshold for depression (American Psychiatric Association [APA], 2000). This definition would ignore the substantial influence of residual symptoms on functional status (Keller, 2003, 2005; Zimmerman et al., 2012) and long-term disease course, as residual symptoms are repeatedly shown to be risk factors for relapse and recurrence (Nierenberg et al., 2010; Paykel et al., 1995).

Conversely, if remission is agreed to mean, "free from depressive symptoms" (Nierenberg & DeCecco, 2001) then full remission would be rare in depression (Nierenberg et al., 2010). As many as 90% of patients have been found to retain at least 1 residual symptom (Nierenberg et al., 2010) and the high incidence of comorbidity present in the depressed population (Rohde, Lewinsohn, & Seeley, 1991) would make it unlikely for many patients to be completely free of symptoms. Such a high threshold for a responder is unrealistic; it would ignore the substantial improvement made by some patients, but also it would label the

majority of patients in clinical trials as non-responders (Nierenberg & DeCecco, 2001). This would clearly impact the outcome of clinical trials and interpretations of efficacy (Tedlow et al., 1998). Consequently, some studies adopt a middle ground for their definition; such as "*no more than 1 clinically significant symptom and no associated functional impairment"* (Vitiello et al., 2011, p389-390). Incorporating a measure of psychosocial functioning into the definition allows flexibility with residual symptoms without compromising on functional recovery: a vital aspect for the clinical meaning of remission (Malhi & Byrow, 2016).

To avoid criticism associated with such harsh cut-offs, trials will often report "response rates" alongside "remission rates" to measure treatment efficacy (Trivedi et al., 2006; Vitiello et al., 2011). This allows studies to capture the complexity and individual variability of recovery from depression. Response is often classified by a percentage reduction in symptom score on standardized depressive scales (Montgomery, 1994); most commonly taken as at least a 50% reduction in depressive scores (Berlim & Turecki, 2007; Montgomery, 1994). This approach allows response to be individualized to each patient, and allows for patients who make a significant improvement to be represented adequately in the research findings. However there are two issues with this approach. Firstly, there is still substantial variation across studies in what constitutes a response (Berlim & Turecki, 2007), with vast disagreement between experts. Empirical studies suggest that a cut-off of 60% is more valid (Mulder, Joyce, & Frampton, 2003) while some authors argue a 30% reduction is clinically meaningful (Shelton et al., 2005). Often, cut-offs are chosen to provide equal group numbers (Uher et al., 2010) and some articles lack an operational definition completely (Berlim & Turecki, 2007). Indecision over how to define this outcome within this literature has already led to some serious consequences. Studies by Emslie and colleagues (2002; 1997), and Keller and colleagues (2001) faced legal review over altering protocol-defined cut-offs in later published work, to make differences between treatments seem larger than reality, or to fit with the authors' belief of treatment efficacy (Jureidini et al., 2004; Jureidini, Mchenry, & Mansfield, 2008). They failed to report with clarity the null results of their original primary outcomes between drug and placebo, highlighting instead the positive results of post-hoc comparisons that were not previously included in the study design (Jureidini et al., 2004, 2008).

The second issue with the adoption of a percentage reduction to indicate response is that the reasons for cut-off selection remain quite arbitrary (Uher et al., 2010). One must remember that despite the advantage of categorical approaches in a clinical setting, categorical approaches face criticisms of data reduction. Researchers suggest that dichotomizing individuals loses up to a third of information about the present variation in response patterns (Cohen, 1983), which substantially reduces one's power to detect significant differences (Altman & Royston, 2006). Therefore, with categorisation coming at such a cost, the boundaries chosen must, at the very least, strive to relate to clinical experience. In using a simple 50% percentage reduction in symptoms to define response, individuals who experience a 49% reduction in symptoms and those who experience a 2% reduction in symptoms are considered equivalent. Intuitively, these patients are likely to have very different attitudes towards their illness at the end of trial (Zimmerman et al., 2012), and may experience differential treatment effects that are masked by taking a group mean (Cuijpers, van Lier, van Straten, & Donker, 2005; Stulz, Thase, Klein, Manber, & Crits-Christoph, 2010).

*Temporal considerations*

Regardless of choice of definition for response, most clinical trials assess this outcome at the end of treatment (Emslie et al., 1997; Goodyer et al., 2007; March et al., 2009) and in situations of percentage reduction, compare it to a baseline assessment. Comparisons of response across only two time-points is argued inadequate to discern true treatment effects for a number of reasons. Firstly, despite a need to conserve data due to the aforementioned data reduction issue of categorical approaches, a large amount of longitudinal data remains unutilised in this method. Secondly, the researchers have made several assumptions regarding the speed at which the treatment will work, the effect size it will have and the trajectory shape it will follow (Montgomery, 1994; Nierenberg & DeCecco, 2001; Rush et al., 2006; Uher et al., 2010). However, to date little is definitively known about any of these components in the treatment of depression (Berlim & Turecki, 2007; Uher et al., 2010). Firstly, the choice of final assessment time-point could dramatically influence the category within which a patient would fall. One study discovered that remission cumulatively increased from 17% of patients at the end of the 12-week blinded treatment, to 61% at a 72-week follow-up (Vitiello et al., 2011), and other authors have also suggested chronic depression may take significantly longer to reach remission (Kornstein & Psychiatry, 2001).

There is therefore an inherent danger in short-term clinical trials of prematurely categorizing patients and underestimating the efficacy of treatment and their mechanisms of action (Shelton et al., 2007). Moreover, depressive symptomatology fluctuates over time, making it particularly difficult to discriminate treatment effects from other confounding factors that would affect any single isolated measurement (Nierenberg & DeCecco, 2001; Uher et al., 2010).

In addition, assessing symptomatology over two time-points assumes a linear trajectory exists between beginning and end of treatment, which is an incorrect assumption, given non-linear patterns have been reported to better fit trial data (Keller et al., 2000). Including data on the temporal pattern of symptom course could advance our understanding of how antidepressants or psychological therapies operate (Uher et al., 2010). Furthermore, one should question the interpretation of findings from clinical trials that use only before and after time-points as reporting true remission rates. Remission implies a time-period where an improvement has not only been achieved but sustained, yet many studies do not reference a timeframe as a requirement (Nierenberg & DeCecco, 2001). Operationally defining the length of time for remission to be confidently declared is highly debated, and ranges from as little as 3 to as much as 8 weeks across studies (Berlim & Turecki, 2007; Vitiello et al., 2011).

*Defining a clinical response to treatment: an empirical approach*

In order to consider heterogeneity in symptom change in defining response, categorization methods must allow for an assessment of longitudinal data. Studies have begun to measure symptom severity systematically following the treatment phases of trials in a naturalistic manner (Goodyer et al., 2008; Keller et al., 2000; March et al., 2009). While this provides a rich dataset for investigating change over time, end-point analyses are still used and the detailed longitudinal information is rarely utilized to its potential (Uher et al., 2010). Empirical modelling techniques differ from this approach in two ways. Firstly, they define their groups post-hoc (Ram & Grimm, 2009), meaning that the data define the subgroups. Secondly, they utilise multiple time-point assessments, which is a major strength of this type of approach. The additional data allow for a further description of the developmental shape of change over time, and the extent of inter-individual difference in this change (Ram & Grimm, 2009). The improved ability to characterize complex trajectory patterns could also

inform our mechanistic understanding of treatment action and consequently the development of new treatments.

Growth curve modelling is one empirical technique, and describes symptom change over time using a regression function. The intercept represents baseline severity and slope describes the change in severity over time (Cuijpers et al., 2005). Provided an adequate number of time points are present, there is flexibility for linear, quadratic, cubic or piecewise shapes to be investigated. Piecewise shapes are particularly useful for hypotheses of clinical trials, as they allow for a break and change in the trajectory, which can capture different rates of change at particular stages in clinical trials (Gueorguieva et al., 2011; Uher et al., 2010). However, growth curve modelling assumes that all individuals are drawn from the same population; that one growth curve adequately fits and describes all individuals in the sample (Ram & Grimm, 2009). As mentioned above, this may not be the best representation of a depressive population.

It is plausible that depression is better characterized in multiple, dissimilar subgroups of patients, each with a distinct phenotype (Thibodeau et al., 2015). However, one issue in depression research is that these subgroups are unknown. For known groupings, such as gender or ethnicity, significant inter-individual differences in change over time can be modelled using multiple-group growth curve modelling: an extension of single growth curve modelling. This allows groups to differ on all three aspects of the growth curve model (Ram & Grimm, 2009). Growth mixture modelling (GMM) is a further extension that allows subgroups to be "unobserved", meaning that groups cannot be distinguished based on something that is overtly measurable (Muthén & Muthén, 2017; Ram & Grimm, 2009; Wickrama, Lee, O'Neal, & Lorenz, 2016), making it a promising tool for studying treatment response in depression. GMMs utilize the heterogeneity in longitudinal data to categorize patients into particular unobserved classes that follow a similar trajectory (Uher et al., 2010). A mean growth curve is then estimated for each class, and information on the probability that a given individual belongs to a certain class is often reported to provide an indication of goodness of fit (Ram & Grimm, 2009; Wickrama et al., 2016). The allocation of patients to groups is therefore decided by the modelling process, based on model fit, and cannot be influenced by the researcher.

In practice, GMM is an exploratory approach in which a series of models are run, based on theoretical hypotheses regarding the shape of the overall trajectory, the number of expected subgroups and how these subgroups are expected to differ (Ram & Grimm, 2009). It is standard practice to run models up to and including at least one more latent group than is expected by theory (Ram & Grimm, 2009). Testing more complex models than hypothesised allows for a more rigorous comparison of changes in fit and convergence between possible models and consequently, better confidence that the chosen model is the best fitting and not overly simplistic (Ram & Grimm, 2009). A final model is selected based on a number of considerations relating to how well the model fits the data, the model's ability to classify individuals correctly into subgroups, class size and the clinical relevance of the different classes and their corresponding trajectories (Cuijpers et al., 2005; Uher et al., 2010; Wickrama et al., 2016). Parsimony is favoured in these models so the addition of an extra subgroup or more complex trajectory shape must add explanatory value to the previous model (Uher et al., 2010). Furthermore, models can be simplified whereby the variance within classes is assumed to be zero if convergence presents a significant issue, providing more flexible options to explore trajectory shape. This type of modelling is termed latent class growth analysis (LCGA).

There are multiple benefits to defining treatment response in depression post-hoc. Firstly, this approach still retains the categorical nature necessary to complement and inform clinical practice (Uher et al., 2010). However, it does so without assumptions on what score, or percentage decrease, will constitute a response, which is a major shortcoming of a priori definitions. Categories instead refer to favourable or unfavourable trajectories, rather than responders and non-responders (Gueorguieva et al., 2011). These are arguably more appropriate terms for patient populations where treatment mechanisms are not well understood. The information provided on trajectory shape can also be useful for clinicians to manage expectations and improve patient compliance and motivation (Stulz et al., 2010). In GMM, as categories are based on what naturally occurs in the data, categorization is less arbitrary, and subgroups of patients more homogenous (Thibodeau et al., 2015). When inter-individual differences are subtle, and the anticipated effect sizes are small, a categorization method that minimizes within-group heterogeneity would be advantageous.

*Growth mixture modelling to measure treatment efficacy in depression.*

A small number of studies have used growth mixture modelling to study depression symptom change over time (Rhebergen et al., 2012), and treatment effects in patients with major depression (Brière, Rohde, Stice, & Morizot, 2016; Cuijpers et al., 2005; Gueorguieva et al., 2011; Stulz et al., 2010; Thibodeau et al., 2015; Uher et al., 2010). A naturalistic study of 804 patients with depression modelled symptom trajectories over two years (Rhebergen et al., 2012). Employing a LCGA model, five distinct classes of patients emerged from their sample, each with a distinct linear trajectory over time. One group displayed a rapid improvement in depressive symptoms over time; two further groups showed a gradual improvement, differing only on baseline severity. Two smaller groupings represented a minority of the sample (20%), showing trajectories indicative of chronic and unchanging depression, again differing on severity levels. Interestingly, this study illustrated that diagnoses often associated with more chronic forms of depression (dysthymia and double depression) actually showed poor correspondence to the chronic classes identified empirically. These findings suggest that diagnostic criteria alone may be insufficient to explain the heterogeneity of patients with depression adequately. A unique aspect of this study was that it provided an insight into how depressive symptoms change over time in a clinical population, without direct effects of a treatment intervention. Consequently, it provided valuable, clinically relevant data on the characteristics of patients who may spontaneously respond without intervention.

Two studies investigating trajectories of symptom change in clinical trial data of antidepressant medications produced comparable findings (Gueorguieva et al., 2011; Uher et al., 2010). These studies had a strong research design with a good number of time-points. This rich amount of information allowed for the *shape* of symptom change over time to be interrogated. Encouragingly, both studies found symptom change for patients across the trial followed trajectories with a distinct break point marking two different rates of change. Moreover, across both studies, these piece-wise functions showed significant variation between individuals and further model testing revealed that two subclasses of patients best described the data, each with their own distinct trajectory. The findings from these empirical studies have provided support for the effectiveness of antidepressant medications. Firstly, treatment increased the likelihood of membership to a responding class (Gueorguieva et al.,

2011), but also, symptom reduction in the responding classes appeared evident and most rapid during the first 3 weeks of treatment (Gueorguieva et al., 2011; Uher et al., 2010).

Both studies mentioned above also agreed that classes showed a 3:1 percentage split of patients (Gueorguieva et al., 2011; Uher et al., 2010). Such imbalance in groupings may explain contradictory results seen in other studies, where categorical definitions fail to reach significance thresholds achieved by continuous outcomes (Emslie et al., 2002). However despite such strong agreement between the two datasets in response shape and number of subgroups, they differed markedly in the reported rates of change of the larger versus smaller groups. Uher and colleagues (2010) stated that their majority group was best defined as gradual responders, while their minority group appeared to respond rapidly, showing a sharp initial decrease in symptoms during the first three weeks of the trial. Conversely, Gueorguieva and colleagues (2011) stated that their groups were best described as a majority class of responders and a minority class of non-responders that failed to show a significant change in their trajectory over the 9-week trial.

Uher (2010) and Gueorguieva and colleagues' (2011) work provided good support for the notion that multiple subgroups of patients (as defined by the association between symptom change and treatment) exist within depression, however the lack of agreement in clinical interpretability implores for more empirical work to be conducted. In addition, both studies only investigated pharmacological intervention. While SSRIs are often prescribed as a first line of action, it is common and best recommended practice for them to be combined with a form of psychotherapy (NICE, 2015, NICE, 2016). Consequently, studies have begun to investigate symptom change trajectories across other treatment modalities, including pharmacological-psychological combination treatment. Two studies have investigated such combination treatments over 12 weeks, in two differing samples of depressed patients (Stulz et al., 2010; Thibodeau et al., 2015). These studies found corresponding results: that patients' symptom change over time appeared to group into three distinct trajectories. In addition, both studies found that baseline severity was a significant factor for group differentiation, and that the majority of patients improve gradually; only the smallest classes showed steeper declining slope values indicative of rapid response. However, both studies applied only a linear function to their models. As described above, investigation of non-linear trajectories would more accurately define trajectory shape in depressive symptom

change (Gueorguieva et al., 2011; Uher et al., 2010), and indeed, evidence from the original trial data for one study (Stulz et al., 2010) supported the presence of non-linear trends (Keller et al., 2000). Restricting models to a simplistic function may bias the models to favour more classes in order to represent true nonlinear functions as linear trends ("Topic 6: Mplus Short Course Videos and Handouts," n.d.; Wickrama et al., 2016) . Consequently, without a thorough investigation of trajectory shape, interpretation of these results is limited.

One strength of one of the above studies mentioned here was that the treatment modality of Thibodeau and colleagues' (2015) research included combination treatments of two antidepressants and three psychotherapies that are commonly prescribed in clinical settings. However, to date, very little research has been conducted using GMM in trials of solely psychotherapy treatments. Only one study known to the current author has investigated this in a clinical population (Cuijpers et al., 2005). Cuijpers and colleagues (2005) randomized patients to either CBT, or treatment as usual (TAU), which consisted of a tailored therapy to the individual's needs. Four trajectory classes emerged from their analysis, distinguished by baseline severity and differential slope values depending on treatment allocation. The two classes with the lowest baseline severity contained the majority of patients. A unique strength of this study was that the authors collected data every 3 months for 1.5 years. Such detailed investigation of long-term trajectories is currently under-studied in this particular research field. Indeed, authors have argued that trends that emerge after 12 weeks cannot be assumed to continue in the same way; the duration of follow-up plays a contributory role in determining response classes (Brière et al., 2016; Thibodeau et al., 2015). Therefore, short-term clinical trials at best provide an incomplete picture of response patterns in depression.

*Impact on clinical trial interpretations*

One of the main advantages to the employment of GMM in clinical trials is the ability to investigate differential treatment effects between subgroups of patients with depression. A number of studies have found that treatment increases the likelihood of patient membership to favourable trajectory classes (Brière et al., 2016; Gueorguieva et al., 2011). Other studies have taken this a step further and demonstrated that certain treatments are more beneficial for particular classes of patients (Cuijpers et al., 2005; Stulz et al., 2010). Stulz and colleagues(2010) found that only within the moderately depressed patient group

was there a significantly benefit of combination treatment compared to monotherapy. The size of this subgroup was likely driving the effect seen in the original clinical trial for the superiority of combination treatment (Keller et al., 2000; Stulz et al., 2010). This study highlights the limitations of end-point analyses. Cautious interpretation must be taken when generalizing group means; the effect may not be true for all patients.

A more detailed clarification of clinical trial results by empirical approaches has led to the question of validity for clinical trial conclusions that dichotomise patients without consideration of trajectory over time. For instance, Uher and colleagues (2010) reported opposing results depending on their choice of categorization method. When patients were classified using GMM, patients in more favourable trajectory classes were more likely to be randomized to nortriptyline than escitalopram, even following sensitivity analyses to account for classification uncertainty. However, when categorizing based on percentage reduction, patients treated with escitalopram were more likely to be classified as a responder. Gueorguieva and colleagues (2011) suggested that the extent of misclassification could be as high as 35%, erroneously classifying good responding patients as false negatives to treatment. Theoretical definitions may therefore be too strict (Gueorguieva et al., 2011) and the choice of methodology to determine outcome is particularly important if homogeneity is a goal for revealing the best group of non-responding individuals.

Some studies have reported that not only can treatment influence trajectory membership, but also trajectory course. Cuijpers and colleagues (2005) fitted separate GMMs to each treatment group and, while finding that the number and division of patients between their 4 trajectories were equivalent for both treatments, the shapes of those trajectories were significantly different. Patients in the highest depressive symptomatology class receiving TAU had no significant improvement in their symptoms over 18 months, while those treated with CBT experienced a clinically significant decline. Furthermore, the class with the second highest baseline symptomatology experienced a more rapid reduction in symptom severity with CBT. However, it is not typical for studies to divide their sample into treatment groups. GMMs are large sample techniques. False impressions of poor model fit could occur when the smaller samples don't contain enough variance to converge properly (Ram & Grimm, 2009; Wickrama et al., 2016). Consequently, very few studies have conducted separate trajectory analyses for treatment groups, with one other known to the current author and

indeed, those authors did not find treatment to affect the nature of trajectory, but only trajectory membership (Brière et al., 2016).

*Moving forwards*

To date, the majority of studies that have implemented empirical techniques such as GMM in MDD have been in adults (Cuijpers et al., 2005; Gueorguieva et al., 2011; Rhebergen et al., 2012; Stulz et al., 2010; Thibodeau et al., 2015; Uher et al., 2010), and trial data has been of primarily pharmacological (Gueorguieva et al., 2011; Uher et al., 2010), or combination interventions (Stulz et al., 2010; Thibodeau et al., 2015). Regardless of treatment modality, all studies supported the presence of multiple, qualitatively distinct classes of symptom trajectories over a single population, not necessarily distinguishable on baseline severity alone. However, studies disagreed on their interpretation of trajectory shape and nomenclature. While psychotherapies are beginning to be investigated (Brière et al., 2016; Cuijpers et al., 2005), there is insufficient research in this treatment sector and a major shortcoming of the current field is the lack of longer-term follow-up trials. However, perhaps the biggest gap in the literature resides in the lack of studies using GMM in clinical samples of adolescents, particularly considering the importance of this developmental group in depression outline in the introduction to this thesis (p8). While treatment effects of clinical depression in adolescence have been investigated, all studies (with one exception (Scott, Lewis, & Marti, 2019)), have used traditional approaches to define treatment response (Goodyer et al., 2008; March et al., 2009).

GMM however, has been used to describe the development of depressive symptomatology in general populations of adolescents in a number of studies (Brendgen, Wanner, Morin, & Vitaro, 2005; Brière, Janosz, Fallu, & Morizot, 2015; Costello, Swendsen, Rose, & Dierker, 2008; Wickrama & Wickrama, 2010). In the absence of intervention, all studies found support for the existence of qualitatively distinct trajectories of symptom change over time in adolescence. These trajectories typically included a group with persistently high symptoms, a group with persistently low symptoms, and a group of increasing symptoms. Brière and colleagues (2016) extended this work in a prevention trial, investigating how a cognitive-behavioural preventative intervention related to trajectories of subclinical depressive symptoms in adolescence. They found that 4 distinct trajectories were present in their adolescent sample over their two-year study, which agrees with naturalistic study

findings (Brendgen et al., 2005; Costello et al., 2008; Wickrama & Wickrama, 2010). Two trajectories possessed similar traits to those reported in clinical adult samples (Stulz et al., 2010; Thibodeau et al., 2015; Uher et al., 2010); they accounted for the majority of the sample and followed a gradual declining trajectory, differing on baseline severity. One group also showed persistent symptoms, which agrees with some adult literature (Cuijpers et al., 2005; Gueorguieva et al., 2011; Thibodeau et al., 2015), and general population adolescent literature (Brendgen et al., 2005; Brière et al., 2015; Wickrama & Wickrama, 2010). However, the final grouping markedly differed from those trajectories reported in adult samples. It was characterized by a strong initial decline in symptoms up to 6 months; the most rapid of all 4 trajectories, however these patients appeared to relapse, experiencing a spike in symptom severity at 12 months that persisted for the remainder of the trial. It is possible this relapsing class was able to emerge due to the length of follow-up, highlighting the danger that can occur when interpreting data from short-term trials.

It would be of interest to establish whether the trajectories highlighted in Briere and colleagues' (2016) paper can be replicated in a clinical sample of adolescents receiving specialist care. Only one study, published this year, has employed such methods in a treatment trial of adolescent depression (Scott et al., 2019). In a re-analysis of the TADS data (March et al., 2009), Scott and colleagues (2019) found that over a 12 week period, three unique trajectory classes were present in their adolescent sample receiving antidepressant, psychological or combination treatment. Two classes mirrored the rapid (minority; 9.2%) and gradual (majority; 75.3%) responder classes discussed in a number of adult trials (Thibodeau et al., 2015; Uher et al., 2010), and a further class showed limited improvement over the trial (15.5% of their sample). In agreement with adult trials, patients allocated to pharmacological treatment were more likely to follow a favourable trajectory of improvement (Gueorguieva et al., 2011).

Unlike Briere and colleagues(2016), no relapsing class emerged in the TADS re-analysis (Scott et al., 2019). However, Briere and colleagues' (2016) relapsing class only emerged after a 6-month period. It is possible that the short duration of the TADS study precluded the emergence of such a class, or indeed, may have even misrepresented some patients. As previously discussed, duration of trial has a significant impact on the emergent classes and

their interpretation (Brière et al., 2016; Thibodeau et al., 2015), therefore future work would benefit from extending these findings in adolescent depression, with longer-term follow up.

The number of time points used in the TADS trial was also a significant limitation. Only three time points meant that only linear trajectories could be investigated. A fuller investigation of alternative trajectory shapes over this developmental age could provide further insight into the dynamics of symptom change during and after treatment for adolescent patients.

*Objectives and Hypotheses*

Consequently, the primary objective of this first piece of work was to conduct a secondary analysis of the IMPACT trial (Goodyer et al., 2011), to investigate trajectories of depressive symptom change in a clinical sample of adolescents, from randomisation to the final assessment one year following end of treatment. The specific aims were to:

1. define the number of longitudinal classes revealed from depression symptoms only and describe the shape of the trajectories for each group;
2. compare the defined groups with standard priori definitions of response/non-response.

Prior literature would suggest that 4 classes could emerge from the data, with favourable trajectories showing rapid and gradual improvement in symptoms(Uher et al., 2010), and unfavourable trajectories showing either no improvement, or a relapsing trajectory shape (Brière et al., 2016).

Chapter 2 is dedicated to describing the demographic and clinical characteristics of the emerging classes from this first piece of work.

## Methods

### *Study Design*

This study was a re-analysis of the IMPACT trial (Goodyer et al., 2011). The IMPACT study was a multicentre, pragmatic, observer-blind, randomised controlled trial investigating whether there was a superior effect for two specialist psychological treatments (cognitive behavioural therapy; CBT, short-term psychoanalytic psychotherapy; STPP) compared with a reference treatment of brief psychosocial intervention (BPI) on reducing self-reported depressive symptoms by end of follow up 12 months after end of treatment (Goodyer et al., 2017). Participants were randomly assigned to one of the three treatment arms, with stochastic minimisation by age, sex, self-reported depression sum score, and region, as per study protocol (Goodyer et al., 2011). The primary findings from the trial demonstrated no differences in depression symptoms sum scores between treatment groups over the course and by the end of the study (Goodyer et al., 2017). Consequently, treatment group was collapsed for the present study, to investigate the symptom trajectories in this whole population. Self-reported depressive symptomatology was measured at 6 nominal time points: baseline, 6, 12, 36, 52 and 86 weeks post-randomisation. The last 2 time points were post-treatment; treatment was completed by 36 weeks in >95% of the cohort (Goodyer et al., 2017). Based on each individual's sum score symptom change over the trial, a series of growth mixture models were conducted to determine the best fitting model, and the number of classes of individuals present within the dataset.

### *Setting*

The IMPACT trial (Goodyer et al., 2011) recruited patients from 15 National Health Service child and adolescent mental health service (CAMHS) clinics across 3 geographical regions in the UK: East Anglia, North London and North-West England covering an estimated 1,000,000 adolescents aged 11-17 years. The study recruited, assessed and followed up all participants between June 29, 2010 and Jan 17, 2013 (Goodyer et al., 2017).

### *Participants*

Adolescents aged between 11 and 17 years, with a current diagnosis of major depression (DSM-IV(APA, 2000)) with moderate to severe impairment were enrolled in the IMPACT trial. Exclusion criteria included generalised learning difficulties or a Pervasive Developmental Disorder, pregnancy, substance abuse, or a primary diagnosis of bipolar type 1,

schizophrenia or an eating disorder. Patients were also excluded if they were currently taking medication that would interact with an SSRI and were unable to stop this medication. Full details of the study protocol can be found in the study protocol (Goodyer et al., 2011). Patients were randomized to CBT, STPP or the reference treatment of BPI, as per study protocol.

*Variables*

Symptom trajectory class membership was defined through GMM (see below) using the self-reported Mood and Feelings Questionnaire (MFQ) score across all time-points. This is a 33-item Questionnaire of depressive symptomatology covering the past 2 weeks (Burleson Daviss et al., 2006). MFQ items were measured on a 3-point scale (almost never, sometimes, often/almost always). Total scores (range of 0-66) were used in GMMs. Higher scores indicated more severe depressive symptoms and were positively correlated with greater psychosocial impairment (Goodyer et al., 2017).

*Bias*

The recruitment sites were dependent on referrals from primary care sources including family physicians, community mental health teams and self-referral. Clinics who participated were invited and not selected randomly. Therefore, we cannot be certain that the sample is necessarily representative of major depression in the adolescent population at large nor of cases usually referred to child and adolescent mental health clinics in the UK. However, there are no other referral options for the primary care services other than their local NHS services and therefore the clinics are likely to be receiving the majority of referrals for major depression. Finally these clinics are part of routine NHS mental health services and not set up solely for the purposes of the IMPACT study.

*Study size*

Of 557 participants screened for eligibility into the IMPACT trial, 87 were excluded (73 did not meet criteria for major depression, 4 had mania as their primary diagnosis, 4 had a primary substance use disorder, 2 had received previous treatment used in the trial, 1 had autism, 1 was pregnant, 1 would not engage and 1 was unable to read or understand information). 470 adolescents were therefore randomised to 3 treatment arms.

Subsequently, 5 withdrew consent to use their data, leaving 465 patients included and in the current analysis (Goodyer et al., 2017).

*Statistical analyses*

*Missing Data*

Multiple imputation was the chosen method to deal with cases of missing data, in order to maximise sample size and achieve convergence for the GMM. Eleonore van Sprang (visiting Masters student) and Sharon Neufeld (Research Associate employed in the Department of Psychiatry, University of Cambridge) created the imputed dataset. The dataset was made available in the IMPACT trial data. A description of the imputation method written by van Sprang and Neufeld can be found in the publication manuscript of this chapter, currently submitted for peer review (See Supplement). It states:

*"Despite excellent follow-up rates (80% at 86-weeks(Goodyer et al., 2017)), multiple imputation was required in order to maximize sample size and achieve convergence for the GMM. Due to a wealth of auxiliary variables predicting missingness, data was [sic] presumed to be missing at random. Also due to these auxiliary variables, multiple imputation was favoured over Full Information Maximum Likelihood as auxiliary variables can easily be incorporated into a multiple imputation model and help decrease bias and increase efficiency (Graham, 2003). Variables at all time points were assessed for inclusion in the imputation model in addition to MFQ items. Those related to outcome ($p<0.05$ or $r>=0.3$) and/or missingness in outcome, and variables used in final analyses, were included in the model (White, Royston, & Wood, 2011). Additional non-missing variables were also included to improve model prediction. This resulted in imputation of 24 variables plus the 33 MFQ items, repeated over six assessments, yielding a dataset too large to impute in wide format. Thus, time-varying data was imputed in long format, a method which is less biased under conditions of less missing data, more repeated measures, and a reliable outcome measure (Gottfredson, Sterba, & Jackson, 2017), as is the case in the present data. For each model, fifty datasets were multiply imputed using chained equations (White et al., 2011). As it is not possible to obtain the VLMR and LMR fit statistics for model comparison in a multiply imputed dataset, multiple imputations were averaged prior to estimation of GMM. While we acknowledge it is more optimal to obtain model estimates from each of the multiply imputed*

*datasets and then combine estimates (White et al., 2011), our approach allowed us to obtain these fit statistics which are crucial for determining the most optimal model."*

*Latent growth curve modelling (LGCM)*

Latent growth curve modelling (LGCM) is a type of modelling that allows for the analysis of the change in MFQ score over time, which in our case is the duration of the IMPACT trial. For growth curve modelling to be viable, the data must suggest that significant variability exists in slope values, and that the slope variance is non-negative. This means that the covariance between adjacent time points must be higher than the covariance between non-adjacent time points (Wickrama et al., 2016). Consequently, the longitudinal correlation patterns of the repeated measures of MFQ score were investigated, to examine the feasibility of estimating growth curves. Results are shown in Table 1.

*Table 1. Correlation matrix among MFQ scores included in the model*

| **Assessment** | *0 (baseline)* | *1 (6 weeks)* | *2 (12 weeks)* | *3 (36 weeks)* | *4 (52 weeks)* | *5 (86 weeks)* |
|---|---|---|---|---|---|---|
| *0 (baseline)* | 1.000 | | | | | |
| *1 (6 weeks)* | .463 | 1.000 | | | | |
| *2 (12 weeks)* | .344 | .609 | 1.000 | | | |
| *3 (36 weeks)* | .307 | .414 | .554 | 1.000 | | |
| *4 (52 weeks)* | .224 | .294 | .416 | .659 | 1.000 | |
| *5 (86 weeks)* | .221 | .272 | .335 | .535 | .613 | 1.000 |

*MFQ= Mood and Feelings Questionnaire. All correlations were significant at p<.001.*

The observed correlation between two adjacent occasions equal .463, .609, .554, .659, and .613. The adjacent correlations are higher than correlations between two non-adjacent time points. This presents preliminary evidence for a non-negative variance of slope parameter. As such, a LGCM will likely fit well with the data structure.

*Step 1: Specify a traditional latent growth curve model (LGCM).*

The trajectory of change in depressive symptoms over the trial was modelled on imputed data from all 465 individuals in the original IMPACT trial. We used MFQ scores to estimate

the latent growth curve, at planned assessment intervals of 0, 6, 12, 36, 52 and 86 weeks post-randomisation. MFQ was treated as a continuous variable. As an aim of this research was to test whether age and gender were predictive of class membership, this information was not included in the models.

The first stage of the analysis began with a growth curve model with one latent class: a conventional latent growth curve model (LGCM). A LGCM describes the average course of depressive symptoms using continuous latent growth factors of intercept, linear and quadratic slopes. The intercept describes the level of depressive symptoms at baseline, a linear slope describes how the symptoms change over time and a quadratic factor describes any curvature in that slope. It is necessary to identify a common mean growth curve to determine the model fit of a single class model. Then successively more classes will be added to determine the best fitting model. LGCMs with linear and quadratic trends over time were considered and tested in the Mplus program version 8.0 (Muthén & Muthén, 2017). We also hypothesised that a change in symptom trajectory might occur when treatment ended for patients. While treatment cessation was planned to end around the fourth assessment for patients (nominally 36 weeks), the average length of treatment was 27 weeks in the IMPACT trial. This falls between the mean of the third assessment (18 weeks from baseline) and the mean of the fourth assessment (43 weeks from baseline; see Table 2). Therefore, two piecewise models with one transition point were considered, one with the transition point at the third assessment, and one at the fourth. These models expressed the separate growth curves as two linear trends. Collectively, the four models tested allowed for a variety of straight and curved trajectories with up to one sharp transition point to capture different rates of improvement at different stages of the trial.

*Time consideration*

The means, standard deviations (SD), medians and inter-quartile ranges for assessment timings are presented in Table 2. As is clear from these values, the timing that each assessment took place varied substantially across individuals. Consequently, three models were tested to assess the degree to which variation in time of assessment would affect model fit. In the first unconditional models, time-points were fixed at the mean time of that assessment from the whole sample. Then two further series of models were tested. The first included time of assessment as a covariate in the models. Finally, a third series of models

were run which allowed for assessment time-point to individually vary across individuals (TSCORES). The rationale for testing both methods of controlling for time was because the option of TSCORES makes models substantially more complex, which can affect model convergence in later stages of GMM.

Maximum Likelihood (Robust) was used as the model estimator for all models. This was chosen because it is robust against deviations of multivariate normality. A large number of start values (5,000 with 100 optimisations) were used to avoid solutions of local maxima (Brière et al., 2016; Hipp & Bauer, 2006).

*Table 2. Descriptive statistics for desired and actual time-point of each assessment*

| **Desired assessment time-point (weeks)** | **Actual assessment time-point (weeks)** | | | | | |
|---|---|---|---|---|---|---|
| | *Mean* | *SD* | *Median* | *IQR* | *Min* | *Max* |
| *Baseline* | 0 | 0 | 0 | 0 | 0 | 0 |
| *6 weeks* | 12.0 | 3.86 | 11.1 | 3.9 | 6.1 | 41.0 |
| *12 weeks* | 18.4 | 4.04 | 17.7 | 4.0 | 11.6 | 38.0 |
| *36 weeks* | 42.7 | 4.09 | 42.0 | 4.2 | 30.6 | 63.1 |
| *52 weeks* | 60.1 | 5.02 | 59.0 | 5.1 | 48.3 | 92.1 |
| *86 weeks* | 95.3 | 9.18 | 93.6 | 6.7 | 69.0 | 149.0 |

*Step 1: Results*

Tables 3, 4 and 5 show the model fit information for the unconditional, conditional and time-varying LGCMs, respectively. Table 6 denotes all intercepts, linear and quadratic trajectories for each model. As TSCORES did not prove to provide a substantially better fit of the data, we took forward for discussion only models where time was included as a covariate (Table 4). Assessment time-points will herein be referred to as the average number of weeks that that assessment took place: 0, 12, 18, 43, 60 and 95 weeks, post-randomisation.

As can been seen in Table 4, and observed in the graphical representation of Figure 1, both quadratic and linear-piecewise models provided a better representation of the data than a simple linear trajectory, suggesting symptom change was not a simple linear function. A piecewise model that separately modelled the change in depressive symptoms, linearly, over the first 18 weeks of treatment (on average; assessments 0-2), and then the remaining period of the trial (assessments 3-5), was identified as the optimal function for the common

mean growth curve of depressive symptoms over the trial ($X^2(37)= 66.570$, p=.002, CFI/TLI=.964/.956, RMSEA=.041 and SRMR=.037, BIC=21098.832, AIC=21015.992).

*Table 3: Fit indices for unconditional LCGMs.*

| **Model Fit Information** | **Linear** | **Quadratic** | **Piecewise(43 week break)** | **Piecewise(18 week break)** |
|---|---|---|---|---|
| ***Information Criteria*** | | | | |
| *Akaike (AIC)* | 21445.476 | 21151.320 | 21213.256 | 21045.437 |
| *Bayesian (BIC)* | 21491.038 | 21213.450 | 21275.386 | 21107.568 |
| ***Chi-square Test of Model Fit*** | | | | |
| *Value* | 425.078 | 176.219 | 230.755 | 80.244 |
| *Degrees of Freedom* | 16 | 12 | 12 | 12 |
| *P-Value* | <.001 | <.001 | <.001 | <.001 |
| ***RMSEA (Root Mean Square of Approximation)*** | | | | |
| *Estimate* | 0.234 | 0.172 | 0.198 | 0.111 |
| *90 Percent C.I.* | N/A | N/A | N/A | N/A |
| *Probability RMSEA <=.05* | N/A | N/A | N/A | N/A |
| ***CFI/TLI*** | | | | |
| *CFI* | 0.475 | 0.789 | 0.719 | 0.912 |
| *TLI* | 0.508 | 0.737 | 0.649 | 0.891 |
| ***SRMR (Standardised Root Square Residual)*** | | | | |
| *Value* | 0.064 | 0.048 | 0.051 | 0.039 |

*CFI: Comparative Fit Index; TLI: Tucker-Lewis Index*

*Table 4: Fit indices for conditional LCGMs.*

| Model Fit Information | Linear | Quadratic | Piecewise(43 week break) | Piecewise(18 week break) |
|---|---|---|---|---|
| ***Information Criteria*** | | | | |
| *Akaike (AIC)* | 21125.041 | 21024.495 | 21027.396 | 21015.992 |
| *Bayesian (BIC)* | 21191.314 | 21107.335 | 21110.237 | 21098.832 |
| ***Chi-square Test of Model Fit*** | | | | |
| *Value* | 164.718 | 73.620 | 75.929 | 66.570 |
| *Degrees of Freedom* | 41 | 37 | 37 | 37 |
| *P-Value* | <.001 | <.001 | <.001 | .002 |
| ***RMSEA (Root Mean Square of Approximation)*** | | | | |
| *Estimate* | 0.081 | 0.046 | 0.048 | 0.041 |
| *90 Percent C.I.* | N/A | N/A | N/A | N/A |
| *Probability RMSEA <=.05* | N/A | N/A | N/A | N/A |
| ***CFI/TLI*** | | | | |
| *CFI* | 0.849 | 0.955 | 0.952 | 0.964 |
| *TLI* | 0.834 | 0.946 | 0.942 | 0.956 |
| ***SRMR (Standardised Root Square Residual)*** | | | | |
| *Value* | 0.063 | 0.035 | 0.037 | 0.037 |

*CFI: Comparative Fit Index; TLI: Tucker-Lewis Index*

*Table 5: Fit indices for time-varying LCGMs.*

| Model Fit Information | Linear | Quadratic | Piecewise(43 week break) | Piecewise(18 week break) |
|---|---|---|---|---|
| ***Information Criteria*** | | | | |
| *Log-likelihood* | -10716.420 | -10572.385 | -10596.524 | -10541.635 |
| *Akaike (AIC)* | 21454.840 | 21174.769 | 21223.048 | 21113.269 |
| *Bayesian (BIC)* | 21500.403 | 21236.900 | 21285.178 | 21175.400 |

*Table 6: Intercept and Slope values for each Model.*

| | Intercept | Slope1 | Quadratic1 | Slope2 |
|---|---|---|---|---|
| **Linear** | | | | |
| *Unconditional* | 38.219*** | -1.887*** | | |
| *Conditional* | 45.538*** | -3.686*** | | |
| *TSCORES* | 38.146*** | -1.872*** | | |
| **Quadratic** | | | | |
| *Unconditional* | 43.276*** | -5.073*** | 0.302*** | |
| *Conditional* | 45.913*** | -7.676*** | 0.590*** | |
| *TSCORES* | 42.964*** | -4.915*** | 0.290*** | |
| **Piecewise(43 week break)** | | | | |
| *Unconditional* | 41.892*** | -3.433*** | | -0.471** |
| *Conditional* | 45.809*** | -5.973*** | | 0.230*** |
| *TSCORES* | 41.820*** | -3.458*** | | -0.470** |
| **Piecewise(18 week break)** | | | | |
| *Unconditional* | 45.650*** | -7.328*** | | -1.101*** |
| *Conditional* | 45.941*** | -6.982*** | | -1.088*** |
| *TSCORES* | 45.091*** | -7.004*** | | -1.111*** |

**Variances were significant at p<.05*

***Variances significant at p<.01*

****Variances significant at p<.001*

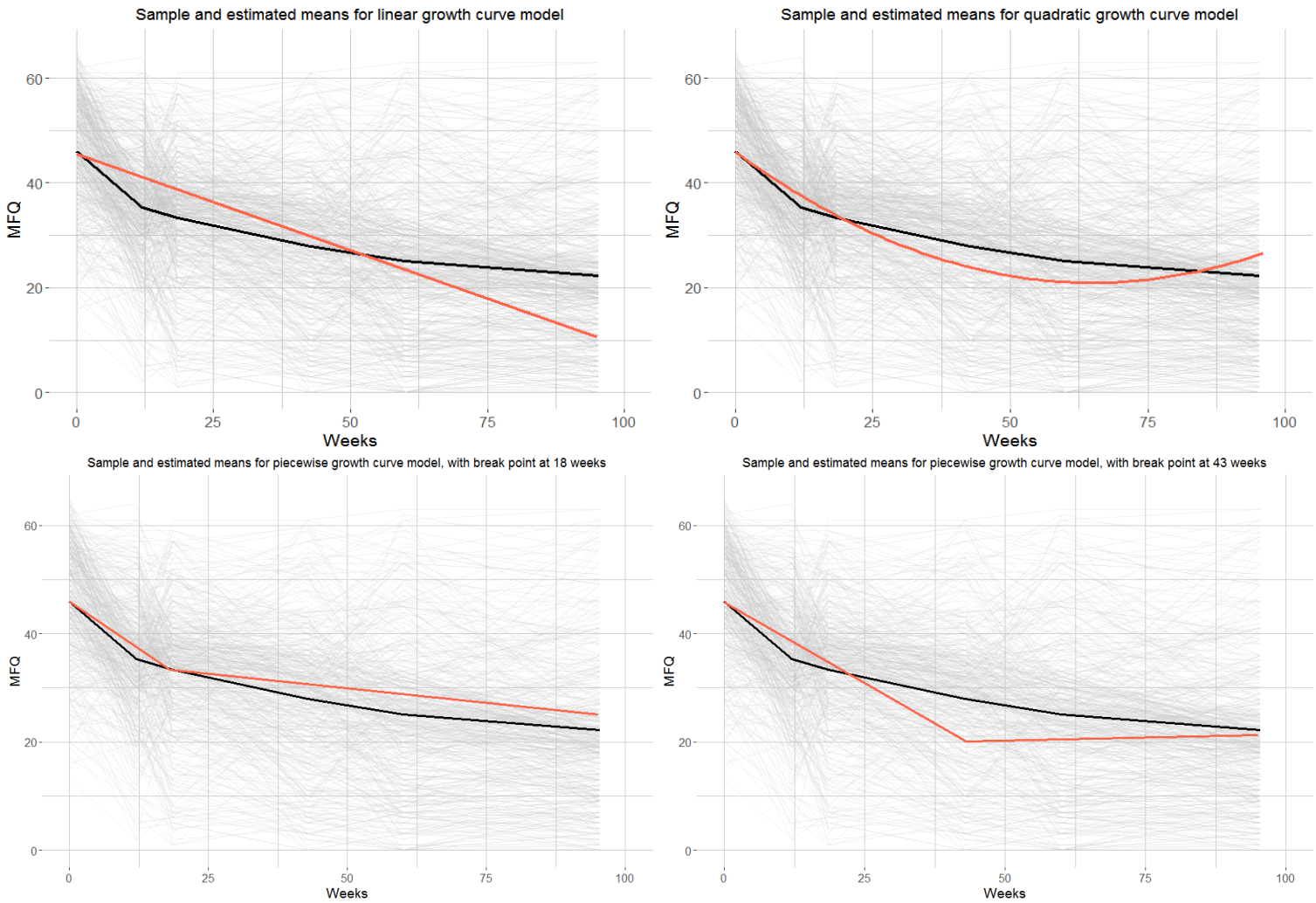



*Figure 1: Competing LGCMs. The black line represents the sample means, while the red line indicates the estimated means by model fit. Behind plots every individual patient's trajectory over time.*

Table 7 shows the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, along with the and AIC and BIC weights of all competing models (Wagenmakers & Farrell, 2004). Comparisons of the AIC and BIC weights allow these values to be interpreted as conditional probabilities (Wagenmakers & Farrell, 2004). This therefore allows us to judge the statistical importance of the differences observed in these values (Wagenmakers & Farrell, 2004). Based on the AIC, the piecewise model described above shows a 98.3% probability that it is the best model. It can further be shown that this model is 75.6 times more likely to be the best model in terms of Kullback-Leibler discrepancy than the next best model (calculated from dividing the AIC weight of best model by the next best model; (Wagenmakers & Farrell, 2004)). Based on the BIC, the piecewise model described above shows a 97.1% probability that it is the best model. It can further be shown that this model is 69.4 times more likely to be the best model in terms of Kullback-Leibler discrepancy than the next best model. Consequently, we can be confident that this is the most optimal trajectory shape.

On inspection of the optimal model (linear-piecewise model with a break point at 18-weeks), we can see that the average intercept, and linear trajectory for the first period were statistically significant (45.941, $p<.001$ and -6.982, $p<.001$ respectively), meaning that the trajectory of depressive symptoms decreased over the first 18 weeks of the trial. The linear trajectory of the remaining period was insignificant (-1.088, $p= .140$) suggesting that depressive symptoms over this part of the trajectory showed no average change and were stable over the follow-up period of the trial.

The results from this model further indicated that there were significant inter-individual differences in all growth factors (intercept: 80.737, $p<.001$; first linear trajectory: 30.642, $p<.001$; and second linear trajectory: 1.557, $p<.001$). This suggests variation exists around baseline levels of depression and rates of change in depressive symptoms over both sections of the model. In addition, a negative covariance was present between the intercept and first linear trajectory (-21.533, $p=.014$), indicating that on average, individuals with a high baseline level of depression were more likely to experience declining symptoms over time compared to other individuals. A negative covariance was also present between the first and second linear trajectories (-2.039, $p=.003$) suggesting that on average, individuals with a steeper decline in their depressive symptoms over the first 18 weeks of the trial were less

likely to have a steep decline in the second part of the trial. No other covariances were statistically significant.

In conclusion, the data suggest that a single class solution provides a good representation of the data, but significant variation exists on all aspects of the trajectory. Consequently, multiple class models were tested to determine the optimal number of classes for this dataset.

*Table 7: AIC and BIC values and weights for all competing models.*

| Model | No. Parameters. | Log-likelihood | AIC | AIC difference | AIC weights | BIC | BIC difference | BIC weights |
|---|---|---|---|---|---|---|---|---|
| **Linear** | | | | | | | | |
| *Unconditional* | 11 | -10711.738 | 21445.476 | 429.484 | <.001 | 21491.038 | 302.206 | <.001 |
| *Conditional* | 16 | -10546.521 | 21125.041 | 109.049 | <.001 | 21191.314 | 92.482 | <.001 |
| *TSCORES* | 11 | -10716.420 | 21454.840 | 438.848 | <.001 | 21500.403 | 401.571 | <.001 |
| **Quadratic** | | | | | | | | |
| *Unconditional* | 15 | -10560.660 | 21151.320 | 135.328 | <.001 | 21213.450 | 114.618 | <.001 |
| *Conditional* | 20 | -10492.247 | 21024.495 | 8.503 | .013 | 21107.335 | 8.503 | .014 |
| *TSCORES* | 15 | -10572.385 | 21174.769 | 158.777 | <.001 | 21236.900 | 138.068 | <.001 |
| **Piecewise(43 week break)** | | | | | | | | |
| *Unconditional* | 15 | -10591.628 | 21213.256 | 197.264 | <.001 | 21275.386 | 176.554 | <.001 |
| *Conditional* | 20 | -10493.698 | 21027.396 | 11.404 | .003 | 21110.237 | 11.405 | .003 |
| *TSCORES* | 15 | -10596.524 | 21223.048 | 207.056 | <.001 | 21285.178 | 186.346 | <.001 |
| **Piecewise(18 week break)** | | | | | | | | |
| *Unconditional* | 15 | -10507.719 | 21045.437 | 29.445 | <.001 | 21107.568 | 8.736 | .012 |
| *Conditional* | 20 | -10487.996 | 21015.992 | 0 | .983 | 21098.832 | 0 | .971 |
| *TSCORES* | 15 | -10541.635 | 21113.269 | 97.277 | <.001 | 21175.400 | 76.568 | <.001 |

*Step 2: Specify a Growth Mixture Model (GMM).*

Analyses thus far suggested that a single class solution provides a good representation of the data. In the second stage, multiple class models were considered to test whether a single class is the optimal number of classes for this dataset. We used GMM to identify distinct classes with different trajectories, allowing for within-class variation. For the best fitting class model, the variance between classes for each growth factor was tested individually. No evidence of significant variation was found (Table 1, Appendix 1A). This means that the variation around each group mean was not significantly different for the respective classes. This suggests that allowing between-class variation in growth factors would not improve the model fit, thus between-class variances were held equal across classes for all growth factors. Only solutions that were replicated with different starting values were accepted.

Classes were incrementally added to the single class model to determine the best fit. We considered models with 1 to 5 trajectory classes due to our a priori hypothesis; that we would expect 4 classes to emerge from the data. The objective was to select the most parsimonious model, based on a number of criteria. First, we examined information criteria to compare the relative fit of trajectory solutions. These included the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The model with the lowest AIC and BIC was retained, as low values indicate better model fit. BIC values were given priority in cases of discrepancy between indices. As BIC adjusts for sample size, balancing goodness of fit and parsimony, it often performs better (Nylund, Asparouhov, & Muthén, 2007). Models were only considered as favourable over another if the BIC difference was 10 or more (Uher et al., 2010). In addition to fit indices, the uncertainty of the model in the classification of subjects into the correct class was considered, and assessed using the entropy values. These range between 0 and 1 and indicate the probability of correct classification; 0 indicates everyone has an equal posterior probability of membership to all classes and 1 indicates that each individual has a posterior probability of 1 of membership to a single class, and probability of 0 to the remaining classes. As such, values closer to 1 are preferred. Finally, clinical interpretability and relevance of the class trajectories, as well as class size, were taken into account. Models where classes contained less than 10% of the sample were rejected as these were not considered stable numerically given our modest sample size (Rhebergen et al., 2012; Uher et al., 2010; Wickrama et al., 2016). Patients were assigned to their most likely class based on model probabilities.

*Step 2: Results*

The model fit information for all GMMs is shown in Table 8. No convergence problems were detected across the class models for the GMMs. Overall, it was decided that the best fit was achieved with a two-class solution, illustrated in Figure 2. The BIC showed a favourable decrease of approximately 42 with the addition of a second class from the single class solution. In addition, the two-class solution showed the best quality of classification of individuals into trajectory classes with an entropy of 0.844. The model estimates were stable across different sets of random starting values and all 100 final optimisations converged to the same solution.

*Table 8. Model fit information for GMMs*

| **Fit Statistics** | | **1 Class** | **2 Classes** | **3 Classes** | **4 Classes** | **5 Classes** |
|---|---|---|---|---|---|---|
| *Log-Likelihood* | | -10487.996 | -10454.836 | -10440.533 | -10431.253 | -10423.269 |
| *AIC* | | 21015.992 | 20957.672 | 20937.065 | 20926.506 | 20918.538 |
| *BIC* | | 21098.832 | 21057.081 | 21053.042 | 21059.051 | 21067.651 |
| *Entropy* | | 1 | .844 | .734 | .718 | .729 |
| *Group size (%)* | C1 | 465(100%) | 74(15.9%) | 329(70.7%) | 56(12.0%) | 109(23.5%) |
| | C2 | - | 391(84.1%) | 77(16.6%) | 161(34.6%) | 56(12.0%) |
| | C3 | - | - | 59(12.7%) | 57(12.3%) | 219(47.1%) |
| | C4 | - | - | - | 191(41.1%) | 54(11.6%) |
| | C5 | - | - | - | - | 27(5.8%) |

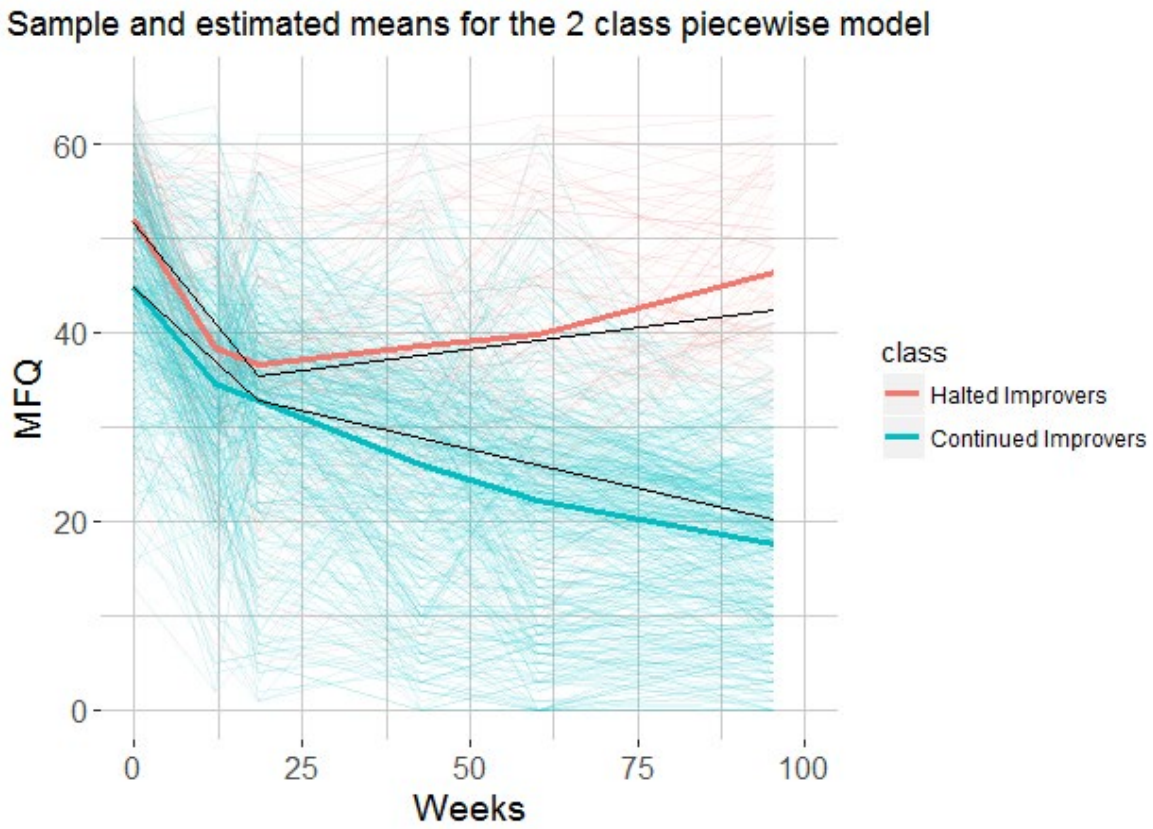



*Figure 2: Sample and estimated means for the 2-class piecewise growth mixture model. Behind plots every individual patient's trajectory over time, colour coded to his or her respective classes.*

The overlap between classes in their posterior probabilities of membership to class 1 and class 2 is shown in Figure 3. Values around 0.5 are where most uncertainty in class allocation lies. 17.6% of class 1 have a posterior probability below 0.75 of being allocated to class 1, and only 4.1% of class 2 have a posterior probability below 0.75 of being allocated to class 2. Overall, class 1 showed an 85.6% probability of correct membership in this class, and class 2 showed a 97.7% probability of correct class membership. This investigation into the posterior probabilities therefore provides additional support for the strength of this model. The Mplus code for the two-class model is provided in the Appendix 1B.

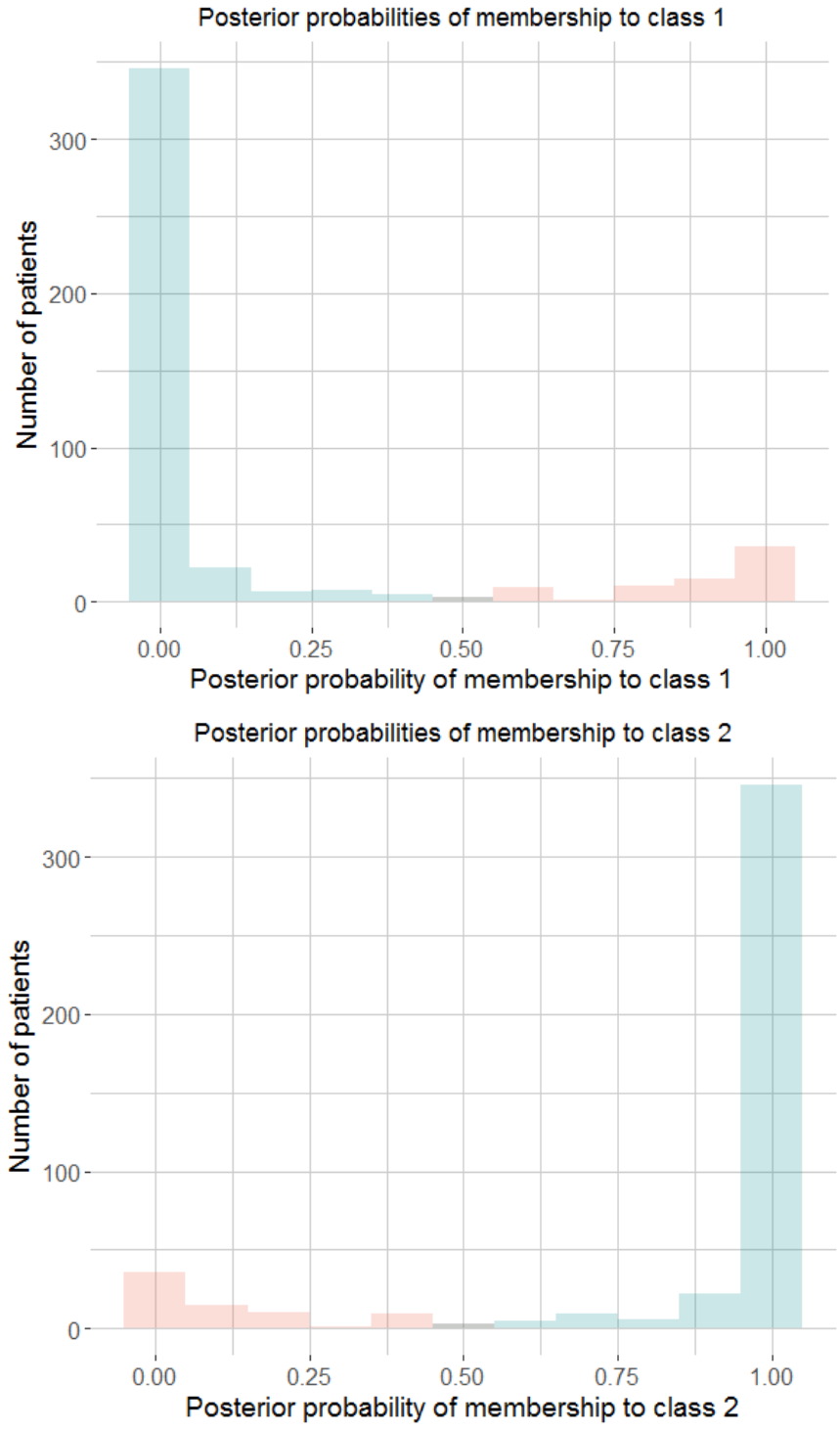



*Figure 3. Posterior probabilities of membership to the two-class model. Red bars indicate patients who were allocated to class 1, and blue bars indicate patients who were allocated to class 2.*

The two-class piecewise GMM divided subjects into a small class of 74 patients (class 1, 15.9% of patients), and a comparatively large class of 391 patients (class 2, 84.1% of patients). Class 1 on average showed a significantly higher baseline level of MFQ scores than class 2 (Wald $X^2(1)=25.577$, p<.001). Both class 1 and class 2 showed a significant decrease in MFQ score over the first 18 weeks of the trial (-8.794, p<.001, and -6.466, p<.001 respectively) however, class 1 showed significantly faster rate of MFQ reduction compared with class 2 (Wald $X^2(1)=5.446$, p=.0196). That is, the class who had a higher initial baseline MFQ improved faster.

However, these two classes appeared to depart more markedly from each other only after this 18-week break point. Class 1 showed a significantly faster rate of improvement over the first 18 weeks of the trial, then showed no further improvement and remained stable over the rest of the trial (0.899, p=.183). Class 2 on average, showed a significant continued, but slower decline in MFQ score than in the initial 18 weeks (-1.639, p=.014). This difference between the second linear slopes was statistically significant (Wald $X^2(1)=167.075$, p<.001).

The mean depression scores and the percentage change between time-points are shown in Table 9. By the end of the trial, Class 2 showed a 60.5% improvement in depressive symptoms, compared with 11.0% in class 1. Given the emerging trajectory shapes, we labelled patients in class 1 as “halted-improvers” and patients in class 2 as “continued-improvers”.

*Table 9. Estimated and observed mean values for MFQ scores and observed mean percentage improvement in MFQ scores for both latent classes.*

| | **Class 1: Halted-improvers (n=74)** | | | | **Class 2: Continual improvers (n=391)** | | | |
|---|---|---|---|---|---|---|---|---|
| | *MFQ scores* | | | | *MFQ scores* | | | |
| **Assessment Point in average weeks from baseline** | *Estimated* | *Weighted Estimates* | *Observed* | *% observed improvement from baseline* | *Estimated* | *Weighted Estimates* | *Observed* | *% observed improvement from baseline* |
| **0** | 51.638 | 51.721 | 52.096 | | 44.828 | 44.810 | 44.774 | |
| **12** | 41.090 | 38.191 | 38.420 | 26.252 | 37.073 | 34.649 | 34.623 | 22.672 |
| **18** | 35.390 | 36.134 | 36.558 | 29.826 | 32.881 | 32.752 | 32.689 | 26.991 |
| **43** | 39.232 | 38.243 | 38.669 | 25.774 | 25.877 | 23.938 | 25.920 | 42.109 |
| **60** | 40.789 | 39.034 | 39.835 | 23.535 | 23.039 | 22.291 | 22.224 | 50.364 |
| **95** | 43.966 | 44.978 | 46.357 | 11.016 | 17.247 | 17.797 | 17.673 | 60.528 |

*Consideration of the three-class model*

In the three-class piecewise GMM, a smaller third class separated out with relatively lower baseline depressive symptoms (MFQ=30.9). Furthermore, this class did not show a significant decrease in depressive symptoms in the first 18 weeks, unlike the other two classes (-2.467, p=.096), nor showed a significant change in the second part of the trial (-0.797, p=.278, Figure 4). While this class is of clinical interest, the addition of this third class did not meet the required reduction in BIC values to be regarded favourable, and the quality of classification suffered (.734). The overlap between classes in their posterior probabilities of membership to class 1, class 2 and class 3 are shown in Figure 5. While only 16% of class 1 had a posterior probability below 0.75 of being allocated to class 1 in the 3-class model, the posterior probabilities of classes 2 and 3 suffered. Twenty-three per cent of class 2 showed a posterior probability of less than 0.75 of being allocated to class 2, and 20% of class 3 showed a posterior probability of less than 0.75 of being allocated to class 3. Overall, class 1 showed a 93.7% probability of correct membership in this class, however class 2 showed only a 73.7% probability of correct membership, and class 3 a 79.7% probability of correct class membership. This investigation of the posterior probabilities highlights the weakness of the three-class model when compared with the two-class model.

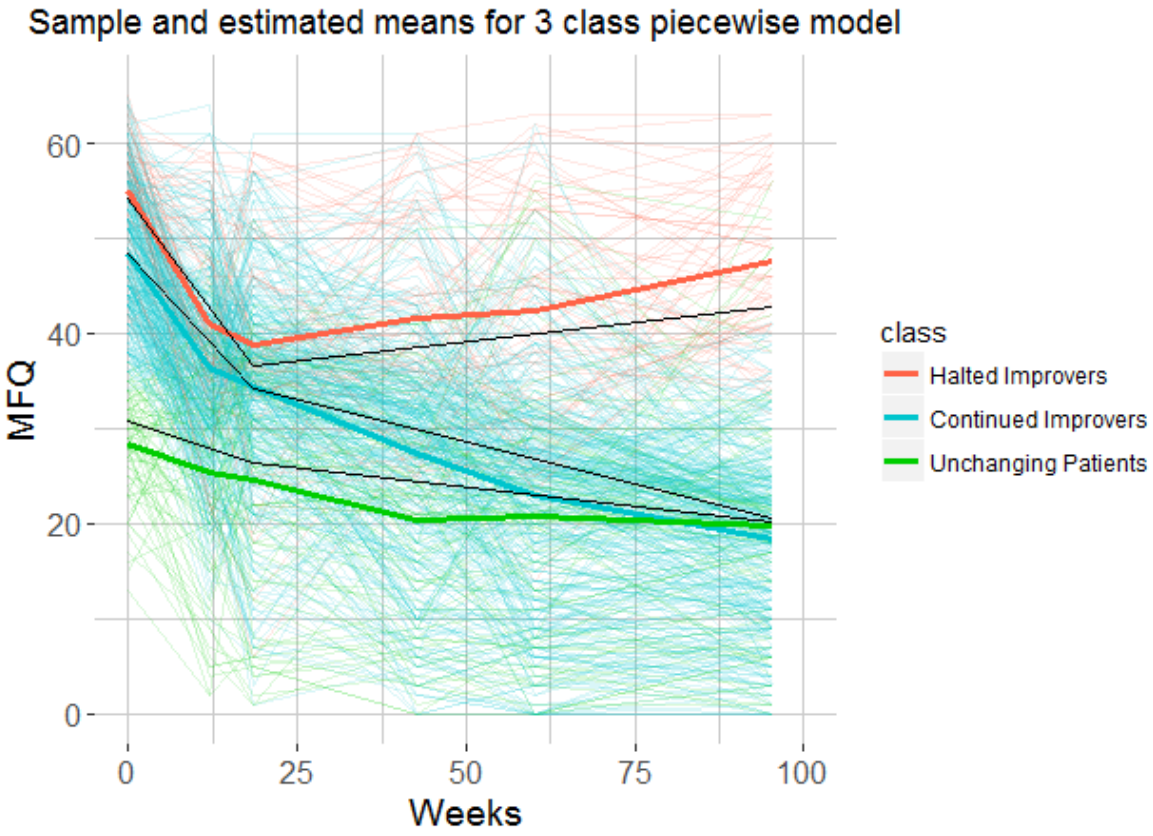



*Figure 4: Sample and estimated means for the 3-class piecewise growth mixture model. Behind plots every individual patient's trajectory over time, colour coded to his or her respective classes.*

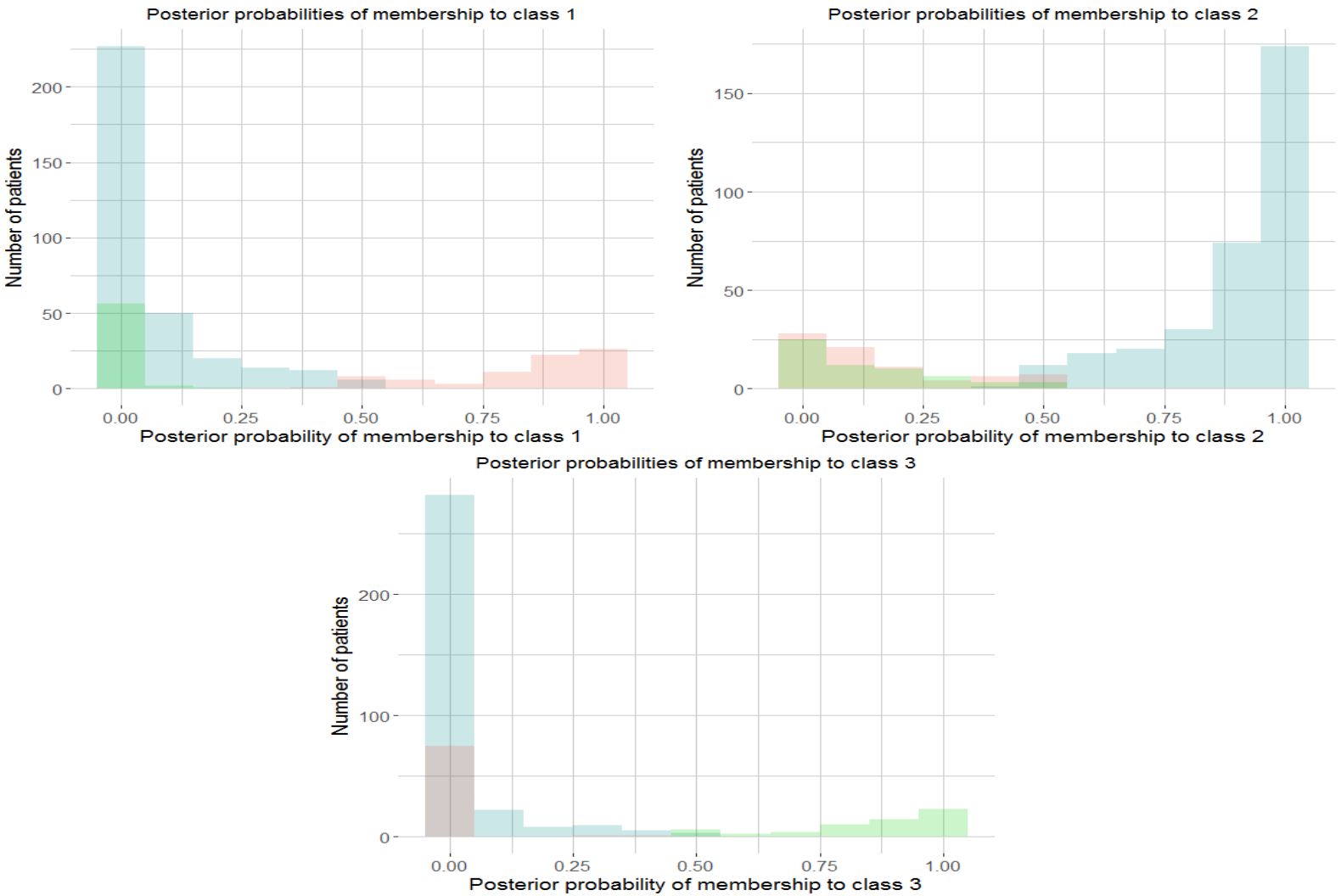



*Figure 5: Posterior probabilities of membership to the three-class model. Red bars indicate patients who were allocated to class 1, blue bars indicate patients who were allocated to class 2, and green bars indicate patients who were allocated to class 3.*

The two-class model offered the most parsimonious model, with a good description of the data and fewer parameters. Therefore, we primarily explored the classification based on a two-class piecewise growth mixture model.

*Predictive Validity*

To provide support for the validity of the two trajectory classes, we examined total scores of the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). HoNOSCA is a putative measure of mental health that includes assessment of symptoms of multiple disorders and functioning. We would expect similarity in trajectory shape, resulting in significant differences between classes at baseline and final time-point, and an insignificant HoNOSCA difference between classes at the breaking point of 18 weeks. In addition, we compared the percentage decline in HoNOSCA score at their final assessment.

Figure 6 plots the mean trajectory of HoNOSCA scores across trajectory classes. Visual inspection suggests general agreement in depressive symptom trajectories and mean scores of impairment across the two classes. As expected, class 1 showed a significantly higher baseline HoNOSCA score (19.9), than class 2 (18.3, $t$= -2.02, $p$=.047). In addition, class 1 showed a significantly higher HoNOSCA score at 95 weeks (14.7) compared with class 2 (6.5, t=-8.15, p<.001). Following the pattern shown in the MFQ trajectories, the difference between HONOSCA scores at the third assessment (18 weeks) was insignificant (class 1: 14.3, class 2: 12.8; $t$=-1.25, $p$=.215). Class 1 showed a 26% improvement in HoNOSCA score by the end of the trial, whereas Class 2 experienced a 64% improvement. Overall, HoNOSCA scores appear to follow similar trajectories to MFQ. Furthermore, linear regression showed that class membership significantly predicted HoNOSCA score at 95 weeks (Table 10). Class 1 on average have a higher HoNOSCA score by 0.5 standard deviations at 95 weeks than class 2. Together, this offers support that the two-class model has a level of predictive validity and is measuring clinically relevant changes.

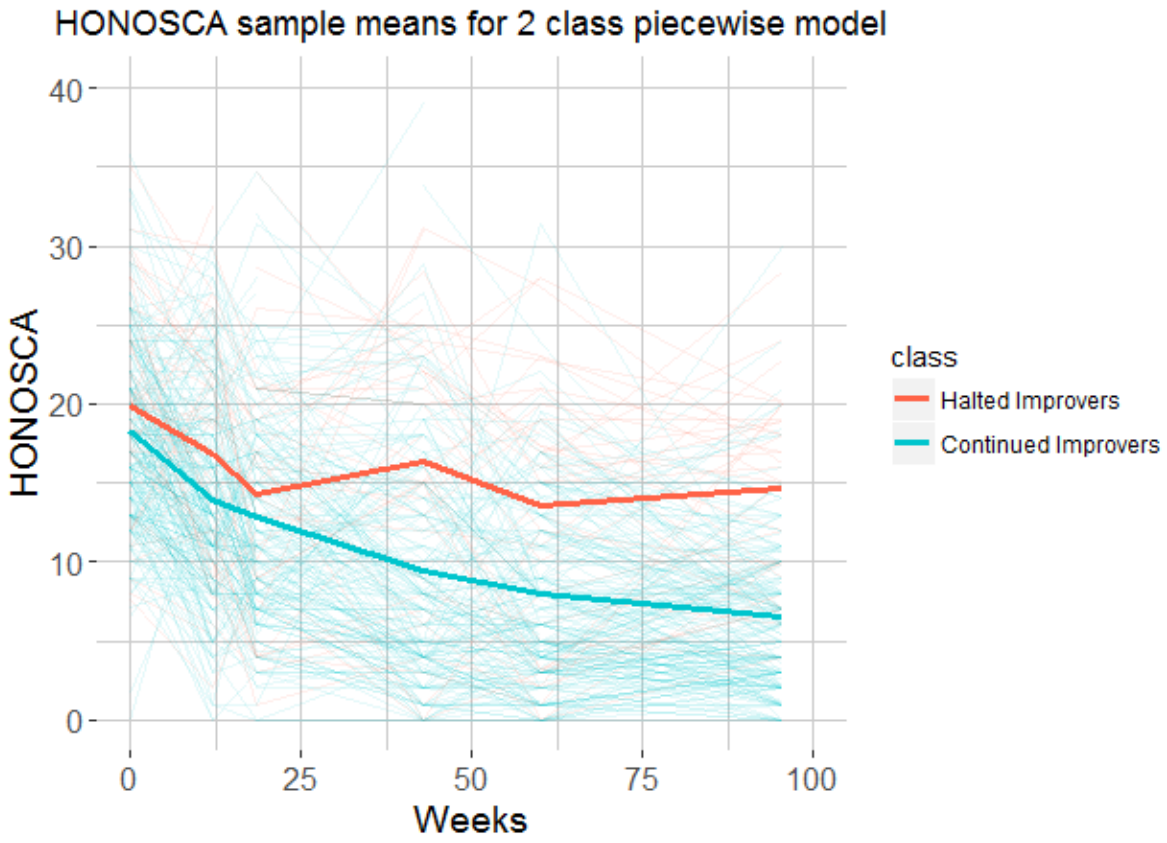



*Figure 6: HONOSCA sample means for the 2-class piecewise growth mixture model, derived from MFQ. Behind plots every individual patient's trajectory over time, colour coded to his or her respective classes.*

*Table 10: Regression results for class membership on HONOSCA scores at 95 weeks.*

| | **R2** | **B** | **SE B** | **β** | **p** |
|---|---|---|---|---|---|
| **Class1** | 0.249 | 8.180 | 0.861 | 0.499 | <.001 |

*Agreement between alternative categorical outcomes*

As a two-class model was the best fitting model for our data, it offers a strong comparison for a priori dichotomous outcome measures. Therefore, the agreement between traditional definitions of response/non-response and these trajectory classes was investigated, using the Cohen's *Kappa* coefficient of agreement (Cohen, 1960). Two of the most widely used definitions for clinical response were compared to the empirically derived trajectory classes described above. The continued-improvers were considered the comparative for "clinical responders", and halted-improvers were considered the comparative for "clinical non-responders".

The first traditional definition categorises patients based on whether they have experienced at least 50% improvement in depressive symptoms (measured by MFQ score here), by the end of trial (in our case, 95 weeks; Nierenberg & DeCecco, 2001). The correspondence between trajectory classes and clinical response categories indicated that 100% of halted-improvers were also clinical non-responders. However, only 269 of 391 (69%) of continued-improvers were also clinical responders. The remaining continued-improvers (122 of 391; 31%) were classified as clinical non-responders. There was moderate agreement between trajectory membership and clinical categorical outcomes ($k=0.412$, $p<.001$).

The second traditional definition categorises patients based on whether they have reached a clinical cut-off below a score considered to indicate free from diagnosis at the end of trial. This was taken as an MFQ score of less than 27 at 95 weeks (Kent, Vostanis, & Feehan, 1997). The correspondence between trajectory classes and clinical cut-off response categories indicated that 100% of halted-improvers were also clinical cut-off non-responders. However, only 332 of 391 (85%) of continued-improvers were also clinical cut-off responders. The remaining trajectory continued-improvers (59 of 391; 15%) were classified as clinical cut-off non-responders. There was stronger, albeit still moderate agreement between trajectory membership and clinical categorical outcomes when defined by a cut-off score at 95 weeks ($k=0.642$, $p<.001$).

Graphical comparison of these three approaches is shown in Figure 7.

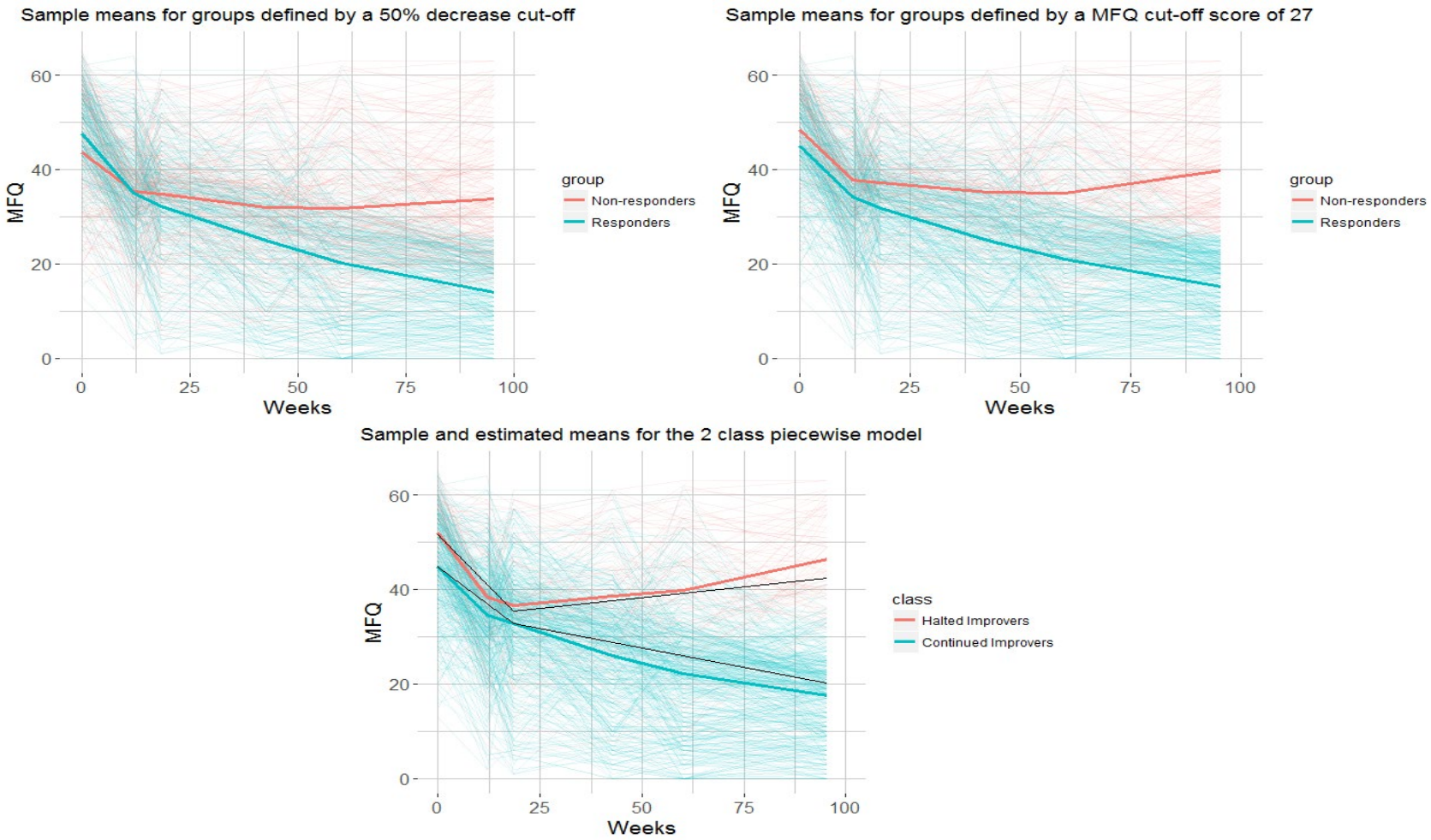



*Figure 7: Sample means for the three categorical approaches. Top-left: Clinical response; Top-right: Clinical cut-off; Bottom: Trajectory classes. Behind plots every individual patient's trajectory over time, colour coded to his or her respective classes.*

## Discussion

### *Key Results and Interpretation*

This empirical work has highlighted that significant heterogeneity exists in patients' patterns of symptom change over time, in a cohort of depressed adolescents. This is consistent with the findings from clinical adult (Cuijpers et al., 2005; Gueorguieva et al., 2011; Rhebergen et al., 2012; Stulz et al., 2010; Thibodeau et al., 2015; Uher et al., 2010) and adolescent studies (Scott et al., 2019), and general population studies of adolescents (Brendgen et al., 2005; Brière et al., 2015, 2016; Costello et al., 2008; Wickrama & Wickrama, 2010). Trajectories of symptom change in this sample agreed with previous literature that suggests symptom change is not a simple linear function over time (Brière et al., 2016; Gueorguieva et al., 2011; Keller et al., 2000; Uher et al., 2010). Here, symptom change followed a piecewise function, with two separate linear trajectories. We identified two classes of individuals that differed on specific characteristics of this piecewise function: a large majority (84.1%) group of continued-improvers and a small minority (15.9%) group of halted-improvers. The halted-improvers demonstrated significantly higher baseline symptom score and therefore illness severity and a faster initial rate of improvement compared with continued-improvers. However, unlike the continued-improvers, who showed a slowed, but significant continual decrease in depressive symptoms to 95 weeks, the halted-improvers showed desistance in improvement from approximately 18 weeks after the beginning of treatment. At the end of the follow up phases of the trial, the percentage reduction in symptoms was markedly different between the continued and halted-improvers, at 60.5% and 11% respectively. Interestingly we did not find support for a class that demonstrated no improvement, as has been reported in other studies (Brière et al., 2016; Gueorguieva et al., 2011; Thibodeau et al., 2015). This suggests that for most adolescent patients, the effects of the treatments used in the IMPACT trial were positive to at least some degree.

We did not observe the emergence of 4 classes in our data as we had hypothesised. This contrasts with previous empirical studies in adolescents (Brendgen et al., 2005; Brière et al., 2016; Costello et al., 2008; Wickrama & Wickrama, 2010). However, these studies were not clinical samples, but samples of the general population. Consequently, those studies covered a wider range of the depressive spectrum, which provides GM models with more variance to explain. Moreover, although there are a number of clinical adult studies which support the existence of more than two classes, these studies applied constraints on investigated

trajectory shape (Stulz et al., 2010; Thibodeau et al., 2015), and eliminated within-class variation from their models (Brière et al., 2015; Rhebergen et al., 2012). These methodological choices are often necessary when models struggle to fit the data, or when there are limited time-points available, and both lead models to favour more classes (Wickrama et al., 2016). Furthermore, the clinical adolescent GMM study also faced constraints over trajectory shape which may have led to favouring more classes (Scott et al., 2019). I discuss this in more detail below, but our results were stable across different sets of random starting values without these constraints, offering a much more representative model of patient experience of depressive symptom change.

It is interesting that the best fitting shape was a piecewise function. Previous studies reporting two classes have also described their trajectory groups as expressing a piecewise function (Gueorguieva et al., 2011; Uher et al., 2010). Quadratic and linear shapes imply smooth change over time, whereas piecewise functions are defined by a sharp change in trajectory. For the present study, this change occurred at 18 weeks, whereby symptom reduction significantly slowed from 18 weeks onwards regardless of class membership. Moreover, the group of halted-improvers actually showed significantly more rapid improvement between the start of treatment and 18-week assessment, but plateaued thereafter. We believe the significance of the 18-week break point might reflect treatment cessation, as the 18-week assessment was the closest time-point to the average length of treatment in the trial (27 weeks). This suggests that therapy may be acting on different mechanisms in the two groups. For the halted-improvers, longer-term clinician involvement might be beneficial to allow these patients time to improve to a point of remission, before stopping active treatment. However, it is important to note that as this group also exhibited a significantly higher baseline level of severity, the rapid initial reduction in symptoms may simply reflect a regression to the mean.

The halted-improver trajectory pattern is a novel finding in the literature. One lesson from this finding is that if only the first 18 weeks were considered in our work, the halted-improvers would show an identical profile to Scott and colleague's rapid responder group (Scott et al., 2019). However, it was found that patients who showed more favourable improvement within the first 18 weeks of treatment actually exhibited a less favourable trajectory pattern over 18 months. This is a stark contrast to Uher and colleagues (2010),

who supported the assertion that improvement within the initial stages of treatment is predictive of favourable long-term outcomes (Szegedi et al., 2009; Uher et al., 2010). However, their trial was limited to a 12 weeks duration, which underscores that to disentangle group trajectory patterns accurately, long-term follow-up in clinical trials is crucial. The early phase of treatment might not be the ideal time to assess treatment response and clinicians should be cautious of their interpretation of patients that appear particularly responsive during this phase of treatment. Indeed, as the greatest percentage reduction for this group was observed at 18 weeks (Table 9), followed by a worsening to only an 11% reduction at 18 months, a descriptive view of the data suggest that these patients may have begun to relapse following the end of their treatment phase. However, it is important to emphasise that the statistical value of the increase in this slope was insignificant. Nevertheless this current work supports the need for at least medium-term follow-up when designing future treatment trials to fully understand the trajectory of symptom change over time.

We considered whether our 2-class solution as a dichotomous outcome agreed with currently used outcomes. The findings showed only moderate agreement between empirical and a priori definitions of symptom change, similar to those of previous studies (Gueorguieva et al., 2011; Uher et al., 2010). The difference resides in between 15% or 31% of improvers (depending on definition) identified through GMM being misclassified as non-responders by a priori methods. This agrees with other authors who suggest that current clinical response definitions may be too strict (Gueorguieva et al., 2011). These differences demonstrate that the choice of methodology to determine outcome is particularly important if homogeneity is a goal for revealing the best group of non-responding individuals. This current work would caution researchers interested in such a sub-population from using cut-off or percentage change measures; it is possible that a significant percentage of potentially good responding patients might be misclassified as false negatives to treatment. The impact of this misclassification in clinical trials has been highlighted in previous GMM work (Cuijpers et al., 2005; Thibodeau et al., 2015; Uher et al., 2010) and as mentioned in the introduction to this chapter, the potential for the choice of definition to overturn conclusions of clinical trials emphasises that the level of correspondence between traditional measures of response and empirically-defined classes is inadequate (Gueorguieva et al., 2011; Uher et al., 2010). Until more is definitively known about response-based sub-classes of depression,

researchers must take care in their choice of outcome measure with respect to their research questions, and in particular try to minimise a false negative result.

*Study Limitations*

There are a number of limitations of the current study. Firstly, it is important to note that GMM are large-sample statistical techniques. While a sample of 465 is sufficient (Nylund et al., 2007), it is modest and many studies possess a sample of 600 or more (Brière et al., 2016; Gueorguieva et al., 2011; Rhebergen et al., 2012; Thibodeau et al., 2015; Uher et al., 2010). Sample size is therefore a limitation of this current work. Indeed, there is a danger of rejecting models with a potentially clinically relevant class purely due to insufficient sample size of the overall clinical trial, and a larger sample in the present study may have seen the emergence of a more stable and clinically relevant 3-class model. The smaller third class separated with significantly lower baseline severity, but no significant change in their depressive symptoms over either part of the model. Consequently, this describes the a group that might be unaffected by treatment, in agreement with other clinical samples in adults (Thibodeau et al., 2015; Uher et al., 2010). The 3-class solution did demonstrate a reduction in BIC, however it did not meet our a priori cut-off and entropy values suffered with the addition of another class. Indeed, our 3-class model was not rejected on insufficient class size, but failure to meet acceptable fit improvement while maintaining adequate class differentiation. This supports the assertion that our two-class model was not erroneously led by a modest sample size. However the possibility that a larger sample may have allowed for fit and entropy criteria of a 3-class model to improve to a point of acceptance remains to be determined.

The low-severity, unchanging group of the three-class model is also significant in the debate over heterogeneity within subgroups of depression. Firstly, it appears that this class is a result of a split of the continued-improvers class in the two-class model. This supports the stability and homogeneity of the halted-improvers class. Secondly, this additional class could help explain why the definition of 50% reduction shows the poorest correspondence with trajectory approaches; when defined through a percentage reduction, halted-improvers and this third group would merge. However, these patients display strikingly different patterns of symptom change over time, and likely illustrate differential underlying mechanisms.

Crucially, end-point symptoms are low in this third group, but high in the halted-improvers group.

A second limitation of this work is the lack of a non-symptom driven end-of-trial (nominally 86 but averaged 95 weeks) outcome validator, such as interpersonal function. However, the investigation of HoNOSCA, a measure of general psychiatric symptomatology and function, showed a certain level of predictive validity. The two classes' HoNSOCA scores showed similar trends to MFQ, with significant differences in baseline severity and end-point scores between the two classes. Furthermore, the two classes did not significantly differ in HoNOSCA scores at the 18-week break point, suggesting that HoNOSCA trajectories for these two classes mirror those of MFQ. Class membership was also a significant predictor of HoNOSCA score at the final assessment, in the expected direction. However, it is important to note that HoNOSCA has been shown to display multidimensionality. It is suggested that its sensitivity to change is better investigated as two subscales for emotional and behavioural symptoms for this scale, rather than total scores (Tiffin & Rolling, 2012). Without a pure singular measure of patient functionality, the extent of predictive validity our HoNOSCA assessment can provide for our work is limited. Unfortunately, data were not available from a separate cohort of sufficient similarity for validation, and our sample size was not large enough to divide into training and test sets. While less optimal, other resampling techniques that use the available data (such as leave-one-out cross validation) would provide an alternative way to investigate how well our model generalises. Future research should focus on validating this model either through these resampling techniques, or ideally, with the collection of a new adolescent cohort. By fixing the parameters of a model on a new dataset, to those of the model described here, one can test the reliability of the model to predict most likely class membership of new patients.

While this study is suggestive of differential underlying mechanisms, both within and between trajectory classes, it is important to note that there was no no-treatment control group used in this study, for ethical reasons. Consequently, the precise psychological treatment implications cannot be conclusively determined. The reference trajectory in those who may show spontaneous remission, or an active "placebo" effect, but have not received the 'active agent' is unknown. While placebos for psychological treatments are theoretically

complicated, the possibility that the effects observed here were not related to treatment must remain an alternative explanation.

Future studies would benefit from the collection of more time-points to allow for a more detailed investigation of trajectory shape both during and following treatment. In addition, an interesting extension of this work would be a latent transition analysis (LTA) (Muthén & Muthén, 2017). LTA would allow for the investigation of the extent of movement of individuals between classes, and predictors of such movement. Understanding the extent to which these classes are fixed provides an understanding of the usefulness of baseline predictors of membership. Moreover, it allows for the investigation of whether intervention can help alter a person's trajectory. For example, it would have been of interest in this current work to investigate whether subsequent SSRI prescription during follow-up increased the likelihood of halted-improvers to change trajectory. However, data collected after psychological treatment cessation was limited and did not allow for this type of exploration.

*Reflection on Methodology*

In longitudinal clinical trials such as IMPACT, an important and vital consideration is the adequate handling of missing data. For missing data handling, end-point analyses in clinical trials often use the last observation carried forward method (LOCF) (Gueorguieva et al., 2011; Lane, 2008). However, this method assumes that patients with missing data have either remained at the same level of severity, or ignores the difference in time between one patient's last assessment and another (Gibbons et al., 1993). Given our understanding of the complexity of drop-out in clinical care (O'Keeffe et al., 2018), these assumptions are short-sighted. Indeed, including length of time in study as a covariate has been shown to diminish previously significant treatment effects (Uher et al., 2010). LOCF methods in treatment studies therefore could bias results in favour of the treatment with less attrition (Lane, 2008). Attrition in longitudinal studies, particularly in mentally unwell populations is expected (O'Keeffe et al., 2018), so methods that are most robust in the presence of missing data should be favoured.

GM modelling itself is statistically stronger than typical end-point analyses that use LOCF, because they are themselves more robust in the presence of missing data (Uher et al., 2010).

However, despite this, and despite the promising retention rate of the study (80% of MFQ outcome data available at 86-weeks), model convergence struggled when using only the available data. Consequently, the imputed dataset available in the study database was used to produce the models (van Sprang, Neufeld & Goodyer; unpublished work). Multiple imputation (MI) differs from LOCF because it uses the observed data to make predictions regarding the likely value for the missing data (Sterne et al., 2009). MI does this multiple times, creating multiple imputed datasets. Consequently, more valid predictions are obtained because the multiple imputed values are averaged over the distribution of missing data, allowing for a measure of uncertainty (Dong & Peng, 2013).

One of the main assumptions of MI is that the data are missing at random (MAR) (Sterne et al., 2009). This means that observed data can explain adequately any systematic differences that exist between the missing and observed values (Sterne et al., 2009). Studies where there is a substantial amount of data missing, or an inadequate amount of supporting data to aid the imputations, would therefore struggle to support this assumption of MI. The IMPACT trial was able to utilise 24 different variables repeated over the 6 time points. Given this wealth of available data, MAR was a plausible assumption for these data, and consequently offers confidence in the validity of the imputed dataset. Moreover, although 5 imputed datasets are arguably sufficient (Schafer & Olsen, 1998), more recent research suggests that larger numbers of imputed datasets are necessary to reduce the potential variability of the MI process and maintain power (Graham, Olchowski, & Gilreath, 2007). Consequently, our study produced 50 imputed datasets, further supporting the validity of the data. Taken together, the MI dataset in IMPACT provided a robust method to address issues of missing data for our GMM models.

Growth mixture modelling is a powerful statistical technique that allows for a great deal of exploration in a dataset (Ram & Grimm, 2009). This flexibility is remarkably useful in developmental and clinical psychology, however it is not without assumptions or constraints (Ram & Grimm, 2009). Firstly, the adoption of GMM methodology inherently assumes a priori that categorically distinct trajectories exist (Bauer, 2007; Bauer & Curran, 2003; Tarpey & Petkova, 2010). This is because GMM assumes non-normality in data is a representation of a mixture of normal distributions ("Topic 6: Mplus Short Course Videos and Handouts," n.d.). However, this is one of a number of explanations for non-normal distributions. There is

therefore a danger of overestimating trajectory classes, and forcing descriptively meaningless classes out of a dataset that is truly one class, but non-normal (Bauer & Curran, 2003). Indeed, many have suggested that disease severity is best represented as a continuous latent predictor (Tarpey & Petkova, 2010). However, the limitation of these models is that they are yet to extend to incorporate longitudinal data (Tarpey & Petkova, 2010), and thus would be inappropriate for our specific research questions. In addition, it is not unheard of in depression literature for single class solutions to emerge through GMM (Gueorguieva et al., 2011). The two-classes solution in this current work is a fairly modest outcome compared to some studies (Brière et al., 2016; Thibodeau et al., 2015). Moreover, the growth functions in this current work make theoretical sense given the design of the clinical trial (Goodyer et al., 2011), and the emergent trajectories and differences between them agree with our current knowledge of response and relapse in adolescent depression (Birmaher et al., 2000; Goodyer & Wilkinson, 2018; Goodyer et al., 2017; Weersing et al., 2017). Furthermore, an investigation of the cross-sectional distribution of MFQ scores revealed that the data did not display significant skew or kurtosis at any time point (Appendix 1C), suggesting that our findings are not a misrepresentation of non-normal data.

Secondly, while GMM is an exploratory approach, it is a constrained one (Ram & Grimm, 2009). That is, GMM will search for unobserved classes, but only classes that fit the criteria of the model specification (Ram & Grimm, 2009). This means that the choices made by researchers and aspects of the clinical trial itself have a pivotal influence over the resultant model and extracted classes. For instance, during the first step of specifying a GM model, the researcher must identify the optimal trajectory shape in a single growth curve. More complex single growth curves will explain more variance in the model and result in the subsequent extraction of fewer classes. Researchers must therefore adequately consider what are plausible trajectory patterns for their data, based on theory, previous empirical findings and the design of the clinical trial itself (Ram & Grimm, 2009). For our work, four models were tested. We saw theoretical justification for investigating linear and quadratic trends based on previous work (Gueorguieva et al., 2011; Keller et al., 2000; Stulz et al., 2010; Thibodeau et al., 2015; Uher et al., 2010), and trial design gave justification for piecewise functions; treatment cessation suggested reasoning for the presence of a sharp break point, and as the average time at which treatment ended in our study fell between two time points of the trial, an investigation of two piecewise models was warranted. There

have been no reports of cubic functions in clinical trials of depression, nor was there reason to believe another sharp change would occur during the trial. More complex piecewise models (such as with quadratic functions) would not have been stable with the number of time points we had available and would risk over-fitting a single growth curve. We therefore investigated a number of growth functions of varying complexity, but those that were theoretically plausible given the study design and population, and reasonable given the available number of measurements.

An addition unique to our study at this stage was the investigation of including time of assessment as a covariate in the models. While GM models are robust against variation in time ("Topic 6: Mplus Short Course Videos and Handouts," n.d.), the extent of variation observed in this particular trial (Table 2) warranted testing models with time of assessment as a covariate. Inclusion of this covariate significantly improved the fit of the growth curve model, and was therefore beneficial in ensuring that additional classes in subsequent steps would not emerge that were erroneously capturing this variation.

Step two involved defining the number of classes present in the dataset, which can be achieved in a number of ways. Latent class growth analysis (LCGA) is a simplification of GMM, whereby the within-class variation is fixed at zero. Consequently, LCGA has a tendency to favour more complex models with more classes, because the models are trying to force individuals into identical trajectories; there is no allowance for flexibility around a particular trajectory class (Wickrama et al., 2016). Where substantial inter-individual variation exists, more classes are necessary to fit the data adequately. LCGA does allow for easier interpretation, and is often favoured when there are instability or convergence problems in GM models (Brière et al., 2016; Rhebergen et al., 2012). However, we decided that, firstly for our analyses, a technique that biased towards favouring more classes would not have been ideal, considering the relatively good fit of the single class solution. Secondly, to assume that no individual variation exists within classes in depressive patients is not representative of real data. We therefore favoured a GMM analysis at outset, which would allow for within-class variation, and our results were stable enough to accept these models. We did not however allow for between-class variation, which is another possibility for defining classes in trajectory modelling. As there was no evidence of significant variation between classes in any growth factors (Appendix 1A), freeing these parameters would have

made the models unnecessarily complex without justification, and confuse interpretation. In addition, there is a lack of sufficient evidence to suggest a theoretical reason for why classes would experience different amounts of variation.

A number of factors that relate to the specific design of the trial itself can also influence both the number of classes that emerge, but also their interpretation. As GMM are often secondary analyses (Gueorguieva et al., 2011; Uher et al., 2010), these factors are usually outside of the researcher's control. However, it is important to consider when interpreting the findings from GMM studies. I have already touched upon sample size, but in clinical trials of adult depression, the length of study has also shown to influence not only the number of classes that emerge from the data, but also their pattern of change over time (Brière et al., 2016; Thibodeau et al., 2015). For instance, Gueorguieva and colleagues (2011) acknowledged that given a longer follow-up period, their "non-responding" patients may actually follow a slow-responding trajectory. Furthermore, Thibodeau and colleagues (2015) showed that the response classes that emerged from their data actually differed depending on the length of follow-up. While over a 12-week period, 3 classes were specified, when this was extended to 6 months, a fourth class emerged from the data. This additional class followed a trajectory pattern that illustrated non-response, which was not present during initial stages of the trial. These findings agree with our assertion that the initial stages of treatment might not be the ideal time to adequately assess treatment response and that short-term trials provide an incomplete picture of response patterns in depression. However, the sensitivity of GM models to trial length means that it is imperative that the context of the trial is retained in interpretation of classification labels, and is why the present author has reframed from labelling their groups beyond a description of the trajectories.

A final point on interpretation considers the measurement tool. As mentioned in the introduction to this chapter, prior to defining what is a reasonable response to treatment, researchers must firstly quantify the symptom change in depression that is associated with the treatment received. As is the case with this current work, symptom change is typically quantified through self-reported outcomes like the MFQ. A reflection on the extent to which self-report measures accurately describe symptom change is a separate research question to the questions addressed in this current thesis. However, for interpretation, one must firstly

remember that in depression these outcomes are typically susceptible to bias from the very illness that they are attempting to measure (Nierenberg & DeCecco, 2001). Secondly, symptom items in these measures are often given equal weighting, but qualitatively, there is a substantial clinical difference in the significance of a reduction of suicidality and the reduction of fatigue, for example (Malhi & Byrow, 2016). Taking sum scores therefore precludes the investigation of whether treatment is differentially sensitive for specific symptoms (Bagby, Ryder, Schuller, & Marshall, 2004), particularly if instruments have shown to possess multidimensionality (Pancheri, Picardi, Pasquini, Gaetano, & Biondi, 2002). However, multidimensionality does not appear to be an issue with the MFQ. The internal construct validity has received support through both item response theory and categorical data factor analysis in measuring a single continuum of depression severity (Sharp, Goodyer, & Croudace, 2006). Nevertheness, the outcome of GMM must be viewed with the consideration of the limitations of the assessment tool used to define the model.

To conclude these reflections, model design and selection in GMM is not clear-cut. It requires a holistic consideration and interpretation of a number of factors, as well as an adequate description of trajectories with respect to the model specifications and the trial design, to ensure the model corresponds with the underlying theory (Curran, Obeidat, & Losardo, 2011; Curran & Willoughby, 2003). As eloquently stated by Ram and Grimm(2009); *"model selection is an art; informed by theory, past findings, past experience and a variety of statistical fit indices"*(p.8). *"There is no substitute for careful definition of the research problem"* (Ram & Grimm, 2009, p.11).

**Appendix 1A**

*Table 1: Wald chi-square test statistics for testing significance of the variances of growth factors between classes for the 2-class model.*

| | **Wald $X^2$** | **df** | **p** |
|---|---|---|---|
| **I** | 0.144 | 1 | .7043 |
| **S1** | 0.012 | 1 | .9144 |
| **S2** | 1.113 | 1 | .2914 |

**Appendix 1B**

*Mplus code.*

```
INPUT INSTRUCTIONS

Title: Two_class_GMM
Data: file is 'Mean_Wide_Imputed_MFQ_FINAL.dat';
Variable:
  Names are
    ID gender ageBase arm region time0 randTime adhere famMed imd ethni
    dura0 mfq0 bc0 loi0 rcmas0 rses0 rrs0 rtshR0 rtshSH0 nssi0 attem0
    tho0 sleep0 BehavD0 ComorD0 AnxD0 AnxOth0 MDD0 ssri0 time6 dura6 mfq6
    bc6 loi6 rcmas6 rses6 rrs6 rtshR6 rtshSH6 nssi6 attem6 tho6 sleep6
    BehavD6 ComorD6 AnxD6 AnxOth6 MDD6 ssri6 time12 dura12 mfq12 bc12
    loi12 rcmas12 rses12 rrs12 rtshR12 rtshSH12 nssi12 attem12 tho12 sleep12
    BehavD12 ComorD12 AnxD12 AnxOth12 MDD12 ssri12 time36 dura36 mfq36
    bc36 loi36 rcmas36 rses36 rrs36 rtshR36 rtshSH36 nssi36 attem36 tho36
    sleep36 BehavD36 ComorD36 AnxD36 AnxOth36 MDD36 ssri36 time52 dura52
    mfq52 bc52 loi52 rcmas52 rses52 rrs52 rtshR52 rtshSH52 nssi52 attem52
    tho52 sleep52 BehavD52 ComorD52 AnxD52 AnxOth52 MDD52 ssri52 time86
    dura86 mfq86 bc86 loi86 rcmas86 rses86 rrs86 rtshR86 rtshSH86 nssi86
    attem86 tho86 sleep86 BehavD86 ComorD86 AnxD86 AnxOth86 MDD86 ssri86
    lgTime6 lgTime12 lgTime36 lgTime52 lgTime86;

USEVAR= mfq0 mfq6 mfq12 mfq36 mfq52
        mfq86 time6 time12 time36
        time52 time86;

IDVARIABLE= ID

Class=c(2);

Analysis:
```

```
    Estimator = MLR;
    Processors=4;
    STARTS=5000 100;
    TYPE=MIXTURE RANDOM;

Model:
 %Overall%
        i s1 | mfq0@0 mfq6@1.2 mfq12@1.8 mfq36@1.8 mfq52@1.8 mfq86@1.8;
        i s2 | mfq0@0 mfq6@0 mfq12@0 mfq36@4.3 mfq52@6 mfq86@9.5;
    i WITH s1;
    i WITH s2;
    s1 WITH s2;

    mfq6 ON time6;
    mfq12 ON time12;
    mfq36 ON time36;
    mfq52 ON time52;
    mfq86 ON time86;

Output:
    sampstat;
    Tech7;

 PLOT:
    TYPE=PLOT3;
    SERIES = mfq0(0) mfq6(1.2) mfq12(1.8)
    mfq36(4.3) mfq52(6) mfq86(9.5);

 SAVEDATA:
 File="C:\Users\sed48\GMM_CI_2_cprob.csv";
 SAVE=CPROB;
```

**Appendix 1C**

Table 1 shows skewness and kurtosis values for MFQ scores at each time point. Skewness is a measure of symmetry, whereby 0 represents that the mean and median are equal: a normal distribution. Negative values indicate that the mean is less than the median, and positive values indicate that the mean is greater than the median. A skewness value greater than ± 1 indicates a significant problem (Bulmer, 1979). No time-point in our data raised concern for skewness.

Kurtosis is a measure of whether the tails of the distributions are over-represented, or under-represented, compared to a normal distribution. Positive kurtosis scores indicate that the data possess more outliers than expected in a normal distribution, whereas negative kurtosis scores indicate that the data possess fewer outliers than expected in a normal distribution. Kurtosis values greater than ± 3 indicate a significant problem (Bulmer, 1979). No time-point in our data raised concern for kurtosis.

*Table 1: Skewness and kurtosis for MFQ scores at each time point*

| MFQ | Mean | SD | Skew | Kurtosis |
|---|---|---|---|---|
| *Baseline* | 45.939 | 10.553 | -0.565 | -0.205 |
| *12 week* | 35.228 | 10.925 | -0.196 | 0.269 |
| *18 week* | 33.305 | 11.953 | -0.414 | 0.211 |
| *43 week* | 27.949 | 13.331 | 0.060 | -0.252 |
| *60 week* | 25.026 | 14.167 | 0.392 | -0.218 |
| *95 week* | 22.238 | 13.827 | 0.798 | 0.353 |

*MFQ: Mood and Feelings Questionnaire*

## Chapter 2: Behavioural predictors of trajectory classes.

### Introduction

To meet DSM5 diagnostic criteria (APA, 2013) for major depressive disorder, endorsement of any 5 of a potential 9 symptoms is necessary, with one being either persistent low mood (or irritability, for children/adolescents), or anhedonia (reduced interest or pleasure). These symptoms, once meeting a threshold level, are then evaluated for level of severity. Consequently, a wide range of phenotypic presentations is possible with depression, and a broad scope of illness severity. This makes depression a condition that is clinically heterogeneous. Patients differ in patterning of symptomatology, duration of illness, comorbidities, coping styles, cognitive abilities and demographics (Katon, Unützer, & Russo, 2010). It is therefore perhaps not surprising that as yet there is little precision in treatment and that any single therapy is only moderately successful (Cuijpers, 2017; Cuijpers et al., 2011; Kessler et al., 2003; Moncrieff, 2018; Trivedi et al., 2016). A substantial amount of research has therefore focused on attempting to identify more specifically which aspects of this heterogenous condition are associated with predicting or moderating clinical response (Bagby, Ryder, & Cristi, 2002; Kemp, Gordon, Rush, & Williams, 2008).

The previous chapter described an empirical method of categorising our adolescent depressive patient sample into trajectories of symptom change over time, following psychological treatment. We saw that agreement between this data-driven approach and a priori categorical methods of defining good and poor responders were moderate at best. The consideration of multiple time-points into the categorisation process for GMM provides a much less arbitrary approach to defining groups relating to treatment outcome, and the results demonstrated a particularly homogenous group of halted-improvers. This places the empirical method at an advantage for investigating potential clinical predictors of this grouping, and this will be the focus of this next chapter's work. A particular interest of this work (discussed below) was the contribution of subclinical psychotic features to treatment prognosis; a currently understudied factor in this field of research. However, I endeavoured to investigate these features within a broader clinical model, allowing adequate consideration of potentially overlapping contributors, such as severity and clinical complexity. The IMPACT trial (Goodyer et al., 2011) was well-suited to investigate such predictors, as the trial collected a large battery of clinical data, including self-reported

symptomatology and clinician-led diagnostic interviews. My model focussed on demographic and clinical characteristics, which I have summarised into three broad categories for this work, following categories outlined by previous authors (Jensen, Hoagwood, & Petti, 1996; Kemp et al., 2008). Predominantly, these will fall into aspects relating to: i) demographics; ii) severity of illness; and iii) clinical complexity of the patient. Psychotic features (iv) will be discussed separately as a distinct area of interest. The introductory review presented below was limited to those variables that were available in the IMPACT cohort. I note the impact this has on interpretation in the discussion.

*Demographics*

Our first demographic consideration was gender. The incidence of depression is known to differ substantially between males and females. During adolescence and adulthood, the prevalence of depression is much higher in females (Goodyer et al., 2017; Sramek, Murphy, & Cutler, 2016; Weissman et al., 1993; Wilkinson, Dubicka, Kelvin, Roberts, & Goodyer, 2009), yet males appear to experience more severe depression and are at a higher risk of developing persistent depression (Dunn & Goodyer, 2006). It was thought that females respond better than males to treatment (Braun, Gregor, & Tran, 2013), yet meta-analytic studies and reviews present conflicting findings on this matter. For instance, one meta-analysis showed that the higher the proportion of females within clinical trials of cognitive-behavioural therapy (CBT), the better the response rates of those trials (Braun, Gregor, & Tran, 2013) and one growth mixture modelling (GMM) study found that males were associated with the poorest symptom courses (Rhebergen et al., 2012). However, reviews of both antidepressant medication and CBT studies have failed to show convincing support that gender is a predictor or moderator of treatment outcome (Cuijpers, Weitz, et al., 2014; Sramek et al., 2016). Moreover, a recent review of this literature in adolescent depression suggested that demographic characteristics have marginal effects compared with clinical characteristics for treatment response (Weersing et al., 2017). Taken together, these results suggest that, despite the large discrepancy in prevalence of depression, gender does not have a substantial influence on treatment outcome. The majority of studies however, investigated gender differences in a post-hoc design, which limits our understanding of the true effect of gender on treatment response.

Following this, a second demographic characteristic for consideration is the influence of age on treatment outcome. A few adolescent studies have supported the finding that younger

age may be associated with more favourable outcomes (Clarke et al., 1992; Curry et al., 2006; Goodyer, Herbert, Secher, & Pearson, 1997; Jayson, Wood, Kroll, Fraser, & Harrington, 1998; Scott et al., 2019), but the literature is small and hindered by the same problem as those studies reporting gender differences; studies were not designed to investigate age effects on treatment response a priori.

Taken together, there is currently little evidence that either of these demographic variables operate as clear-cut predictors of treatment response, but their value is still to be determined appropriately. Consequently, models should still consider their potential influence in analyses, yet their prognostic value is likely to be modest at best (Weersing et al., 2017).

*Severity of illness*

Variability in severity; (a higher number of, and/or an increased intensity of symptoms) is one of the most consistently reported associations with non-response in the treatment of depression, across age ranges (Curry et al., 2006; Goodyer et al., 1997; Katon et al., 2010; Wilkinson et al., 2009), and is considered one of the most reliable predictors of treatment outcome (Kemp et al., 2008). Indeed, a study of 702 adults with depression, half of which were already resistant to two antidepressant treatments, showed that severity was associated with a 1.7 times increased risk of being categorised as resistant to another antidepressant (Souery et al., 2007). Furthermore, an exploratory study of adults found that, among other predictors discussed later, severity predicted non-remission, regardless of treatment modality (Frank et al., 2011). Severity also appears important in predicting response in late-life depression (Katon et al., 2010). A study of 871 patients aged over 60 found that patients with higher initial levels of severity, despite receiving more intense and longer treatment, still experienced a poorer outcome both at the end of trial, and during the 12 month intervention period (Katon et al., 2010). The authors concluded that severity is a robust predictor of non-remission that strongly associated with other variables indicative of non-response (Katon et al., 2010).

Empirical studies of treatment response, such as that outlined in Chapter 1, have also invariably found that baseline severity is a significant driver of trajectory groupings (Brière et al., 2016; Cuijpers et al., 2005; Gueorguieva et al., 2011; Rhebergen et al., 2012; Stulz et al.,

2010; Thibodeau et al., 2015). Moreover, studies have further associated these differential severity classes with differential responses to treatments (Brière et al., 2016; Cuijpers et al., 2005; Gueorguieva et al., 2011; Rhebergen et al., 2012; Stulz et al., 2010). For example, Gueorguieva and colleagues (2011) reported that higher baseline depression severity was associated with an increased likelihood of membership to their non-responder (as opposed to responder) trajectory class. This finding across both traditional and empirical studies of treatment response highlights the prognostic value of initial severity in the treatment of adult depression.

While the majority of empirical studies have thus far been conducted on adults, a number of traditional studies have associated greater initial severity with poorer outcome to CBT in adolescent depression (Brent et al., 1998; Clarke et al., 1992; Goodyer et al., 1997; Jayson et al., 1998). Authors have suggested that these more severely-ill patients may require more intensive treatment or additional pharmacological interventions at outset to aid their recovery (Brent et al., 1998). However, the TADS study reported that illness severity itself demonstrated a moderating, rather than predictive effect (Curry et al., 2006). While the presence of a more chronic form of depression at baseline showed to be a predictor of a poor response, patients with mild or moderate depression benefitted more from combination treatment, while severely depressed individuals saw no benefit of the addition of psychotherapy to pharmacological treatment. The authors hypothesised that severely depressed adolescents may require pharmacological treatment to improve to a level where psychological therapies can prove effective. Indeed, this fits with personal discussions the current author has had with patients about their experience of treatment efficacy. For studies investigating psychological interventions alone in adolescents, such as the IMPACT study, the literature would suggest that severity would play a key role in a predictive model of outcome and indeed, this is what the findings of Chapter 1 demonstrated; baseline severity as measured by self-report MFQ differentiated trajectory classes in this cohort.

Typically considered one of the most severe symptoms of depression is suicidal ideation, and many studies have suggested associations between this specific symptom and treatment non-response is independent of severity (Curry et al., 2006; Frank et al., 2011; Scott et al., 2019; Souery et al., 2007; Wilkinson et al., 2009). For example, Souery and colleagues (2007) found that suicidal risk demonstrated a stronger association with treatment resistance than

general severity, showing a 2.2 fold increase in risk of resistance. Frank and colleagues (2011) also associated suicidality with increased likelihood of resistance to monotherapy. This association has been replicated in adolescent samples as well. The TADS study reported that adolescents with more severe suicidal ideation benefitted less from any treatment (Curry et al., 2006) and Wilkinson and colleagues (2009) found that in the Adolescent Depression Antidepressants and Psychotherapy Trial ((ADAPT); Goodyer et al., 2007), severity, comorbid obsessive-compulsive disorder and suicidal ideation together predicted continued depression in adolescents after 28 weeks of treatment with SSRIs, a brief psychosocial intervention plus/minus CBT. ADAPT also reported that adolescents presenting with non-suicidal self-harming behaviours showed a three-fold increased risk of suicidal attempts (Wilkinson, Kelvin, Roberts, Dubicka, & Goodyer, 2011); this was later replicated in an adolescent sample of treatment resistant depression (Asarnow et al., 2011). Both suicidality and self-harm may therefore index unique moderating influences on treatment response, which is potentially independent of overall severity and personal impairment of the depression.

While symptomatology is an easily collectable indicator of illness severity, in clinical assessments of depressed cases clinicians often gauge severity on the extent to which their illness is impacting their daily function. A number of studies have consequently investigated the effects of non-symptom driven functional impairment on prognosis as a separate indicator relating to severity, such as The National Institute for Mental Health Treatment of Depression Collaborative Research Program (Sotsky et al., 1991). These authors highlighted the importance of function in their multi-treatment placebo-controlled trial, finding that social, cognitive and workplace function all differentially predicted outcome of specific treatments. In addition, Frank and colleagues (2011) found that in adults, greater severity of work and social dysfunction predicted a more prolonged illness duration before remission was achieved. In adolescents, the TADS study reported that poorer initial levels of global functioning resulted in a less favourable outcome to all treatment modalities (Curry et al., 2006). Levels of function are therefore likely related to severity of symptoms. Overall these findings show that potential indicators of poor prognosis may extend beyond symptomatic variables, to impairment of daily function and these should be considered as potential predictors or moderators of symptom change and treatment response.

*Clinical Complexity*

As we have already discussed, depressed patients can exhibit a wide range of phenotypic heterogeneity in depressive symptoms. However, these symptoms are not necessarily limited to symptoms of depression. Indeed, Goodyer and colleagues (1997) found that 93% of their sample of depressed adolescents were diagnosed with at least 1 other psychiatric disorder and conversely, depression is one of the most common comorbid disorders in patients with other mental health disorders as their primary diagnosis (Braun, Sunday, & Halmi, 1994; Buckley, Miller, Lehrer, & Castle, 2009; Garber & Weersing, 2010). Consequently, the level of clinical complexity a patient presents with may complicate treatment procedures and serve as a marker for vulnerability to treatment resistance (Brent et al., 1998; Curry et al., 2006; Frank et al., 2011). Indeed, the TADS study showed that adolescents with more than 1 comorbid disorder were less likely to benefit from treatment (Curry et al., 2006), as was also reported in the open-treatment trial; the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study of adults (Trivedi et al., 2006). Furthermore, a number of review articles that attempt to categorise these predictive characteristics of non-response invariability produce a grouping related to clinical complexity or comorbidity (Bagby, Ryder, & Cristi, 2002; Brent et al., 1998; Curry et al., 2006; Kemp et al., 2008). As such, clinical complexity, as indexed by comorbidity, is an important patient characteristic that could affect trajectory membership of patients in the IMPACT study.

The most common comorbidity associated with depression is generalised anxiety disorder together with panic and social phobia (Kessler et al., 2003); and in adolescents separation anxiety (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015). Unsurprisingly therefore, the presentation of comorbid anxiety specifically has been reported as either a predictor of poorer treatment outcome (Clarke et al., 1992; Curry et al., 2006; Frank et al., 2011), or moderator of differential outcome (Brent et al., 1998; Jensen et al., 1996; March et al., 2009) in depressed patients. For instance, a multicentre study found that comorbid panic or anxiety disorder showed a greater increased risk of resistance to treatment than either suicidality or depression severity (Souery et al., 2007). Two studies of adults by Frank and colleagues (2011; 2000) have also supported these findings. Both studies found longer remission times for patients with comorbid panic disorder or phobias, social phobia in particular (Frank et al., 2011). The link with social phobia is interesting as it may relate to the significance of social function on treatment outcome discussed earlier (Jayson et al., 1998;

Sotsky et al., 1991). The authors highlighted that these effects were not specific to any treatment and suggested that this complexity at baseline presentation may indicate that multiple treatment modalities are required to provide more adequate response rate than was achieved with the monotherapeutic approaches used in their studies (Frank et al., 2011; Frank et al., 2000).

Indeed investigations of adolescent populations have also yielded similar results with respect to anxiety. Curry and colleagues (2006) found that adolescents who were diagnosed with an anxiety disorder in particular benefitted less from either pharmacological, psychological or combination treatment. Brent and colleagues (1998) also found that the presence of comorbid anxiety disorder in depressed adolescents predicted a higher likelihood of a depression diagnosis after 16 weeks of treatment, to any of the three different psychosocial treatments utilised in their study. Interestingly, within this group of anxious-depressed teens, differential treatment effects were observed. Adolescents with anxious-depression responded better to CBT than systemic-behavioural family therapy or non-directive supportive therapy. This supports the assertion that comorbid conditions may require more tailored treatment approaches than cases of single diagnoses, resulting in poorer outcomes in studies that investigate a single type of treatment approach. The relationship between comorbid anxiety and treatment resistance in adolescent depression has also been supported in longitudinal studies (Sanford et al., 1995). Adolescents with persistent depression one-year post-treatment were more likely to have a comorbid anxiety or substance use disorder diagnosed at baseline (among other prognostic factors, such as older age and family relationship difficulties; (Sanford et al., 1995)). Interestingly, baseline severity and functioning did not influence the power of these relationships in this study, highlighting that comorbidity, while related, is a distinct factor that has its own influence on longer-term treatment response (Sanford et al., 1995). The UK pragmatic trials however show no specific moderator effects of anxiety disorders on any specific treatment suggesting that whatever the effects of anxiety on depression treatment it is unlikely to be differentiated by the type of therapy given (Goodyer et al., 2017).

Although less consistently, other comorbid disorders have been associated with difficulties to achieve remission in depression. Firstly, a naturalistic study of adolescent depression associated comorbid obsessive compulsive disorder (OCD) with persistence of depression at

36 weeks (Goodyer et al., 1997), along with depression severity and age. These findings have been supported in larger samples of adolescent depression as well (Wilkinson et al., 2009). Interestingly, without OCD, the predictive effects of age and severity were noticeably weakened. Moreover, without these factors together, adolescents were highly likely to respond to their given treatment (Goodyer et al., 1997). This illustrates the need for a multifactorial approach to the question of treatment resistance, as effects may be additive.

A number of studies have also associated comorbid bipolar disorder or bipolar symptomatology with a delay in treatment response to both medication and psychotherapies (Dudek et al., 2010; Frank et al., 2011). Frank and colleagues (2011) found that, along with panic disorder and social phobia, aspects relating to sub-threshold bipolarity, such as psychotic features, psychomotor retardation, lifetime suicidality and neurovegetative symptoms, were all associated with a longer time to remission in adult outpatients with depression. In addition, patients with bipolar symptoms appeared to show differential treatment effects, favouring SSRIs over psychological treatment (Frank et al., 2011). Indeed, Dudek and colleagues (2010) proposed that undetected bipolar features might explain treatment resistance in some cases. Their work found that, even after exclusion of patients with diagnosed bipolar disorder, higher scores of bipolarity were associated with an increased likelihood of failing to achieve remission with multiple antidepressant medications in depressed adults. The presentation of symptoms that do not necessarily meet threshold for diagnosis may therefore still serve as indicators of clinical complexity, and thus influence treatment response to therapies optimised for predominantly depression.

Overall, comorbidity, whether symptomatic or diagnostic, has emerged as a relatively consistent indicator of poor response, however inconsistencies do exist. For instance, Clarke and colleagues (1992) compared the agreement of continuous and diagnostic outcomes. They found that lower trait anxiety emerged as a predictor of better outcome to group CBT on continuous measures of depression, but conversely, an increasing number of comorbid diagnoses were associated with a better outcome on a diagnostic level. This contradiction underscores the importance of accurately defining recovery, as the authors further found that not one predictor was consistent across dichotomous and continuous outcome measures (Clarke et al., 1992). Empirical methods of defining treatment response have, however, largely supported the assertion that clinical complexity, as described through

measures of comorbid conditions, often hold prognostic value for unfavourable outcomes (Cuijpers et al., 2005; Gueorguieva et al., 2011; Scott et al., 2019; Stulz et al., 2010; Thibodeau et al., 2015). Along with severity, a commonality in GMM studies has been that higher baseline anxious traits, either measured by diagnostic (Cuijpers et al., 2005) or self-report scales (Gueorguieva et al., 2011; Stulz et al., 2010), associated with the undesirable trajectories across studies. These trajectories were often those most severe (Stulz et al., 2010), or those showing little (Cuijpers et al., 2005) or no symptom improvement over time (Gueorguieva et al., 2011). Although one study did not find any baseline differences in axis 1 disorder diagnoses (which includes anxious disorders) between their responder groups (Thibodeau et al., 2015), patients with diagnosed personality disorders showed almost a two-fold increased likelihood of membership to the non-responder class, compared with rapid responders (Thibodeau et al., 2015). Taken together, empirical work largely supports the notion that comorbid conditions or subclinical symptomatology of other psychiatric disorders are factors that can negatively influence a patient's trajectory during and after treatment.

*Psychotic features*

As touched upon above, there is a growing belief that subclinical symptoms have an influential effect on treatment effectiveness. Subclinical psychotic features, or psychotic experiences (PEs) in particular, have gained increasing attention in the context of depression in recent years and were of particular interest for this current work due to a number of sample characteristics. Firstly, PEs are generally associated with younger age (Kelleher et al., 2012; Wigman et al., 2012). Indeed, the incidence of PEs in adolescence populations is found to be higher than in adult samples (Kelleher et al., 2012). Moreover, PEs are found to co-occur with sleep disturbance (Jeppesen et al., 2015), which was one of the highest reported symptoms of the IMPACT cohort (Goodyer et al., 2017), even above low mood. Taken together, investigating the presence of PEs may be an important consideration when discussing prognostic questions of outcome for this sample of depressed adolescents.

The prevalence of PEs in depressed populations is remarkably high; presenting in as much as 30% of depression cases (Perlis et al., 2010; Wigman et al., 2012). Perez and colleagues (2018) also found an increased likelihood of PEs in 30% of a clinical sample of patients presenting with anxiety and depressive disorders to UK Improving Access to Psychological

Therapies (IAPT) services. Conversely, studies have shown that up to 80% of adolescents presenting with PEs (termed "at-risk-mental-state" (ARMS) patients) exhibit at least one diagnosis of mental illness (Kelleher et al., 2012); often anxiety or depression (Hui et al., 2013). MDD and clinical psychosis are known to share predisposing risk factors (Björkenstam, Burström, Vinnerljung, & Kosidou, 2016; Egan et al., 2001; Massat et al., 2005; Sundquist, Frank, & Sundquist, 2004) and the relationship between PEs and mental illness is as strong for emotional disorders (Jeppesen et al., 2015) as it is for psychotic disorders (Dominguez, Wichers, Lieb, Wittchen, & van Os, 2011).

In line with these findings, Stochl and colleagues' (2015) further advocated that PEs should be considered as existing on a single continuum of depression severity, marking the upper extreme. In their latent class analysis of two community cohorts, a single dimension of general distress emerged, rather than two separate dimensions of depression and psychosis. The study by Perez and colleagues (2018) mentioned above supports this view of PEs, as do a number of studies of ARMS patients (Fusar-Poli, Nelson, Valmaggia, Yung, & McGuire, 2014; Hui et al., 2013). Comorbid depression and anxiety in these ARMS patients was often associated with indicators of depression severity, such as increased suicidal ideation, self-harm and more impaired global functioning (Fusar-Poli et al., 2014; Hui et al., 2013). Current diagnostic tools however, do not relate the presence of PEs to depression or anxiety disorders (APA, 2013; World Health Organisation [WHO], 1992). Moreover, at least in the UK, IAPT services do not make any formal assessment of PEs in their depressed or anxious patients; thus rendering PEs a potentially untreated aspect of patients' illness (Perez et al., 2018).

The notion that PEs could indicate severity is appealing in the context of treatment response, as the most consistently found predictor for non-response in MDD is severity (Curry et al., 2006; Goodyer et al., 1997; Wilkinson et al., 2009). Indeed, Perez and colleagues (2018) found that both the frequency and distress of PEs were significantly and highly correlated with depression and anxiety scores. The authors found that these correlations remained strong even after an initial period of treatment, which was taken as indicative of a poorer long-term prognosis. Moreover, subclinical PEs in patients with MDD have been found to be strong indicators of relapse and resistance to both pharmacological and psychological interventions (Perlis et al., 2010; Wigman et al., 2014). Wigman and

colleagues (2014) found that even after controlling for baseline severity of depression, patients who more often endorsed experiencing PEs (without meeting criteria for psychotic depression) were 7 times less likely to achieve remission, and almost 4 times more likely to relapse. Perlis and colleagues (2010) also found that the endorsement of as little as one PE was sufficient to associate with non-response to 4 different types of antidepressants. The authors highlight the subtlety of these findings; while a single PE is not sufficient for a diagnosis of psychotic depression, it has a strong association with treatment outcome (Perlis et al., 2010). Interestingly, and in contrast to previous studies (Dudek et al., 2010; Frank et al., 2011), both studies here found that symptoms of bipolar spectrum disorder, or manic-like symptoms, failed to show such strong associations with resistance and relapse as those of psychotic symptoms (Perlis et al., 2010; Wigman et al., 2014). This subtle difference may be important for guiding therapeutic choice.

Taken together, these results demonstrate the importance of appropriate recognition of PEs in treating depression, and their potential affect on treatment efficacy. PEs may require tailored treatment to incorporate aspects of psychosis or trauma treatment to augment therapeutic response, which are currently unavailable in IAPT clinics (Perez et al., 2018). The distress associated with untreated PEs could result in a worsening of a patient's depressive condition, despite adequate treatment for this primary condition (Perez et al., 2018), thus rendering improvement limited. No study to date has, however, investigated the prognostic potential of subclinical PEs in clinical samples of depressed adults or adolescents, where response has been defined empirically.

*Objectives and Hypotheses*

The objective of this next piece of work was therefore to conduct a secondary analysis on the full IMPACT sample, investigating the prognostic ability of a battery of baseline demographic and clinical characteristics on empirically-derived trajectory class membership of this clinical sample of adolescents, as outlined in chapter 1. Prior literature would suggest that a number of demographic and clinical variables discussed above should be included in the model and may differ between classes in prevalence. Subclinical PEs were a predictor of interest above the battery of clinical variables noted for importance. We specifically hypothesised that subclinical PEs at baseline would be associated with unfavourable trajectory classes.

**Methods**

*Study Design and Size*

In chapter 1, the best fitting trajectory model was selected and the information on most likely class membership was saved for all patients. This chapter builds on this work, by investigating a battery of baseline demographic and clinical characteristics of trajectory class membership. Consequently, full details of the study design, setting and participant criteria are outlined in chapter 1.

*Variables*

Symptom trajectory class membership was defined through growth mixture modelling using the self-reported Mood and Feelings Questionnaire (MFQ) score across all time-points, as outlined in chapter 1. This is a 33-item Questionnaire (Burleson Daviss et al., 2006) of depressive symptomatology covering the past 2 weeks. MFQ items were measured on a 3-point scale (almost never, sometimes, often/almost always). Total scores with a theoretical range of 0-66 were used in GMMs. Higher scores indicated more severe depressive symptoms and were positively correlated with greater psychosocial impairment (Goodyer et al., 2017). As depression severity was the variable with which the two groupings were defined, MFQ was not a variable of interest for these predictor analyses.

A number of baseline clinical variables available in the IMPACT cohort were investigated for their potential predictive value over class membership. These included measures of anxiety, obsessional traits, overall psychiatric symptomatology/function, lifetime suicidal attempts, non-suicidal self-injury, and psychotic symptoms. Derived sum scores from self-report measures were used for anxiety (the Revised Children's Manifest Anxiety Scale, [RCMAS]; Reynolds & Richmond, 1978), obsessionality (the short Leyton Obsessional Inventory for adolescents, [LOI]; Bamber, Tamplin, Park, Kyte, & Goodyer, 2002) and overall psychiatric symptomatology/function (the Health of the Nation Outcome Scales for Children and Adolescents, [HoNOSCA]; Gowers et al., 1999). Lifetime suicide attempts were defined as a binary variable (yes, no) from data derived from the Columbia Suicide Severity Rating Scale [CSSRS]; Posner et al., 2011). Lifetime non-suicidal self-injury (NSSI) was measured using the self-report Risk and Self Harm Inventory (Vrouva, Fonagy, Fearon, & Roussow, 2010). The Kiddie-Schedule for Affective Disorders and Schizophrenia [k-SADS-PL]; Kaufman et al., 1997) interview was used to assess the presence of psychiatric diagnoses and psychotic symptoms

at baseline. Comorbidity was defined on an ordinal scale, as the number of additional mental illnesses other than depression that met threshold criteria during interview. Psychotic symptoms were also defined on an ordinal scale (absent, present: subthreshold, or present: threshold), as answering positively to either of the two screening questions for psychosis present in the k-SADS-PL interview.

*Statistical analyses*

The class allocations from chapter 1 were saved and then used to conduct analyses of associations in a separate step. This prevented predictors from modifying trajectory solutions.

The first goal was to explain how the classes differed on demographics and baseline clinical characteristics. For this, a series of univariate analyses were conducted (chi-square, or t-tests) to determine whether there were significant differences between classes on a battery of variables.

A second goal of this analysis was to determine whether baseline clinical characteristics could predict whether a patient would belong to class 1 or class 2. For this, a logistic regression was conducted to determine how well the baseline clinical variables collectively predicted class membership and assess the relative contribution of psychotic symptoms. Gender and age were also included to control for any variation associated with these factors. A chi-square test was used to determine whether the model was significantly better than the null at predicting class membership, and the R-squared statistic indicated how much variance the model explained. Logistic regressions involve comparison to a specific reference class. We took class 2 as the reference class, which was the most likely class allocation for patients.

Receiver-operating characteristics (ROC) curve analysis was conducted to calculate the area under the curve (AUC) and evaluate the discriminatory ability of the model.

**Results**

The characteristics of patients following each trajectory of the best-fitting two-class piecewise model are described in Table 1.

*Table 1: Characteristics of patients following the two latent trajectories*

| | **Class 1: Halted-improvers** | | **Class 2: Continued-improvers** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | *Mean(n)* | *S.D(%)* | *Mean(n)* | *S.D(%)* | $X^2/t$ | *p* |
| **Demographics** | | | | | | |
| *Female* | 63 | 85.1% | 285 | 72.8% | 4.955 | .026 |
| *Age* | 15.7 | 1.3 | 15.6 | 1.4 | 0.459 | .647 |
| *Region* | - | - | - | - | 2.035 | .361 |
| *East Anglia* | 24 | 32.4% | 161 | 41.2% | - | - |
| *North London* | 22 | 29.7% | 105 | 26.9% | - | - |
| *North-West England* | 28 | 37.8% | 125 | 32.0% | - | - |
| *Ethnicity(white)* | 65 | 87.8% | 325 | 83.1% | 1.024 | .312 |
| *IMD* | 29.2 | 19.3 | 26.2 | 16.9 | 1.250 | .214 |
| **Baseline clinical characteristics** | | | | | | |
| *RCMAS* | 42.3 | 6.7 | 40.7 | 7.3 | 1.863 | .065 |
| *LOI* | 11.8 | 5.6 | 9.6 | 5.1 | 3.124 | .002 |
| *Suicidal thoughts* | 69 | 93.2% | 345 | 88.2% | 1.600 | .206 |
| *Suicidal attempts* | 28 | 37.8% | 131 | 33.5% | 0.519 | .471 |
| *NSSI* | 53 | 71.6% | 218 | 55.8% | 6.443 | .011 |
| *HoNOSCA* | 19.9 | 6.3 | 18.3 | 6.0 | 2.018 | .046 |
| *Comorbidity* | - | - | - | - | 15.46 | .004 |
| *1* | 26 | 35.1% | 121 | 30.9% | - | - |
| *2* | 18 | 24.3% | 59 | 15.1% | - | - |
| *3* | 3 | 4.1% | 5 | 1.3% | - | - |
| *4* | 1 | 1.4% | 0 | 0% | - | - |
| *Psychotic symptoms* | - | - | - | - | 2.024 | .363 |
| *Subthreshold* | 16 | 22.5% | 87 | 23.6% | - | - |
| *Threshold* | 10 | 14.1% | 32 | 8.8% | - | - |
| **Treatment characteristics** | | | | | | |
| *Treatment arm:* | - | - | - | - | 2.463 | .292 |
| *BPI* | 28 | 37.8% | 127 | 32.5% | - | - |
| *CBT* | 27 | 36.5% | 127 | 32.5% | - | - |
| *STPP* | 19 | 25.7% | 137 | 35.0% | - | - |
| *Baseline SSRI prescription* | 10 | 13.5% | 87 | 22.3% | 2.877 | .090 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

Demographic characteristics showed that the distribution of males and females was significantly different between the classes. Eight-Five percent of class 1 were female, compared with only 73% of class 2. Conversely, 15% of class 1 were male, compared with 27% of class 2. No other demographic characteristics were significantly different between groups.

In terms of baseline clinical characteristics, class 1 on average showed higher LOI and HoNOSCA scores, indicating more obsessional traits and more severe overall psychiatric symptomatology and function than class 2 at baseline. Furthermore, 72% of patients in class 1 reported non-suicidal self-injury at baseline, compared with only 56% of class 2, and significantly more of class 1 were diagnosed with one or more additional comorbidities, compared with class 2. The two classes did not significantly differ on treatment arm or SSRI prescription at baseline.

*Predictors of Trajectory Class Membership*

Investigation of the correlations between clinical variables included in the model revealed the strength of collinearity between variables. Results are shown in Table 2. Variance inflation factor scores however, indicated that multicollinearity for these data was not a concern (all VIF values <10, all tolerance values >.2, see Table 1 in Appendix 2A). Further, the data also met the assumption of independent errors (Durbin-Watson value= 2.08, p=.42), and linearity of the logit using the Box-Tidwell test, indicating that logistic regression is appropriate.

Results from the logistic regression are shown in Table 3. This model produced a significant improvement in the fit of the model over the constant ($X^2(4)=46.03$, $p<.001$). This model explained 5.4% of the total variance in class membership allocation (Cox and Snell $R^2=$ 0.054). It was found that the presence of one or more co-morbid diagnoses significantly predicted a higher probability of membership to class 1 (halted-improvers) compared with class 2 (continued-improvers; Table 3). With each increasing number of co-morbid diagnoses, the odds of a patient belonging to the “halted-improvers” class compared to the “continued-improvers” class increased by a factor of 1.4.

*Table 2: Correlation matrix of relationships between clinical variables.*

| | Age | RCMAS | LOI | HONOSCA | Gender | Suicide Attempts | NSSI | Comorbidity | Psychotic Symptoms |
|---|---|---|---|---|---|---|---|---|---|
| **Age** | - | 0.01 | -0.01 | 0.07 | -0.14** | 0.05 | 0.05 | 0.04 | -0.03 |
| **RCMAS** | - | - | 0.51*** | 0.11* | -0.16*** | 0.07 | 0.18*** | 0.14** | 0.13** |
| **LOI** | - | - | - | 0.12* | -0.07 | 0.04 | 0.12** | 0.26*** | 0.13** |
| **HONOSCA** | - | - | - | - | 0.04 | 0.15** | 0.20*** | 0.20*** | 0.14** |
| **Gender** | - | - | - | - | - | 0.13** | 0.07 | 0.04 | 0.05 |
| **Attempts** | - | - | - | - | - | - | 0.35*** | 0.05 | 0.17*** |
| **NSSI** | - | - | - | - | - | - | - | 0.05 | 0.21*** |
| **Comorbidity** | - | - | - | - | - | - | - | - | 0.15** |
| **Psychotic Symptoms** | - | - | - | - | - | - | - | - | - |

*RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents*
*Pearson correlation and Spearman's rank correlation were used where appropriate.*

**p<.05*

*** p<.01*

**** p<.001*

*Table 3: Baseline predictors of trajectory class membership: clinical characteristics.*

| | ***Class 1: Halted-improvers*** | |
|---|---|---|
| | *OR* | *95%CI* |
| *Gender* | 0.50 | 0.23-1.02 |
| *Age* | 1.02 | 0.84-1.26 |
| *RCMAS* | 0.98 | 0.94-1.03 |
| *LOI* | 1.06 | 0.99-1.12 |
| *HONOSCA* | 1.03 | 0.98-1.08 |
| *Attempts* | 0.98 | 0.53-1.79 |
| *NSSI* | 1.77 | 0.95-3.41 |
| *Psychotic Symptoms* | 0.97 | 0.64-1.45 |
| *Comorbidity* | 1.40* | 1.00-1.96* |

*RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents*
*Class 2 is taken as the reference class.*

**<.05*

Standardised residuals of the final model were inspected to assess for whether the model contained any outliers of concern. There were no cases in this sample of residuals larger than ±2.58, and fewer than 5% (21 cases) of the total sample showed residuals greater than ±1.96. Therefore the model can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases are exerting undue influence over the model. It has been recommended to investigate cases where leverage values are greater than twice (Hoaglin & Welsch, 1978) or three times (Stevens, 2009) the average. Four cases in the final model indicated a potential problem with 3 times greater than average leverage for the model, and a further 32 cases showed double the average leverage. However, the Cook's distance of each case was also inspected to assess whether removal of each case would significantly change the fit of a model. There were no cases in the final model with a Cook's distance greater than 1 (Cook & Weisberg, 1982), indicating that the fit would not significantly change upon removal of these cases.

*Receiver-operating characteristics*

Receiver-operating characteristics (ROC) curve analysis was conducted to calculate the area under the curve (AUC), to evaluate the discriminatory ability of the regression model. The AUC for the full model was 0.69 (95% CI, 0.62-0.76), with a mean sensitivity of 66% and mean specificity of 53%. (Figure 1). Conventionally it is accepted that the AUC in a ROC analysis should be >0.75 to be of potential clinical value (Fan, Upadhye, & Worster, 2006). The AUC falls short of this threshold. Therefore sensitivity and specificity of this model is not sufficient to provide a valid predictive model of class membership.

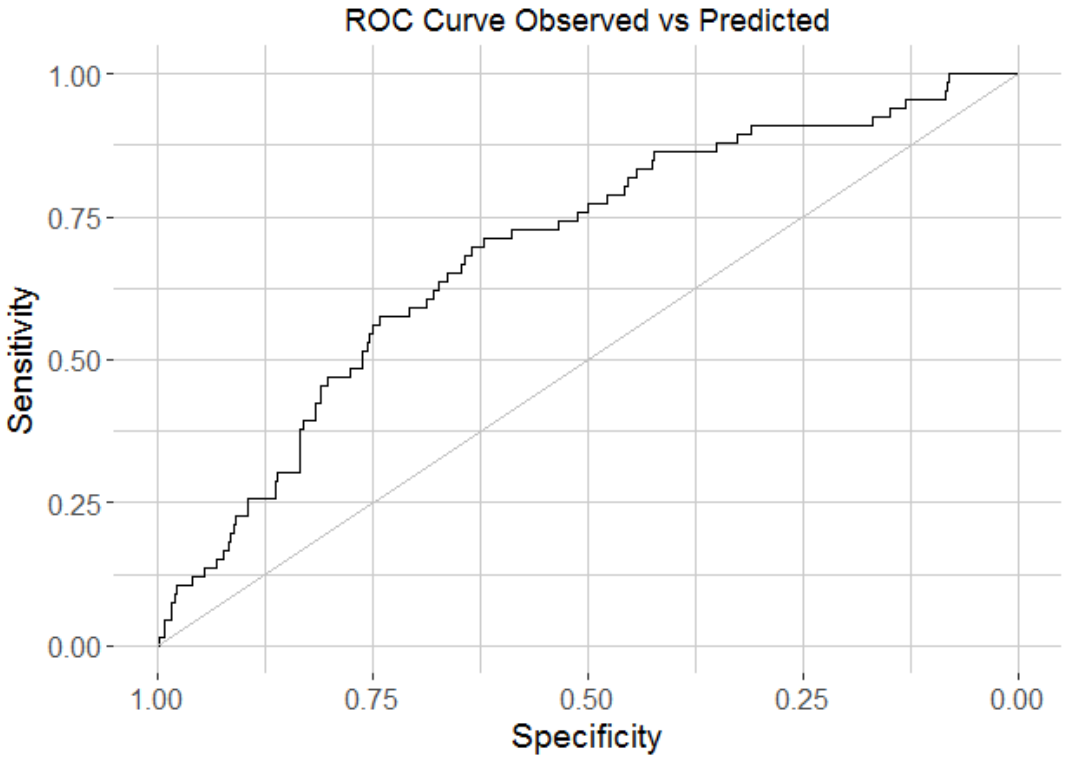



*Figure 1. ROC Curve.*

## Discussion

*Key Results and Interpretation*

This work extended that of Chapter 1, by describing the characteristics of trajectory classes, and investigating potential baseline predictors of class membership. In addition to baseline depression severity, which was a defining feature of the class trajectories discussed in Chapter 1, univariate analyses showed that the group of halted-improvers were disproportionately female, had higher current obsessionality, self-harming behaviours, more psychiatric symptomatology and impairment and experienced greater comorbidity at baseline than continued-improvers. These findings are consistent with prior reports (Curry et al., 2006; Goodyer et al., 1997; Wilkinson et al., 2009) suggesting that indicators relating to severity and clinical complexity are associated with poorer outcome. However, the gender finding reported here contrasts some prior research that found either a positive association with female gender and better outcomes (Braun, Gregor, & Tran, 2013) or no relationship at all (Cuijpers, Weitz, et al., 2014; Sramek et al., 2016). Interestingly, and in contrast to previous reports (Curry et al., 2006; Frank et al., 2011; Perez et al., 2018; Perlis et al., 2010; Scott et al., 2019; Souery et al., 2007; Wigman et al., 2014; Wilkinson et al., 2009), suicide attempts and psychotic symptoms were not significantly different between classes at the univariate level. The reason for dissonance with prior research may be methodological; relating to the difference in defining response classes, or the significantly shorter follow-up of those studies, especially terminating before 18 weeks, as this denotes a distinct divergence in outcome in this clinical cohort.

Given meaningful correlations between the univariate predictors, it is possible that some 'significant' predictors of poor outcome were due to confounding. A multivariate predictive model of demographic and clinical characteristics showed that only the presence of comorbidities provided a significant independent predictor. This is quite an interesting result, given that the continuous measures of anxious and obsessional traits were insignificant. Dichotomous variables like comorbidity are often criticised for reducing the power of effect, and so it would have been more likely for us to find insignificance with the comorbidity variable. However, it is likely that the variance associated with both RCMAS and LOI outcomes were explained by a comorbid diagnosis. Reaching diagnostic criteria for a second illness marks the upper extreme of these continuous measures, so it is possible that this allowed for a distinction between classes to occur with this variable. These results

support previous findings that suggest that depressed patients with comorbid conditions struggle to achieve adequate remission with these therapies (Bagby, Ryder, & Cristi, 2002; Brent et al., 1998; Curry et al., 2006; Kemp et al., 2008). Treatment plans too focussed on depressive symptoms alone may therefore limit the effectiveness of treatment in the longer term. A plateauing effect in symptom trajectories like those observed in this study may result, and increase the risk of relapse. Future work may benefit from designing randomised controlled trials investigating the effectiveness of adjunctive therapies (versus depression treatment alone) for patients with comorbidities over time.

*Alternative explanations*

An important facet of this study was the collapsing of treatment arms in the IMPACT trial, to create a single group of patients receiving psychological treatment. There is evidence from empirical studies that different treatments have proved more effective for different trajectory classes (Cuijpers et al., 2005; Gueorguieva et al., 2011; Uher et al., 2010). Furthermore, some studies have associated certain demographic and clinical characteristics with moderating response to different treatment modalities. For example, the severity of a patient's illness was found to moderate the efficacy of combination treatment in adolescents (Curry et al., 2006) and adolescents with comorbid anxiety have shown differential efficacy rates to psychological treatments (Brent et al., 1998). Consequently, we cannot exclude the possibility that our null results for predictors other than comorbidity might be due to differential treatment effects within trajectory groupings moderating outcome. However, a number of studies into predictors and moderators of treatment outcome have concluded that the majority of demographic and clinical variables that associate with treatment outcome tend to act as non-specific predictors, rather than treatment-specific moderators of outcome (Curry et al., 2006; Frank et al., 2011). There is a great difficulty in recruiting sample sizes large enough to detect significant differences for moderation analyses (Frank et al., 2011), and the sample size per treatment arm in IMPACT was not sufficient to carry out such analyses with adequate reliability.

*Limitations*

Our results must be viewed in the context of several limitations of this work. Firstly, our specific hypothesis that subclinical psychotic symptoms would independently predict poorer outcome was unsupported. This suggests that in adolescent depression, symptoms of

psychosis do not have an overriding, unique effect on the trajectory of depressive symptoms during and after psychological treatment. However, as our analysis was secondary, our measurement of psychotic features was limited to the two questions posed in the k-SADS-PL screening interview relating to hallucinations and delusions. A more in-depth measure of psychotic experiences was not available in this current work. While research has demonstrated that endorsing as little as one single psychotic symptom is enough to observe a difference in remission rates (Perlis et al., 2010), thus providing support for our methodology, it is possible that our null results are a reflection of a lack of power of the measurement tool to detect differences between classes. Indeed, while Perlis and colleagues (2010) used a similarly simplistic method in their work, theirs still included assessment of symptoms beyond hallucinations and delusions (Perlis et al., 2010). The inclusion of a scale such as the Community Assessment of Psychic Experiences (CAPE-P15; (Perez et al., 2018) or the Psychotic-like Experiences Semi-Structured Interview (PLIKSi; Niarchou, Zammit, & Lewis, 2015) may have allowed a more in-depth investigation of subclinical psychotic symptomatology and subtle differences that might exist between the two trajectory classes. Perez and colleagues' (2018) work has advocated the ease with which the CAPE-P15 can be administered in a clinical setting. As such, we would encourage future primary work to utilise such instruments to allow for a fuller investigation of psychotic symptoms in depressed adolescents. Related to measurement limitations, it is important to note that both suicide attempts and non-suicidal self-injury were measured on lifetime scales. It might be that these variables only show a distinct association with response trajectories if they are experienced near the beginning of treatment.

While our sample size was relatively large for clinical trials in adolescent depression, compared with other studies (Brent et al., 1997; Emslie et al., 2002; Goodyer et al., 2008), as noted in chapter 1, it was quite modest for growth mixture modelling. When discussing markers of a minority group in this respect, the frequency of those markers, even in a large overall sample, becomes notably reduced. Thus, one must mention sample size as a limitation of this research. For instance, our sample size reduced by half when considering patients with comorbid disorders, and these disorders ranged from behavioural (29% of those with at least 1 comorbidty), anxiety-related (77% of those with at least 1 comorbidity) or other conditions such as eating disorders or full psychosis (12% of those with at least 1 comorbidity). Given the small percentages associated with each disorder category, it would

not have been good practice to further divide comorbidity into its respective conditions or categories, thus all comorbidities were grouped into a single variable. Our approach for this variable worked to maximise power while retaining clinical relevance for the variable. However, we cannot draw conclusions over which specific conditions associated with halted-improvement, only that with increasing clinical complexity, as defined by the presence of an increasing number of comorbid conditions, the greater the likelihood to follow that trajectory. That said, we could speculate that anxiety may have been driving this effect, given that the majority of comorbid disorders were anxiety-related (77% of those with at least 1 comorbidity). However, further research that is adequately powered to address this question is needed before definitive conclusions can be made. At an individual study level, this scale of clinical research presents a significant barrier. Realistically, this is likely to require large-scale international collaboration in order to gain sufficient numbers, but such collaborations are very much a focus of current efforts to advance our understanding of mental health disorders (Schmaal et al., 2016, 2017a).

*Reflections on Methodology*

It is important to consider the robustness of our results, in light of a number of methodological choices. Firstly, comorbidity was assessed using the Schedule for Affective Disorders and Schizophrenia for School Aged Children (k-SADS-PL), which uses the Diagnostic and Statistical Manual version 4 (DSM-IV) criteria for diagnosis (APA, 2000), rather than the International Classification of Diseases (ICD)(World Health Organisation, 1992). It has been suggested that the use of the DSM often leads to an increased number of comorbidities reported than when using the ICD (Goodyer et al., 1997; Tyrer, 2014). It is therefore possible that the number of comorbidities in our sample is an overestimation of prevalence rates that would be normally reported in clinic. However, these diagnostic tools are optimised for different purposes: the ICD is preferred in clinical settings, as its definitions of conditions are more descriptive, and thus allow flexibility for clinical discretion (Tyrer, 2014). Conversely, the DSM has a distinct advantage in research settings, as it relies on operational criteria for diagnosis. Consequently, the use of DSM would minimise any clinician-related bias that might have emerged from an overreliance on clinical discretion using the ICD. Furthermore, while actual diagnoses may be overrepresented, we would argue that these numbers are still reflective of clinical complexity even if they did not meet threshold under another diagnostic tool, and thus represent a reliable marker of unfavourable trajectories.

Secondly, the analytical methods employed in this current work were conducted at the group-level; that is, we did not test the accuracy at which the model can categorise new patients. As stated in Chapter 1, our sample size was not large enough to divide into training and test sets for GMM, and data from a separate yet comparative cohort was unavailable. However, one previous GMM study has shown the potential for empirical approaches to predict outcome at an individual level (Stulz et al., 2010). The authors found that a predictive model of characteristics, including consideration of baseline measure of anxiety, history of antidepressant medication, patient reported depression severity and family function, was able to provide an accuracy of 61% in categorising new patients (Stulz et al., 2010). This accuracy level is not sufficient to give confidence in clinical application, but more studies are needed that incorporate such approaches to validate the trajectory classes described in chapter 1, and further test the accuracy of our clinical model to categorise new patients.

Finally, it is important to note that the model produced here is comprehensive only in respect to the demographic and clinical characteristics that were available for secondary analysis from the IMPACT trial. There are a number of aspects of behaviour that were not measured in IMPACT, but have been found to significantly contribute to a persons' ability to engage in and respond to treatment. These include, but are not limited to, personality traits and coping styles (Thibodeau et al., 2015), cognitive distortions and family-related factors (Curry et al., 2006) and treatment expectations (Curry et al., 2006). A much fuller inclusion of social and environmental factors should be considered in future primary investigations of predictors and moderators of response. The results however do suggest that including non-depressive symptoms in a more multidimensional longitudinal analysis to further disaggregate the behavioural phenotypes over time may be of value. Such an analysis may improve the signal of putative predictors for treatment response.

Overall, this current work has shown that the greater the clinical complexity of the patient, as measured by the presence of comorbidity, the more likely that patient is to experience halted-improvement trajectory of symptom change during and following a course of psychological treatment. It is important for future work to replicate these findings in primary datasets designed for this purpose, and provide an assessment of model accuracy at an individual level. However, the predictive model itself explained very little of the variance

between these two classes (5.4%), with poor sensitivity and specificity. This small percentage of explained variance suggests that, despite the methodological limitations of this work, it is likely that demographic and clinical observations alone are insufficient for fully predicting symptom change over time. Indeed, it is possible that the mere self-report nature of these measures limits their reliability as predictors (Nierenberg & DeCecco, 2001). In line with the Research Domain Criteria (RDoC) framework (Insel et al., 2010), future work should investigate the significance of biological predictors of such trajectories, and incorporate a multidimensional approach to fully assess factors that influence trajectory membership (Kennedy et al., 2012; Kemp et al., 2008).

**Appendix 2A**

*Table 1. Variance inflation factor and tolerance scores for logistic regression models.*

| | **VIF** | **Tolerance** |
|---|---|---|
| *Gender* | 1.06 | 0.94 |
| *Age* | 1.03 | 0.97 |
| *RCMAS* | 1.36 | 0.74 |
| *LOI* | 1.38 | 0.73 |
| *HoNOSCA* | 1.15 | 0.87 |
| *Suicidal attempts* | 1.18 | 0.85 |
| *NSSI* | 1.19 | 0.84 |
| *Psychotic symptoms* | 1.09 | 0.92 |
| *Comorbidity* | 1.17 | 0.86 |

*RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents.*

## Chapter 3: Neurological predictors of trajectory classes.

The analyses conducted thus far have illustrated that clinical observations are insufficient in predicting likely class membership of a patient's symptom change over time. Clinical characteristics only explained 5% of the variance between trajectory classes in adolescent MDD. Our results agree with the Research Domain Criteria, that posit that symptoms alone may not reflect underlying disease processes with enough accuracy to act as predictors of overall response (Insel et al., 2010). These authors call for diagnostic and prognostic work to look beyond symptom classifications and consider the role of various biological components that affect mental health.

Maladaptive thoughts, feelings and behaviours are all governed by brain circuitry. Consequently, neural systems that generate the core symptoms of depression are a clear target for investigation. Over a decade of studies, reviews and prominent meta-analytic papers such as the ENIGMA-MDD consortium (Schmaal et al., 2016, 2017a) have shown that in depressed cases, significant structural and functional differences exist in regions that regulate disease-relevant functions such as emotion, reward processing and cognitive biases (Belzung, Willner, & Philippot, 2015; Gong & He, 2015; Miller, Hamilton, Sacchet, & Gotlib, 2015). Variability within these regions has also shown promise in differentiating between those who respond well and poorly to treatment (Fu, Steiner, & Costafreda, 2013; Lener & Iosifescu, 2015; Phillips et al., 2015). Therefore, my next line of work will investigate whether trajectory classes demonstrate neurological differences in specific regions of interest, and whether these differences possess predictive value. While I will focus on the investigation of brain structure in this chapter, I acknowledge that altered structure is likely only one of a multitude of biological abnormalities present in depression, including aberrancies in brain function (Forbes et al., 2010; Fu et al., 2013; Gong & He, 2015; Miller et al., 2015). Consequently, I have drawn on functional imaging studies for supportive evidence. My investigations will also focus on the cortical regions emerging as important from prior work, and I discuss each region of interest in turn below.

*Anterior Cingulate Cortex (ACC)*

The network that governs emotion incorporates an intricate group of regions in the brain, including different sub-regions of the ACC (Boes, McCormick, Coryell, & Nopoulos, 2008;

Kross, Davidson, Weber, & Ochsner, 2009; Yoshimura et al., 2010) and the medial prefrontal cortex (mPFC; Lemogne et al., 2009; Yoshimura et al., 2010), governing emotional stability, internal self-referential processes and rumination (defined as deep, considered thought of cause and consequence; usually about one's own feelings or distress). The ACC in particular has a diverse range of functions, involving emotion processing, but also reward processing and cognitive control (Bush, Luu, & Posner, 2000; Devinsky, Morrell, & Vogt, 1995; Mohanty et al., 2007). Various subdivisions of the ACC have been documented as aberrant in MDD and moreover, specific regions have been associated with different symptomatology (Grimm et al., 2009). For instance, the pregenual ACC is known as the "affective division". Activation of this region has been suggested to induce sadness, resulting in associations with depression and its severity (Bush et al., 2000; Grimm et al., 2009). On the other hand, the dorsal ACC is known as the "cognitive division". Reduced blood flow in this region has been linked to impairments in attention and executive functioning in depression (Fossati, Ergis, & Allilaire, 2002), and control over the affective subdivision (Mayberg et al., 1999). Taken together, the ACC appears to function in the regulation of emotional and cognitive processes, which are specifically implicated in depressive symptomatology.

Structurally, reductions of the ACC grey matter volume (GMV) in depressed cases have been largely supported by a number of meta-analytic studies to date (Arnone, McIntosh, Ebmeier, Munafò, & Anderson, 2012; Bora, Fornito, Pantelis, & Yücel, 2012; Koolschijn, van Haren, Lensvelt-Mulders, Hulshoff Pol, & Kahn, 2009; Lai, 2013). In particular, these meta-analytic studies have collectively interrogated the reliability of these findings, showing that reductions in ACC volume are apparent across different analytical approaches (Bora et al., 2012), and are not a result of data duplication or specific age of patients (Arnone et al., 2012). However, the magnitude of affect did vary by these variables. Abnormalities in ACC regions appear to be a trait-related feature (van Eijndhoven et al., 2013) and also a predisposing factor to depression rather than a pathological consequence of the illness itself. For example, the ACC showed a strong contribution to models that differentiated adolescents who later developed depression from those who remained well (Foland-Ross et al., 2015). However, some authors have suggested ACC differences may be a result of medication, as there was a failure to replicate the reduced ACC volumes in one meta-analysis of medication-naïve patients (Zhao et al., 2014).

Surface-based investigations of brain structure have largely supported the literature of ACC GMV in depressed cases. The major difference between surface-based and voxel-based methods is that the former map brain structure to the cortical surface, while the latter map structures based on the intensity of a given voxel. Consequently, agreement between the two imaging methods is encouraging, because it validates voxel-based findings with more anatomically accurate methods (Fischl & Dale, 2000). The ENIGMA consortium conducted a meta-analytical investigation of cortical thickness and surface area differences, using surface-based methods (Schmaal et al., 2017a). They found that adult patients displayed cortical thinning in the orbitofrontal cortex (OFC), anterior (ACC) and posterior cingulate (PCC), insular cortex and temporal lobes. While the OFC demonstrated the largest effect size in this study, their parcellation method meant that the medial wall of the OFC contained the subgenual ACC. Furthermore, cortical thinning in prefrontal and ACC regions have been found to associate with specific symptomatology, including irritability, fatigue (Lener et al., 2016), and suicidal behaviour (Wagner et al., 2012). However, the direction of effect is somewhat debated. While the ENIGMA consortium reported that cortical thinning was apparent even in first episode patients (Schmaal et al., 2017a), other studies have reported that first-episode depression is associated with cortical thickening in the ACC (van Eijndhoven et al., 2013). It is possible that thicker cortices might predispose individuals to depression, while over the course of the illness, physiological processes of the disease state result in accelerated cortical thinning (Foland-Ross et al., 2015). Nevertheless, the ACC appears to be an important biological substrate for the experience of depression and it is likely to be a viable region for further investigation in treatment response.

A number of functional imaging studies have provided support for the prognostic ability of the fronto-cingulate pathways in predicting treatment outcome for patients (Fu et al., 2013; Pizzagalli, 2011). A meta-analysis of 23 studies advocated that hyperactivity of the rostral ACC is a robust predictor of treatment response in depression (Pizzagalli, 2011). The effect was evident across a range of functional imaging (EEG/MEG, fMRI, PET and SPECT), and also generalized to a number of depressive treatments including pharmacological, rTMS and sleep deprivation. However, one of the difficulties of this field is that there is a great deal of methodological variability across studies. These include brain parameters, imaging technique, study design and type of therapy. Indeed, findings of whether ACC activity is predictive of response to CBT have been mixed; with disagreement on whether hyper or

hypo-activation within the ACC is predictive of a positive CBT response (Fu et al., 2008; Konarski et al., 2009; Siegle, Carter, & Thase, 2006). Also, some other studies found that the ACC did not show predictive value (Ritchey, Dolcos, Eddington, Strauman, & Cabeza, 2011). It is plausible that predicting clinical response to less physical treatments with fMRI may require more power (Costafreda, Chu, Ashburner, & Fu, 2009), but more research in psychological therapies is necessary before definitive conclusions can be drawn.

The field of research investigating structural neurological predictors of response to treatment is small. However, there is a pattern emerging that responsive patients tend to display qualitatively different neural structure to those who are unresponsive (Liu et al., 2012). For example, the extent of ACC GMV reductions observed in depressed patients has been shown to associate with resistance to treatment (Chen et al., 2007). Faster improvement was observed in those patients with greater ACC volume and greater ACC task-based activation (Chen et al., 2007). Interestingly, this study found dissociation between subregions of the ACC. The affective division showed predictive value for treatment response to antidepressants, while the midcingulate regions related to baseline severity (Chen et al., 2007). However, for their functional work, the prognostic value of the subgenual ACC was only observed through ROI analyses. When stricter corrections for multiple comparisons were applied, no differences in function between good and poor responders were reported (Chen et al., 2007), suggesting that structural investigations may be a more promising avenue in this field. White matter volume in the ACC has also been shown to discriminate between treatment-resistant and treatment-sensitive depression with comparable accuracy to GMV (Liu et al., 2012).

Other structural work employing whole-brain analyses have supported the significance of the ACC in predicting response (Costafreda et al., 2009). Costafreda and colleagues (2009) found that ACC neuroanatomy provided the strongest weighting in prediction models, although it was noted that the posterior cingulate was also a prominent structure for the prediction model. Their models reported 88.9% accuracy in correctly predicting clinical remission to antidepressants. In contrast, the accuracy of whole-brain structure in diagnosis was only 67.6%, prompting the authors to suggest that structural neuroanatomy may have better utility as a prognostic, rather than diagnostic marker for depression (Costafreda et al., 2009). This assertion was supported by inpatient research (Frodl et al., 2008). These authors

found no difference between patients and healthy controls in the ACC volume but a significant negative correlation between ACC volume and number of hospitalizations was observed. It is important to note however, that the number of hospitalizations is a somewhat crude measure for clinical response. Furthermore, in echo of the functional literature (Pizzagalli, 2011), Costafreda and colleagues (2009) stated that their sample may have been too small to have adequate power to detect whether structural differences predict response to psychological treatments, so our understanding of this line of treatment remains incomplete.

There is a growing body of research investigating the link between structure and function in the brain (Honey, Thivierge, & Sporns, 2010). However, the dearth of literature of the prognostic potential of brain structure to psychological therapies makes generating hypotheses regarding the direction of the expected influence that brain structure would have on treatment efficacy difficult. However, one study theorized that regional cortical thinning could impair a patient's ability of adaptive rumination, and thus ability to engage in psychological therapies that utilize such skills (Späti et al., 2015). These authors found that functional connectivity of the subgenual ACC correlated with thickness of the PFC (Späti et al., 2015) and moreover, this increased connectivity was associated with higher scores on adaptive rumination. Cortical thinning of the ACC present prior to psychological treatment may therefore predict a poor outcome to treatment in adult patients.

*Prefrontal Cortices*

While the ACC has a large corpus of work supporting its involvement in MDD, many studies emphasise that alterations in depression span further than a single region (Grieve, Korgaonkar, Koslow, Gordon, & Williams, 2013; Schmaal et al., 2017a; Zhao et al., 2014). Three key regions of the prefrontal cortex (PFC) also appear important in MDD and treatment response; namely, the medial PFC (mPFC), dorso-lateral PFC (dlPFC) and the orbitofrontal cortex (OFC) (Kaiser, Andrews-Hanna, Wager, & Pizzagalli, 2015; Zheng et al., 2015). These regions play different roles within the affective network, acting as moderators over areas that generate emotion, directing cognition and processing reward (Ochsner et al., 2004). Consequently, these regions show different aberrancies in MDD.

The mPFC has been associated with up-regulation of negative emotion, and is particularly involved in internal processing (Ochsner et al., 2004). Hyperconnectivity demonstrated in this region in MDD could, along with aberrancies in the OFC, contribute to low mood and increased self-referential symptoms in MDD (Kaiser et al., 2015; Zheng et al., 2015). The degree of aberrant functional connectivity was further found to directly correlate with cortical thickness of the dorso-mPFC in MDD (van Tol et al., 2014). However, studies have illustrated functional changes occur in prefrontal regions following treatment (Chuang et al., 2016), and medication status of patients at the time of scanning in this trial was a confound (van Tol et al., 2014). In this context, care must be taken to consider medication status of patients when research question concern aetiology or prediction (van Tol et al., 2014).

The OFC and dlPFC have also exhibited reduced connectivity in depression (Kaiser et al., 2015; Zheng et al., 2015). The activations of the OFC have been associated with the down-regulation of amygdala response to negative emotions (Ochsner et al., 2004), and the dlPFC is particularly involved in the control over network that process external stimuli (Seeley et al., 2007). Therefore, reduced activity in these regions could underpin attentional biases and low mood observed in depression (Kaiser et al., 2015; Zheng et al., 2015). Furthermore, the intricate involvement of the OFC in reward processing (Elliott, Dolan, & Frith, 2000) may provide a plausible substrate underlying anhedonia (Gorwood, 2008), one of the core symptoms of depression (APA, 2013).

Structural imaging modalities have provided complements for the aberrant functioning of prefrontal regions in MDD. As previously mentioned, the ENIGMA consortium found compelling evidence for the importance of the OFC in MDD, whether first episode or recurrent (Schmaal et al., 2017a). The largest effect sizes between depressed cases and healthy controls were reported for thinning of the OFC (Schmaal et al., 2017a). Foland-Ross and colleagues (2015) also found that, while ACC thickness contributed to the classification of adolescent girls who went on to develop depression, it was thinning of the OFC that provided the strongest weighting to these machine learning algorithms. These findings concur with effect sizes reported in volumetric meta-analyses (Koolschijn et al., 2009). Meta-analytic findings have also shown reports of robust GMV reductions in the dlPFC in medication-naïve patients (Zhao et al., 2014) as well as cortical thinning of this region (Späti et al., 2015; van Tol et al., 2014). Morphometric studies have associated this thinning of the

dlPFC and the OFC with reductions in neuronal size, and neuronal and glial cell densities (Rajkowska et al., 1999). Furthermore, one study found that OFC thinning correlated with serum cortisol levels in first-episode drug-naïve patients, suggesting a possible mechanism of observed atrophy (Liu et al., 2015). However, these structural findings are not wholly consistent; a study of familial risk concluded that thickening of the OFC and ACC regions identified high-risk individuals (Peterson et al., 2009). No longitudinal follow-up was conducted on these participants, so it is unknown if high-risk individuals subsequently developed the illness, which consequently, may explain the disparity between studies. Nevertheless, abnormalities in these prefrontal systems may provide significant contribution to the development of the depressive state (Kaiser et al., 2015).

Further variation within already abnormal structure might explain the heterogeneity we see between patients, and relate to their response to treatment. For instance, volume reductions in the dlPFC, along with reductions in the mid-cingulate, have been shown to correlate with depression severity (Chen et al., 2007; Li et al., 2010); the most consistently found behavioural predictor of non-response (Curry et al., 2006; Goodyer et al., 1997; Wilkinson et al., 2009). Volumetric analyses have also showed that greater OFC volumes associated with symptom improvement (Chen et al., 2007). Moreover, the extent of cortical thinning in the dlPFC has differentiated patients with and without risk of suicide (Wagner et al., 2012). This thinning in suicidal patients has been associated with observed decreases in the density of serotonin axons in this region (Austin, Whitehead, Edgar, & JE Janosky, 2002). Although, the small numbers of studies and the inclusion of patients who have already begun treatment within those studies, have limited meta-analytic conclusions of prognostic studies of structure (Fu et al., 2013). Fu and colleagues (2013) found only a trend towards significance of decreased grey matter volume in the dlPFC, associating with a poorer outcome to antidepressant treatment. The effect was also not strong enough to survive corrections for multiple comparisons. However, functionally, meta analytic results found that the positive association between subgenual ACC activation and clinical response to antidepressants extended to include the mPFC and OFC (Fu et al., 2013).

At an individual study level, grey matter volume reductions in the dlPFC have been noted in patients who fail to remit after 6-week antidepressant treatment (Li et al., 2010). Furthermore, predictive models of whole-brain structure were found to discriminate

between patients with treatment-resistant and treatment-sensitive depression with 82.9% accuracy, and that the most important frontal regions discriminating between these groups corresponded to the dlPFC (Liu et al., 2012). This finding appears relatively robust, as it has been replicated in a larger study of drug-naïve patients, albeit with a lower prognostic accuracy (69.6%) (Gong et al., 2011). However, for some studies, it appears that it is patients with treatment-sensitive depression who display the greatest differences compared to controls (Li et al., 2010; Liu et al., 2012; Lui et al., 2011). For instance, Gong and colleagues (2011) found that the accuracy of predicting responsive patients with brain structure was higher than predicting non-responsive patients. It might be that patients with these brain abnormalities are those for which these treatments are effective. Indeed, some functional work appears to support this theory. Patients with the strongest negativity bias in one functional study, and strongest activity in the dlPFC, ventro-lateral PFC and anterior temporal lobe, showed the greatest symptomatic improvement with CBT (Ritchey et al., 2011). It is argued that this increased activation may serve to help combat anhedonic symptoms of the depressive state, due to the role of these areas in reward processing (Fu et al., 2013; Keedwell et al., 2010). Those labelled as resistant may therefore have other abnormalities beyond the brain that require a different treatment process. Alternatively, those with the greatest deviations have greater improvements to make with treatment (Ritchey et al., 2011). Mechanistic theories cannot be investigated until the direction of effect in these data becomes clearer and less reliant on correlational data.

*Insular Cortex*

The anterior region of the insula has a prominent role in monitoring the salience of internal and external stimuli (Seeley et al., 2007), processing reward (Zhang, Chang, Guo, Zhang, & Wang, 2013), and is intricately involved in social emotions such as empathy, compassion and guilt (Lamm & Singer, 2010). Consequently, this region has been referred to as a centre for interoception (Lee & Siegle, 2012). However, the insular cortex as a whole is extremely multifunctional; with subdivisions involved in everything from pain, chemosensation, auditory processing, executive control (Chang, Yarkoni, Khaw, & Sanfey, 2013) and even the experience of conscious awareness (Craig, 2009). This multi-functionality makes the insula a significant hub for interfacing between thoughts, feelings (including physical arousal and mental states) and behaviour (Chang et al., 2013). Considering that these features are of key

interest in depression aetiology, this makes the insula a neural structure of considerable interest when designing imaging psychopathology studies.

A number of review articles implicate the anterior subdivision of the insula in the depressive state (Dunlop & Mayberg, 2014; Lamm & Singer, 2010), due to its extensive connections with the ACC and OFC (Augustine, 1996). One of the largest and most recent meta-analytic studies on neural markers for depression found that the insular cortex showed significant reductions in cortical thickness between patients and controls, with effect sizes comparable to the ACC (Schmaal et al., 2017a). These findings, as with the ACC and OFC, were present in both first-episode and recurrent patients in the ENIGMA sample. Indeed, abnormalities of the insula are argued to be a trait-related feature of depression across both structural and functional neuroimaging (Liu et al., 2014; Takahashi et al., 2010). Two studies have reported that reduced anterior insula volume was present both in patients with past and current depression (Liu et al., 2014; Takahashi et al., 2010), and further, Liu and colleagues (2010) have found that both medication-free, first-episode patients and first-degree relatives showed decreased regional homogeneity in the insula in resting-state fMRI. Regional homogeneity of the anterior insula was also found as one of seven regions that differentiated cases of unipolar from bipolar depression (Liu et al., 2012). The insula demonstrated one of the highest AUC values of these regions, prompting authors to suggest it as a distinguishing feature of MDD. Interestingly, the other key regions mentioned as significant for MDD were those of the PFC and cingulate cortex, mentioned previously in this chapter (Liu et al., 2012). However, it is important to note that discrepancies do exist in the literature. Firstly, one VBM study showed that it was decreased insula volume in the posterior, not anterior, subdivision that associated with melancholic depression, and that these reductions were predictive of slower treatment response (Soriano-Mas et al., 2011). In addition, Foland-Ross and colleagues (2015) reported that thickening of the insula that was predictive of later developing depression in adolescent girls. This discrepancy of thicker cortices in at-risk individuals is similar to that previously discuss with regard to the ACC (van Eijndhoven et al., 2013), and as previously noted, may not necessarily conflict per se, but rather reflect different time-points in the course of disease.

To date it appears the insula plays some contributory role in depression aetiology, but as with other regions, the details require further clarity. Dunlop and Mayberg (2014) argued

that it is precisely this inconsistency in depressive aetiology that suggests the insula may be important for differentiating within a group of depressed patients, rather than between a heterogenous group of patients and controls. As with other brain regions discussed here, there is limited data investigating how structural abnormalities of the insula relate to treatment outcome. Fu and colleagues' (2013) meta-analysis highlighted this: only 9 studies were found investigating structure in medication-free patients, 3 of which were ROI studies focusing on the hippocampus. However, functionally, increased baseline activation of the insula and striatum have associated with poorer response (Fu et al., 2013; Rizvi et al., 2013). Abnormal metabolism of the insula has also been associated with treatment response and remission in a number of PET studies (McGrath et al., 2013; Paillè Re Martinot et al., 2011). Furthermore, this region has been suggested to exhibit the most robust discriminatory power for treatment-specific differentiation (McGrath et al., 2013). A variety of treatments have shown to alter insula activity, including antidepressants (Kennedy et al., 2001), mindfulness (Farb et al., 2007) and deep brain stimulation (Mayberg et al., 2005). Collectively, this suggests that aberrancies of the insula, and their correction, may be key to recovery across therapeutic methods. However, Rizvi and colleagues (2013) emphasised that discrepancies in the outcome of task-based functional studies do still exist, and might well be due to task-based differences in function of that region. Consequently, the utility of fMRI as a biomarker for treatment response in depression continues to be debated (Bullmore, 2012).

Work by Chen and colleagues discussed earlier has demonstrated that there was a strong correspondence between those regions that displayed functional aberrancies and those regions that displayed altered structure (Chen et al., 2007). Symptom improvement following antidepressant medication showed a strong association with GMV in the insula cortex, in addition to those discussed previously (Chen et al., 2007). In line with these findings, a longitudinal study has found that decreased GMV in the left insula was predictive of a slower response to treatment (Soriano-Mas et al., 2011). This reduced GMV further correlated with the number of relapses experienced by depressed patients over two years (Soriano-Mas et al., 2011). However, there are a number of exceptions in the literature. For instance, while depression severity is usually associated with poorer treatment outcome (Curry et al., 2006; Goodyer et al., 1997; Wilkinson et al., 2009), one study found a positive correlation with severity and insula volume (Liu et al., 2014). Given the above, this is

somewhat unexpected. However the correlation between volume and severity is not consistently reported (Takahashi et al., 2010), and one must keep in mind that severity and symptom change are two dissociable concepts. Nevertheless, abnormal structure of the insula cortex marks an important avenue of further investigation.

*Adolescence*

Thus far, unless otherwise stated, all studies discussed here have been conducted on adult samples of patients with depression. However, neurologically, adolescent depression could be an entirely different beast (Kaufman, Martin, King, & Charney, 2001). Neuroimaging studies have shown that throughout development, neurological systems that control certain behaviours become more and more refined. For instance, Casey and colleagues (2000) found that in studies of younger participants, much more diverse patterns of activation were present in cognitive tasks compared with adult samples. In addition, not all brain regions develop at the same time (Gogtay et al., 2004). Indeed, prefrontal regions that regulate emotion and reward processing are the latest to mature (Giedd, 2008; Gogtay et al., 2004; Spear, 2000), which are the very systems implicated in adult depression (Kaufman et al., 2001). Consequently, neurological studies in adolescence require careful consideration when investigating disease processes. One cannot simply extrapolate findings from adult studies onto a developing brain (Kaufman et al., 2001). Furthermore, it cannot be assumed that the mechanistic action of treatment is the same for adolescents, as studies have illustrated differences in sensitivities to antidepressants across development (Bylund & Reed, 2007). As such, predictors of response might also differ in adolescence from adult samples.

There is a scarcity of studies investigating the neural architecture of depression during adolescence, and even fewer investigating neural predictors of response. Of those that have, whether one can say that the depressed adolescent brain resembles that of adult depression is still debatable. For example, in line with adult findings, correlations have been observed between depressive symptoms and increased activity of the rostral ACC and ventral medial PFC (William & Yurgelun-Todd, 2006). Cullen and colleagues (2009) also found similarities with adult studies, reporting decreased functional connectivity between the subgenual ACC and areas of the cortex, including the mPFC and insula in depressed adolescents. Interestingly, most patients in this study had moderate to severe depression despite

receiving antidepressant medication, so these areas may have additional importance as a biomarker of risk for disorder or predicting treatment resistance.

Structurally however, the adolescent literature appears to depart somewhat from adult studies, with a number of inconsistencies. For instance, one longitudinal study suggested that subcortical abnormalities may play a more prominent role in adolescent depression than cortical abnormalities (Whittle et al., 2014). The authors found that altered growth patterns of the hippocampus, amygdala and putamen associated with depression onset in adolescence between 12 and 16 years, but found no association in cortical regions (Whittle et al., 2014). However, other studies have supported the significance of cortical structural abnormalities in depressed adolescents (Ducharme et al., 2014; Foland-Ross et al., 2015; Hagan et al., 2015; Reynolds et al., 2014). Cross-sectional work has suggested that age-related changes in GMV in the ACC differ for adolescents with depression compared to healthy controls (Hagan et al., 2015). Together, these studies suggest that pathology may interfere with the way in which the adolescent brain matures, and this development may impact the differences we observe between adult and adolescent depression (Koenig et al., 2018). Indeed, in contrast to the adult literature, a number of studies have found that depressed adolescents exhibited thicker cortices in prefrontal, cingulate and insula regions compared with controls (Koenig et al., 2018; Reynolds et al., 2014). Longitudinal studies have also found associations between cortical thickness and depression (Ducharme et al., 2014; Foland-Ross et al., 2015). Ducharme and colleagues (2014) found a slower rate of thinning in the ventral medial PFC related to depressive symptomatology in their healthy adolescent sample. This study also reported that cortical thickness of the ACC and OFC positively correlated with depressive symptoms, although not all studies support the direction of effect for these regions (Foland-Ross et al., 2015). Increased thickness of the insular cortex has been found to associate with depression onset (Foland-Ross et al., 2015) and the interaction between depressive symptomatology and cortical thickness of this region has additionally been shown to predict autonomic function (Koenig et al., 2018). Koenig and colleagues (2018) found that for depressed adolescents (and controls) with lower levels of depression, increased insular thickness associated with lower regional resting-state vagal activity. Conversely, for those with more severe depression, increased insular thickness associated with higher resting-state vagal activity (Koenig et al., 2018). This suggests that depressive pathology can influence the relationships between neurological parameters. The

authors postulate that the abnormal cortical thickness observed in adolescent depression might be a compensation to maintain appropriate levels of autonomic function (Koenig et al., 2018).

The literature discussed thus far suggests that, while the direction of effect may contrast with adult depression, cortical thickness of prefrontal and cingulate regions appear important in the pathology of adolescent depression. However, robust meta-analytic data in this field is lacking, and results from the ENIGMA consortium cause further debate (Schmaal et al., 2017a). Contrary to adult studies included in their analysis, and previous adolescent research, their analysis suggested that cortical thickness did not illustrate case-control differences in adolescence. Adolescents in this study showed no differences in cortical thickness when compared to controls, but instead had lower total surface area, and regional reductions in surface areas in the OFC and superior frontal gyrus (an area that forms part of the dlPFC) (Schmaal et al., 2017a). However, their adolescent sample were not of the typical age-range associated with the term "adolescence": 70% of their adolescent sample was over 18 years of age (Schmaal et al., 2017a). Very few studies have investigated surface area as their parameter of interest in depression, with only one in the typical adolescent age range (Schmaal et al., 2017b). This study investigated this relationship between depressive symptom trajectories and cortical structural parameters longitudinally, in adolescents aged between 12 and 19 years (Schmaal et al., 2017b). Furthermore, a unique characteristic of this study was their employment of GMM to define their comparison groups. The authors found three types of patients emerged in their cohort: a group reporting early onset depressive symptoms which decline over time, a group with low and stable depressive symptoms over time, and a group reporting depressive symptoms, although later in adolescence. Interestingly, and in line with the ENIGMA consortium, the authors found no evidence of cortical thickness differences between different symptom trajectories, but there were significant differences in surface area (Schmaal et al., 2017b). Females who reported symptoms of depression early in adolescence displayed significantly reduced ACC and OFC surface area across development, compared to the other two groups. Furthermore, males in this early-onset group showed a lower rate of expansion in cortical surface area of the OFC compared to the other two groups. These findings emphasise that early onset of depressive conditions might show different relationships with brain structure than if the condition

occurred later in development. More research is necessary to understand how neurological parameters differ in depression and across development.

A major strength of Schmaal and colleagues' work (2017b) is their incorporation of empirical approaches to defining their groups for comparison. As noted in chapter 1, defining subgroups of individuals based on percentage reduction at a single time point, the common method of most studies discussed here is arbitrary and allows for heterogeneity to exist within groupings. This could easily cloud significant differences on an individual study level and may be a contributing factor to the lack of consensus of the field. It is imperative that response groups are categorised as homogenously as possible, based on the experiences of the patient, and GMM nicely allows for this type of grouping. However, this is the only study to date to incorporate GMM in neuroimaging research in depression, and there is a lack of structural studies extending this into clinical samples receiving treatment. On the other hand, one study has investigated associations between brain function and treatment response in adolescents, using a similar approach of growth curve modelling (a full description of these approaches was outlined in chapter 1). Forbes and colleagues (2010) found that the activity of reward-related brain function in the striatum and mPFC prior to treatment predicted both final depression levels, and the change in anxiety symptoms over time in adolescents. Specifically, greater activity in the mPFC is associated with a poorer rate of symptom reduction. Unfortunately, this study is limited by its small sample size: with only 13 patients, statistical modelling to investigate the presence of sub-trajectories was not possible. Larger studies are needed if neuroimaging research is to adopt more meaningful approaches to defining groups based on symptom change.

In conclusion, there appears a network of cortical regions that are implicated in depression aetiology and treatment response, most consistently including the ACC, dlPFC, OFC and insular cortex. These regions also appear significant for adolescent depression. However, the field of adolescent depression is currently in equipoise regarding the direction of effect in its structural investigations, and there is a scarcity of studies investigating the prognostic potential of pre-treatment brain structure in this population.

*Objectives and Hypotheses*

The objective of this next piece of work was to conduct a secondary analysis of a subsample of adolescents enrolled into the IMPACT trial. The Magnetic-Resonance-IMPACT (MR-IMPACT) subsample contained 109 patients who had additional structural MRI scan data collected at treatment randomisation. The specific aim was to investigate the relationship between empirically-derived trajectory classes, as defined in Chapter 1, and baseline cortical thickness and surface area, as measured using the surface-based analysis software (FreeSurfer). Prior literature is currently inconclusive to predict a direction of effect with confidence, however we hypothesise that the two trajectory classes would show significantly different cortical thickness and surface area in specific regions of the ACC, OFC, dlPFC and insular cortices.

**Methods**

*Study Design and Size*

This study was a re-analysis of the MR-IMPACT trial (Hagan et al., 2013). MR-IMPACT consisted of a subgroup of patients who, in addition to the data of the IMPACT trial, also underwent structural MRI scans. Full details of the protocol are outlined by Hagan and colleagues (2013) and eligibility criteria for IMPACT is outlined in Chapter 1.

*Setting*

The MR-IMPACT trial (Hagan et al., 2013) recruited 128 patients from the IMPACT trial. One hundred and ten of these patients were recruited from East Anglia, and a remaining 18 were recruited from North London.

*Bias*

Bias associated with the main IMPACT sample is outlined in Chapter 1. The sample used for this study was primarily collected from only one of the three geographical regions of recruitment for IMPACT. Therefore, we cannot be certain that the sample is necessarily representative of the full IMPACT sample. However, comparisons between the sub-sample and main sample were conducted and discrepancies reported below.

*Participants*

Of the 128 patients recruited to MR-IMPACT, 11 patients were excluded for brace or retaining wire MRI artefacts, 2 for brain abnormalities and 6 because their scans occurred after psychological treatment had commenced. Thus, 109 patient scans were available for MRI analysis. Full details of the demographic characteristics of these 109 patients can be found in prior work by Hagan and colleagues (2015).

The GMM longitudinal classes described in Chapter 1 were used as a binomial variable to examine imaging characteristics between continued and halted-improvers. Of the 109 patients, 19 (17.4%) were allocated to class 1 and 90 patients were allocated to class 2 (82.5%). While class 1 contained a slightly higher proportion of patients and class 2 a slightly lower proportion of patients than observed in the full IMPACT sample (15.9% and 84.1% respectively), this difference was non-significant ($X^2(1)= 0.244$, $p=.621$) .

Analyses of associations were conducted in a separate step. Table 1 presents the demographic and clinical data of each trajectory class in the MR-IMPACT cohort. Independent sample t-tests were performed to explore differences between the trajectory classes on all continuous demographic and clinical variables. Chi-square tests were performed on all categorical variables. Mann-Whitney-Wilcoxon tests were performed where data were non-normal.

No significant differences were present between classes on any demographic characteristics investigated in the MR-IMPACT sample. This differs slightly from the full IMPACT sample, which found a significant difference between the proportion of males and females allocated to their respective classes (See Chapter 2). This difference between samples is likely due to the small sample size of MR-IMPACT.

In terms of clinical characteristics, firstly, class 1 showed on average higher depression severity (MFQ). This however, was expected, as MFQ was the variable used to define the two classes (See Chapter 1). Class 1 also showed higher LOI scores, indicating more obsessional traits than class 2 at baseline. This was also true for the full IMPACT sample, however the MR-IMPACT sample did not demonstrate any other significant differences on clinical measures measured in the full IMPACT sample. Measures of state and trait anxiety were additionally investigated in the MR-IMPACT subsample, using the State-Trait Anxiety Inventory (STAI). Here, class 1 displayed higher STAI trait scores, indicating that patients in class 1 display higher trait anxiety than patients in class 2.

*Table 1: Demographic and clinical characteristics of trajectory classes in MR-IMPACT*

| | **Class 1: Halted-improvers (n=19)** | | **Class 2: Continued-improvers (n=90)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** | ***$X^2$/OR*** | ***p*** |
| *Female[%]* | 15 | 73% | 66 | 79% | 0.74 | .78 |
| *Ethnicity(white)[%]* | 18 | 95% | 81 | 90% | 1.99 | .99 |
| *Suicidal attempts* | 6 | 32% | 28 | 31% | 0.002 | .97 |
| *Suicidal thoughts[%]* | 18 | 95% | 82 | 91% | 1.75 | .99 |
| *NSSI[%]* | 15 | 79% | 62 | 69% | 1.69 | .58 |
| *Comorbidity[+]* | - | - | - | - | 1.09 | .30 |
| *1* | 7 | 37% | 33 | 37% | - | - |
| *2* | 4 | 21% | 12 | 13% | - | - |
| *3* | 1 | 5% | 0 | 0% | - | - |
| *Treatment arm:* | - | - | - | - | 1.81 | .40 |
| *BPI* | 8 | 42% | 24 | 27% | - | - |
| *CBT* | 6 | 32% | 37 | 41% | - | - |
| *STPP* | 5 | 26% | 29 | 32% | - | - |
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***t*** | ***p*** |
| *Age* | 15.3 | 1.3 | 15.6 | 1.3 | 0.96 | .35 |
| *IQ (available for 5 in class 1 and 17 in class 2)* | 99.6 | 10.7 | 95.7 | 11.8 | -0.70 | .51 |
| *MFQ* | 51.8 | 10.2 | 44.7 | 10.3 | -2.77 | .01 |
| *RCMAS* | 42.7 | 6.1 | 41.2 | 6.1 | -0.95 | .35 |
| *LOI* | 13.4 | 5.2 | 9.5 | 4.8 | -3.00 | .006 |
| *HONOSCA (available for 16 in class 1 and 82 in class 2)* | 18.3 | 6.7 | 19.1 | 5.5 | 0.45 | .66 |
| *STAI state score* | 49.5 | 10.0 | 44.8 | 11.0 | -1.81 | .08 |
| *STAI trait score* | 64.3 | 8.1 | 59.5 | 7.9 | -2.35 | .03 |
| | ***Median*** | ***IQR*** | ***Median*** | ***IQR*** | ***W*** | ***p*** |
| *Handedness* | 60 | 140 | 80 | 30 | 1083.5 | .07 |
| *IMD* | 17.5 | 18.4 | 13.8 | 13.0 | 732 | .328 |

*[%]Fishers exact test conducted on variables with insufficient cell size for chi-square test.*

*[+]Variable recorded as binary to meet assumptions of chi-square test.*

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, IQ; Intellegence Quotent, STAI; State-Trait Anxiety Inventory.*

The two classes did not significantly differ in their treatment arm allocation within the MR-IMPACT subsample. However, 34% patients were taking a selective serotonin re-uptake inhibitor (SSRI) antidepressant medication at the time of scanning. The numbers per trajectory class is detailed in Table 2. Fluoxetine was the most commonly prescribed SSRI in these patients (32 patients, 86% of those taking SSRIs). Of those taking anti-depressant medication at the time of scanning, the majority followed the continued-improvers trajectory (36, 97%). This difference was statistically significant ($X^2(1)=8.44$, $p=.004$). Of note, there was no significant difference in SSRI prescription rate between latent classes across the whole IMPACT sample (halted-improvers: 14%, continued-improvers: 22%, p = 0.09). Consequently, we are confident that this difference is due to a sampling bias, rather than any other underlying association, as it is known that SSRI prescription varied between study sites in IMPACT (Cousins et al., 2016). However, additional analyses were run to investigate the main effect of SSRI prescription (described below). If any main effects of SSRIs emerged, analyses would be repeated excluding those patients, as a sensitivity check.

*Table 2: Medication use at time of scan of trajectory classes in MR-IMPACT*

| | **Class 1: Halted-improvers (n=19)** | | **Class 2: Continued-improvers (n=90)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | N | % | N | % | *OR* | *p* |
| *SSRI at scan*[%] | 1 | 5% | 36 | 40% | 0.08 | .002 |
| ***SSRI medication use:*** | | | | | | |
| *Fluoxetine* | 1 | 5% | 31 | 34% | - | - |
| *Citalopram* | - | - | 4 | 4% | - | - |
| *Sertraline* | - | - | 1 | 1% | - | - |
| ***Other neuroactive medication use:*** | | | | | | |
| *Propanolol* | 1 | 5% | 1 | 1% | - | - |
| *Risperidone* | 0 | 0% | 3 | 3% | - | - |
| *Gabapentin* | 0 | 0% | 1 | 1% | - | - |
| *Lamotrigine* | 0 | 0% | 1 | 1% | - | - |
| *Zolpidem* | 0 | 0% | 1 | 1% | - | - |
| *Circadin* | 0 | 0% | 1 | 1% | - | - |
| *Diazepam* | 0 | 0% | 1 | 1% | - | - |

*SSRI; Selective Serotonin Reuptake Inhibitor*

*[%]Fishers exact test conducted on variables with insufficient cell size for chi-square test.*

*Structural MRI (sMRI) data acquisition*

All patients were scanned at the Wolfson Brain Imaging Centre, University of Cambridge, UK. The study was conducted on a 3.0 Tesla Magnetom Trio Tim scanner (Siemens, Surrey, England) fitted with a quadrature birdcage head coil. High-resolution T1-weighted images were acquired in the sagittal plane using a 3-dimensional magnetically-prepared rapid acquisition gradient echo sequence (3D-MPRAGE). Acquisition parameters were: 176 slices of 1.00mm thickness, echo time =2.98ms, repetition time =2.30s, inversion time=900ms, flip angle =9°, field of view =240 x 256mm$^2$, isometric voxel-size =1.0mm$^3$ with an interleaved slice acquisition). Brain abnormalities were screened with a high-resolution dual echo-proton density and T2-weighted sequence, by a consultant radiologist specialising in neuroanatomy (Hagan et al., 2013).

*Pre-processing of sMRI data*

The pre-processing and quality checks of the sMRI data were conducted by a researcher blind to class allocations. Automated parcellation of the cortex and segmentation of sub-volumes were performed using the FreeSurfer software package (Fischl, 2012). This image reconstruction of the cortical surface involves several stages. Firstly, the intensity variations caused from inhomogeneities in the magnetic field during scanning are corrected, images are affine registered to the Talairach atlas (Talairach & Tournoux, 1988) and voxels that contain the skull and dura are removed (Dale, Fischl, & Sereno, 1999). In the segmentation step, white matter voxels are classified based on intensity differences, and hemispheres are separated for preliminary segmentation. The grey-white matter boundary is constructed by the application of a triangular surface tessellation to each hemisphere, and smoothed with a Gaussian kernel. The pial surface is then constructed from expanding the white matter surface outwards to the edge of the brain; defined by intensity gradients between grey matter and CSF (Dale et al., 1999; Fischl, 2012). Finally, this surface is then inflated to provide a parameterizable surface for inter-subject registration. Inflation of the cortical surface removes the interference of cortical folding so that the cortical area within sulci can be appropriately represented and accurately matched to a reference template (Fischl, Sereno, & Dale, 1999). FSaverage was used as the reference brain template ("FsAverage - Free Surfer Wiki," n.d.), and images are non-linearly transformed to align with this template based on the positioning of gyri and sulci along two dimensions (latitude and longitude) (Fischl et al., 1999). Sulci and gyri that have low inter-individual variability, such as the

central sulcus and slyvian fissure, are given stronger weighting in the registration algorithm than those that are known to have greater inter-individual variability. Each image is then parcellated into 34 cortical regions per hemisphere, based in the Desikan-Killiany atlas (Desikan et al., 2006), and mapped back onto each individual subject's surface. Cortical thickness measurements were taken as the smallest distance between the pial surface and white matter surface. Thickness values were generated at each vertex and measured twice, once from grey-white matter boundary to pial, and then reversed. No subjects were excluded following quality-checks. Minimal editing of white matter and non-brain tissue was conducted where necessary.

*Statistical analysis of sMRI data*: *vertex-wise analysis*

Comparisons between patients in class 1 and class 2 were carried out, controlling for differences in gender and age. As total brain surface area is related to morphometric measures like thickness, white matter surface area was also added as a covariate of no interest.

Imaging analyses followed the pipeline created by Dr Kirstie Whitaker (Whitaker, 2015). Briefly, images were first resampled to fsaverage, and then concatenated into one file using the mris_preproc command in FreeSurfer. A general linear model (GLM) analysis was conducted on the surface using mri_glmfit command. This command runs the model at each individual vertex. Correction for multiple comparisons was conducted using cluster correction. This calculates the significance of clusters based on the number of contiguous significant vertices, and the chosen amount of smoothing. We used a 15mm Full Width at Half Maximum (FWHM). Cluster forming threshold was set at 0.01 and significance threshold for clusters was set at 0.05. The –2spaces flag was also used in the mri_glmfit-sim command to correct for doing tests on both the left and right hemispheres.

*Sensitivity analysis for SSRIs: vertex-wise analysis*

Due to insufficient numbers within each cell of a 3-factor model, it was not possible to enter SSRI prescription as another categorical covariate in the whole-brain model, and an appropriate continuous variable was not available. Therefore, the difference in thickness and surface area, controlling for gender, age and white matter total surface area, between those who were and were not taking SSRIs at baseline was tested separately per hemisphere. No

whole-brain differences were found in cortical thickness or surface area, between those who were taking SSRIs at the time of scanning and those who were not. Therefore, as there was no main effect of SSRI prescription on whole-brain results, patients taking SSRIs (n=37, 34%) at the time of scanning remained in the analysis (total n=109).

*Statistical analysis of sMRI data*: *region of Interest (ROI) analyses*

Regions of interest (ROIs) were defined for each hemisphere, by FreeSurfer's automated cortical parcellation procedure (Fischl et al., 2004), using the Desikan-Killiany atlas (Desikan et al., 2006). The insula cortex and rostral ACC, are parcellated individually per hemisphere. The OFC is subdivided into medial and lateral components. There is not sufficient evidence to date to exclude either component from the analyses, and some studies combine these divisions to create one single value (Dotson et al., 2016). However, as averaging across the two regions might wash out any significant, more specific effects, we decided to include both as separate regions of interest. Similarly, the dlPFC has been defined as the combination of the rostral and caudal middle frontal gyrus (Garber & Weersing, 2010) and superior frontal gyrus (Yamagishi et al., 2016). As there was no compelling evidence in favour of a single subdivision, each of these regions were also treated as separate ROIs. Consequently, 7 ROIs were tested in total, per hemisphere. These are shown in Figure 1.

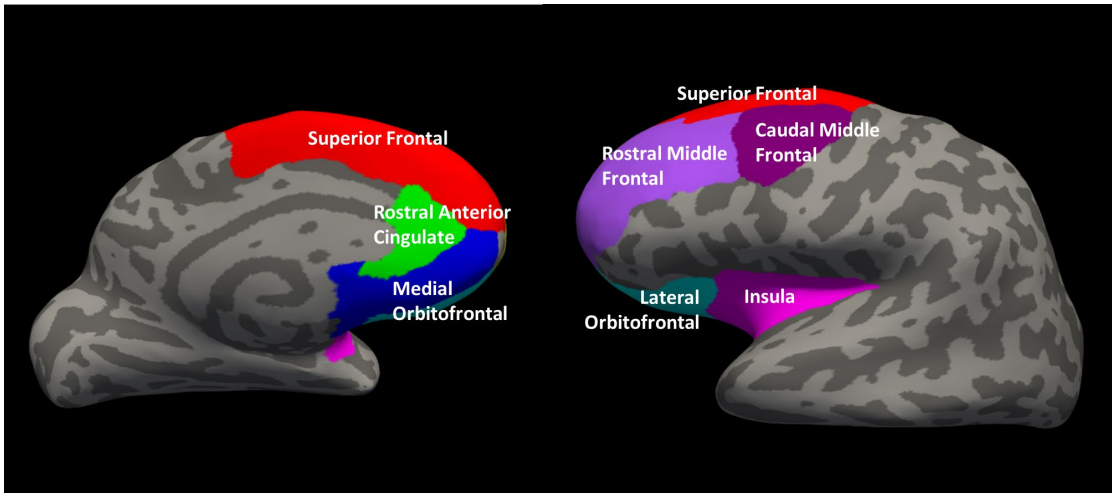



*Figure 1: Cortical regions of interest (ROIs) mapped onto inflated brain.*

Logistic regressions were conducted to investigate whether measures at each of the ROIs could predict class membership, controlling for age, gender and white matter total surface area. Regressions were conducted using the *glm()* function in the core *stats* package in R,

version 3.3.3 (“R: The R Project for Statistical Computing,” n.d.). Both cortical thickness and surface area were investigated separately as variables of interest. The assumptions of multicollinearity, independent errors and linearity of logit were tested for each ROI. In cases of violations, permutation testing at the class level was also conducted. Permutation testing reshuffles the true class allocation across patients and produces a null distribution of the dataset through 10,000 iterations (Ge, Yeo, & Winkler, 2018). Assumptions of logistic regressions therefore no longer apply. If the observed p-value and the p-value obtained through permutation testing agree, the results can be trusted in the presence of the violation (Ge et al., 2018).

A false discovery rate (FDR) threshold of <.05 was used to account for multiple comparisons in ROI analyses (Benjamini & Hochberg, 1995). Unlike family-wise error rate corrections (such as the Bonferroni adjustment) that control for the possibility of at least 1 false positive, FDR procedures control for the expected proportion of false positives. FDR corrections are therefore less stringent. However, Bonferroni correction in analyses that require large numbers of multiple tests would result in too strict a threshold and risk rejection of true positives. Consequently, in such cases FDR correction is preferred (Genovese, Lazar, & Nichols, 2002). FDR correction was conducted using the *p.adjust()* function of the *stats* package in R, version 3.3.3.

*Sensitivity analysis for SSRIs: ROI analyses*

Similar to the vertex-wise analyses, the effect of SSRI prescription on thickness and surface area, controlling for gender, age and white matter total surface area was tested separately for each ROI, per hemisphere, correcting for multiple comparisons with FDR. After correction, there was no significant effect of SSRIs on cortical thickness or surface area of any ROI (details found in Appendix 3A). Consequently, as there was no main effect of SSRI prescription on any ROI for either brain structural variables of interest, patients taking SSRIs at the time of scanning remained in the analyses (n=109).

**Results**

*Group differences across the whole-brain*

Vertex-wise analyses revealed no significant differences between trajectory classes of symptom change in cortical thickness or surface area, in any region of the cortex in both hemispheres. This was apparent before and after cluster correction for multiple comparisons.

*Region of Interest*

Means and SDs for cortical thickness and surface area of each region of interest (ROI), per class are shown in Table 3.

*Table 3: Mean cortical thickness and surface area of each ROI per class.*

| | **Cortical Thickness (mm)** | | | | **Surface Area ($mm^2$)** | | | |
|---|---|---|---|---|---|---|---|---|
| | *Class 1* | | *Class 2* | | *Class 1* | | *Class 2* | |
| *Region* | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* |
| *Medial OFC* | | | | | | | | |
| *Left* | 2.60 | 0.19 | 2.58 | 0.14 | 1943.32 | 265.57 | 1916.34 | 279.49 |
| *Right* | 2.66 | 0.16 | 2.56 | 0.16 | 1839.53 | 199.43 | 1838.12 | 205.90 |
| *Lateral OFC* | | | | | | | | |
| *Left* | 2.87 | 0.17 | 2.84 | 0.14 | 2817.90 | 212.35 | 2775.91 | 252.41 |
| *Right* | 2.83 | 0.15 | 2.80 | 0.15 | 2573.11 | 259.29 | 2623.5 | 262.91 |
| *SFG* | | | | | | | | |
| *Left* | 2.91 | 0.13 | 2.91 | 0.14 | 7186.63 | 766.11 | 7485.68 | 889.29 |
| *Right* | 2.81 | 0.13 | 2.80 | 0.14 | 6862.26 | 831.78 | 7272.88 | 760.41 |
| *rMFG* | | | | | | | | |
| *Left* | 2.49 | 0.12 | 2.44 | 0.14 | 5714.26 | 670.88 | 6094.40 | 703.46 |
| *Right* | 2.36 | 0.12 | 2.36 | 0.13 | 6153.32 | 566.33 | 6322.06 | 919.78 |
| *cMFG* | | | | | | | | |
| *Left* | 2.69 | 0.14 | 2.67 | 0.15 | 2334.11 | 430.23 | 2383.96 | 407.15 |
| *Right* | 2.58 | 0.16 | 2.62 | 0.14 | 2228.63 | 503.50 | 2192.01 | 375.84 |
| *rACC* | | | | | | | | |
| *Left* | 3.11 | 0.20 | 3.08 | 0.24 | 873.94 | 196.42 | 880.19 | 177.47 |
| *Right* | 3.08 | 0.16 | 3.06 | 0.21 | 631.00 | 115.04 | 677.72 | 111.15 |
| *Insula* | | | | | | | | |
| *Left* | 3.27 | 0.09 | 3.27 | 0.13 | 2154.31 | 207.88 | 2215.17 | 251.12 |
| *Right* | 3.29 | 0.10 | 3.27 | 0.12 | 2175.68 | 236.15 | 2237.73 | 262.70 |

*OFC: Orbitofrontal Cortex; SFG: Superior Frontal Gyrus; rMFG: Rostral Middle Frontal Gyrus; cMFG: Caudal Middle Frontal Gyrus; rACC: Rostral Anterior Cingulate Cortex*

No region of interest raised concerns of multicollinearity or independent errors for either variable of brain structure, supporting the use of logistic regressions with these data. All regression models also met the linearity of logit assumption for cortical thickness analyses. While two regions (right rMFG and right cMFG) failed this assumption for surface area analyses, permutation testing results did not differ from those reported below (Table 1; Appendix 3B).

Results from each logistic regression (FDR corrected) are shown in Table 4.
After controlling for white matter total surface area, age and gender, the odds of being in class 1 decreased with increasing cortical thickness in the right mOFC, however this did not survive FDR correction (Table 4). Similarly, the odds of being in class 1 decreased with increasing surface area in the right SFG but this also did not survive FDR correction. After applying the appropriate multiple comparison corrections, no ROI significantly predicted trajectory class membership, in either variable of brain structure.

Standardised residuals of the final model were inspected to assess for whether the model contained any outliers of concern. Less than 1% of this sample had residuals larger than ±2.58. For the majority of regions, the percentage of the total sample of cases showing residuals greater than ±1.96 was less than 5%, and never exceeded 8% for either structural variable. Therefore the models can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases are exerting undue influence over the models. It has been recommended to investigate cases where leverage values are greater than twice (Hoaglin & Welsch, 1978) or three time (Stevens, 2009) the average. While there were a number of cases per regions and structure variables that suggested a potential problem with 2 or 3 times greater than average leverage, the Cook's distance never exceeded 1 (Cook & Weisberg, 1982), indicating that the fit would not significantly change upon removal of these cases.

*Table 4: ROI predictors of trajectory class membership, adjusted for age, gender and total white matter surface area.*

| | **Cortical Thickness (mm)** | | | | **Surface area (mm$^2$)** | | | |
|---|---|---|---|---|---|---|---|---|
| *Region* | *OR* | *95%CI* | *p* | $p_{(FDR)}$ | *OR* | *95%CI* | *p* | $p_{(FDR)}$ |
| *Medial OFC* | | | | | | | | |
| *Left* | 2.90 | 0.07-108.93 | .56 | .96 | 1.00 | 0.99-1.00 | .24 | .56 |
| *Right* | 4.49 | 1.57-1852.74 | .03 | .46 | 1.00 | 0.99-1.00 | .40 | .61 |
| *Lateral OFC* | | | | | | | | |
| *Left* | 2.32 | 0.06-87.42 | .64 | .96 | 1.00 | 0.99-1.00 | .15 | .43 |
| *Right* | 3.24 | 0.10-120.58 | .51 | .96 | 1.00 | 0.99-1.00 | .74 | .83 |
| *SFG* | | | | | | | | |
| *Left* | 0.31 | <.01-16.59 | .57 | .96 | 1.00 | 0.99-1.00 | .13 | .43 |
| *Right* | 1.30 | 0.02-75.81 | .90 | .96 | 1.00 | 0.99-1.00 | .02 | .18 |
| *rMFG* | | | | | | | | |
| *Left* | 8.86 | 0.18-518.16 | .28 | .96 | 1.00 | 0.99-1.00 | .03 | .18 |
| *Right* | 0.55 | 0.01-34.59 | .78 | .96 | 1.00 | 0.99-1.00 | .77 | .83 |
| *cMFG* | | | | | | | | |
| *Left* | 1.84 | 0.05-79.02 | .75 | .96 | 1.00 | 0.99-1.00 | .98 | .98 |
| *Right* | 0.16 | <.01-5.40 | .31 | .96 | 1.00 | 0.99-1.00 | .36 | .61 |
| *rACC* | | | | | | | | |
| *Left* | 1.53 | 0.15-16.91 | .72 | .96 | 1.00 | 1.00-1.01 | .58 | .74 |
| *Right* | 1.23 | 0.10-15.33 | .87 | .96 | 1.00 | 0.99-1.00 | .15 | .43 |
| *Insula* | | | | | | | | |
| *Left* | 0.90 | 0.01-56.68 | .96 | .96 | 1.00 | 0.99-1.00 | .44 | .62 |
| *Right* | 3.95 | 0.04-435.80 | .56 | .96 | 1.00 | 0.99-1.00 | .37 | .61 |

*OFC: Orbitofrontal Cortex; SFG: Superior Frontal Gyrus; rMFG: Rostral Middle Frontal Gyrus; cMFG: Caudal Middle Frontal Gyrus; rACC: Rostral Anterior Cingulate Cortex*

## Discussion

*Key Results and Interpretation*

This piece of work was the first to our knowledge to investigate the relationships between structural brain variables and empirically-derived trajectory classes of symptom change over time, in a cohort of clinically depressed adolescents receiving treatment. Vertex-wise analyses of the whole-brain failed to show any significant differences in cortical thickness or surface area between trajectory classes, when controlling for gender, age and total white matter surface area. Furthermore, neither cortical thickness nor surface area of any region of interest demonstrated predictive value for identification of class membership when controlling for the aforementioned variables. This is contrary what was expected in an adolescent sample, given our understanding of the neurobiology of adolescent depression (Ducharme et al., 2014; Foland-Ross et al., 2015; Hagan et al., 2015; Reynolds et al., 2014). Our results also oppose a number of previous studies of treatment response in adults which suggest that structural abnormalities, particularly in these areas, can identify between favourable and unfavourable outcomes (Costafreda et al., 2009; Dunlop & Mayberg, 2014; Fu et al., 2013; Gong et al., 2011; Liu et al., 2012; Pizzagalli, 2011; Soriano-Mas et al., 2011). As our groups were derived from longitudinal data modelling, these findings tentatively suggest that when groups are defined in a data-driven but more clinically meaningful way, the structural differences between outcome-related subgroups may become reduced. It is possible that the measures of brain structure used in this study may not serve a strong role in differentiating between adolescent patients' symptom trajectories.

This current analysis demonstrated a good amount of internal validity with respect to the full IMPACT sample. Firstly, no significant differences were reported between the proportions of patients in trajectory classes of the full IMPACT sample and the MR-IMPACT subsample. Furthermore, the depiction of classes in the main sample suggests that halted-improvers were overall clinically more severe at baseline. While the MR-IMPACT sample showed fewer differences between classes in terms of demographic and clinical characteristics, the sample did demonstrate agreement with this premise. Obsessional traits and trait anxiety were found to be higher in halted-improvers than in continued-improvers in the MR-IMPACT cohort. It is likely that the fewer significant differences are a result of the significantly smaller sample size in this cohort. Overall, we are confident that the findings here can generalise to the main IMPACT cohort.

The MR-IMPACT sample also found no differences in treatment arm allocation between classes. However, a significant difference did emerge between classes on their SSRI prescription rate at baseline. The majority of patients (97%) prescribed antidepressants at baseline subsequently followed a continued-improver trajectory. Intuitively, this might suggest that actually SSRI prescription alongside psychological treatment is a driving factor in determining a favourable symptom trajectory, and indeed, a number of studies have supported the benefits of combination treatment over monotherapy (Cuijpers, van Straten, Warmerdam, & Andersson, 2009b; Khan, Faucett, Lichtenberg, Kirsch, & Brown, 2012; March et al., 2004). However, such a conclusion cannot be drawn from these current results. Firstly, we did not find any main effect of SSRIs on any structural parameter investigated, suggesting that SSRIs cannot account for the results we report here. Secondly, it is important to note that a unique characteristic of this sample was that they were almost solely recruited from the East Anglia clinics of the trial (N=92). Previous authors have published works on the discrepancies that emerged in the trial between SSRI prescription in East Anglia and other regions (Cousins et al., 2016), advising that associations between this East Anglia sample and SSRI prescription rates should be taken with caution. Indeed, we did not observe this class difference in SSRI prescription in the full, substantially larger, IMPACT sample (reported in Chapter 2). We therefore believe these effects are likely a result of chance or the regional selection bias of MR-IMPACT and not illustrative of a real effect of SSRIs on trajectory class membership.

*Alternative explanations*

Our study did not hypothesise a direction of effect for our particular research questions. This is because prior literature is inconclusive. Research has both supported findings of cortical thickness differences between depressed adolescents and healthy controls (Foland-Ross et al., 2015; Koenig et al., 2018; Reynolds et al., 2014), and also rejected these claims, stating that surface area shows more promising prognostic value (Schmaal et al., 2017a; Schmaal et al., 2017b). Moreover, studies supporting cortical thickness differences in our regions of interest further disagree over whether increased or decreased thickness is observed in adolescent depression (Ducharme et al., 2014; Foland-Ross et al., 2015; Koenig et al., 2018; Reynolds et al., 2014). These discrepancies in prior work, coupled with our null results highlight the complexity of investigating brain structure in a developing sample. This sample is characterised by large changes in cortical development, which are not uniform across the

brain, nor between individuals (Tamnes et al., 2017). Investigating trajectories of symptom change on top of trajectories of neurological change, therefore means that it is possible that individual differences in neurological development may have masked significant differences between our groups. This point is discussed in more detail in Chapter 5.

However, this is unlikely to fully account for our null results, given that our classes did not significantly differ in age, and further, we appropriately controlled for age in both vertex-wise and ROI analyses. Our sample size was not sufficient to allow for additional investigations of interaction effects between age, brain structure and class membership. However, it would be an important avenue to consider in future adequately powered research, given our extensive knowledge of neurological development at this age (Giedd, 2008; Gogtay et al., 2004; Kaufman et al., 2001; Spear, 2000).

Another important distinction between this work and others of the field is that we only investigated pre-treatment differences in brain structure. While the investigation of treatment-related effects on brain structure is a different question to the one investigated here, it is possible that structural alterations between halted and continued-improvers may only become apparent once the intervention is introduced. Subsequently, early identification of such changes could act as a predictor of longer-term response. There have been a number of functional studies that have illustrated changes in the activity of prefrontal and cingulate regions following treatment with both pharmacological and psychological therapies (Keedwell et al., 2010; Kennedy et al., 2001, 2007; Ruhé, Booij, Veltman, Michel, & Schene, 2008). For example, Keedwell and colleagues (2010) found that greater increases in activation of the subgenual cingulate gyrus, among other regions, after 12 weeks of antidepressant treatment was correlated with increasing response rates. Metabolic studies have also supported these findings (Kennedy et al., 2001, 2007). Antidepressant response was associated with increased glucose metabolism of the dlPFC, ACC and other regions of the prefrontal cortex, and decreased metabolism of the insular cortex and subcortical regions after 6 weeks of treatment (Kennedy et al., 2001). Furthermore, differential functional alterations have been observed across treatments. Kennedy and colleagues (2007) found that a reduction in glucose metabolism in the OFC and mPFC associated with general response after 16 weeks of any treatment, whereas the direction of change in the posterior cingulate differentiated between response to

medication and CBT. Non-response was also differentially associated with changes in insula and thalamic metabolism. Finally, Ruhe and colleagues (2008) showed that while final responders and non-responders did not differ in their baseline amygdala activation levels, only responders showed an activation decrease in this region. However, for adolescent depression, there is a lack of literature investigating whether treatment affects cortical thickness and surface area. Future work would benefit from identifying whether structural change accompanies functional alterations, and whether these changes might act as a better predictor of longer-term response than pre-treatment measurement alone.

It is possible that the lack of positive results may be in part due to the inclusion of patients with comorbidities in both groups. In agreement with the full IMPACT sample, approximately 50% of the MR-IMPACT sample presented with at least 1 comorbid mental illness, meeting clinical threshold for diagnosis. Comorbid anxiety presented as the most common diagnosis. Anxiety disorders have been associated with altered brain structure and function from what is typical in healthy controls (Etkin & Wager, 2007; Protopopescu et al., 2006; Uchida et al., 2008). A meta-analysis of functional work in anxiety disorders showed evidence of greater activity in the amygdala and insular cortex across anxious conditions (Etkin & Wager, 2007). Furthermore, structural studies have found GMV reductions in prefrontal (Protopopescu et al., 2006), ACC and insular regions in panic disorder (Uchida et al., 2008). In terms of patients with depression, those presenting with anxious symptoms have also demonstrated additional structural anomalies than depressed patients without anxious symptoms, particularly in the caudate nucleus (Zhao et al., 2017). Abnormalities have also associated with treatment response in anxiety disorders (Shin, Davis, VanElzakker, Dahlgren, & Dubois, 2013; Whitfield-Gabrieli et al., 2015). Greater functional connectivity between the amygdala and prefrontal/ACC regions was predictive of a better outcome to CBT for social anxiety (Whitfield-Gabrieli et al., 2015). Furthermore, a review on this subject found that lower grey matter density in the vlPFC predicted better response to SSRIs for OCD, while increased grey matter density in the rACC correlated with better improvement in PTSD symptoms (Shin et al., 2013).

Taken together, the literature suggests that there are overlapping areas of the brain involved in the manifestation of both depression and anxiety. Consequently, one could argue that comorbidity may actually serve to enhance any structural differences between classes,

rather than diminish them. Indeed, some authors have stated that it is possible that comorbid anxiety could represent more severe psychopathology (Forbes et al., 2010), and our previous results would support this; halted-improvers demonstrated more severe depressive symptoms at outset and throughout the trial, and comorbidity predicted likelihood of membership to this class (See Chapter 2). Moreover, trait anxiety was significantly higher in halted-improvers in MR-IMPACT. However, much like the neuroimaging literature on depression, studies of anxiety disorders often present mixed results, which is in part dependent on the specific anxiety disorder under investigation (Duval, Javanbakht, & Liberzon, 2015). Consequently, we cannot discount the possibility that the unequal presence of trait anxiety and comorbid conditions across our classes may have disproportionately introduced heterogeneity to our sample and affected the results. However, the small sample size precluded further investigation; excluding such a large number of patients would have resulted in too small a group of halted-improvers without comorbidity to provide meaningful analyses. Moreover, given that such a large percentage of patients presented with at least one comorbid disorder, the exclusion of patients with comorbidities would question the generalisability of the results to clinical settings.

*Limitations*

One of the biggest limitations of this current work is our sample size. Imaging data were not available on the full IMPACT sample, such that the MR-IMPACT sample represented a significantly smaller subgroup. Our study may have therefore lacked sufficient power to detect true differences between classes. However, a large proportion of imaging studies to date have reported results in much smaller overall samples (Costafreda et al., 2009; Foland-Ross et al., 2015; Gong et al., 2011; Liu et al., 2012; Reynolds et al., 2014). Our small sample therefore may not be the only reason for our null effects. Furthermore, given that the empirical modelling of 465 patients resulted in a separation of our sample into two substantially different sized classes, a further reduction of this sample to 109 resulted in only 19 patients present in the halted-improvers class. Consequently, it is possible that these proportions were too imbalanced to elicit predictive differences in neuronal structure. However, our halted-improvers class was identical in size to the smallest class produced in Schmaal and colleagues' GMM work (Schmaal et al., 2017b). Nevertheless, we caution conclusive interpretations from this work, and ask for the findings to be interpreted as exploratory, with predominantly hypothesis-informing implications. It will be important for

future work to recruit much larger samples to increase power for these specific research questions.

There is second difficulty with the interpretation of this work relating to sample size. That is; the subsequent impact of large inter-individual variability in brain variables clouding significant differences. For example, there is a great deal of variability in gyrification of the brain, and variation in gyrification has been significantly associated with both cortical thickness (negatively) and surface area measures (positively) (Gautam, Anstey, Wen, Sachdev, & Cherbuin, 2015; Hogstrom, Westlye, Walhovd, & Fjell, 2013). In particular, the presence or absence of the paracingulate sulcus varies greatly across individuals, and even within individuals per hemisphere (Fornito et al., 2008). The presence of this sulcus has been associated with differences in cortical thickness and surface area of the anterior cingulate (Fornito et al., 2008), one of the regions of interest in this work. This was not controlled for in our study. Consequently, it is likely that there was additional variance between individuals' cortical morphometry that may have impacted our results. Our sample size here limited the extent we could investigate and control for such large individual variability. However, this exploratory work highlights the importance for future work to adequately control for these factors in study design.

Finally, the investigation of structure in this present chapter served to identify an easily observable target that could differentiate patient trajectories at outset. Investigations of brain structure in this context cannot inform questions regarding the mechanisms of action underlying how any potential difference might result in symptom change for patients. While this is a critical question, it is a very different one, which IMPACT was not designed to answer. However, future work should prioritise understanding these mechanisms.

*Reflections on Methodology*

Our work decided to focus specifically on brain structure of the cortex. Within the neurological literature, there is a large corpus of work on functional magnetic resonance imaging (fMRI) in depression that show a number of illustrations of cognitive and attentional biases associating with the condition (Hamilton et al., 2012; Kaiser et al., 2015; Miller et al., 2015). There is also a growing field of fMRI studies investigating treatment response (Fu et al., 2013; Pizzagalli, 2011) and treatment effects in depression (Keedwell et al., 2010;

Kennedy et al., 2001, 2007; Ruhé et al., 2008). However, the interpretation of fMRI is still highly debated (Lener et al., 2016; van Eijndhoven et al., 2013). Particularly for task-based imaging, detection of group differences relies the test's ability to accurately challenge the desired function, and also on the chosen function eliciting a difference between groups (Duncan & Owen, 2000). Furthermore, many diverse cognitive tasks appear to recruit the same brain areas (Duncan & Owen, 2000), so the extent to which one can claim that differences relate to specific tasks is limited. More specifically, the depressive state itself is also highly variable (Lener et al., 2016; van Eijndhoven et al., 2013) so whether fMRI findings are a state-dependent feature is also unknown (Lener et al., 2016; van Eijndhoven et al., 2013). The extrapolation of reliable functional markers that are predictive of treatment response in a clinical setting is therefore quite limited (Bullmore, 2012), and argued will unlikely reach the levels of specificity and sensitivity required in clinical practice (Ritchey et al., 2011). Anatomical investigations, however, are argued a more stable avenue of investigation (Lener et al., 2016) and can be validated with histological studies (Cotter, 2002; Ongur, Drevets, & Price, 1998; Rajkowska, 2000). Consequently, results are more interpretable and can more readily apply to clinical settings (Bullmore, 2012). For this reason, I chose to focus my investigation on brain structure.

A further a priori decision was to investigate cortical thickness and surface area over GMV. There are a number of studies that have investigated the association between depression and GMV (Arnone et al., 2012; Bora et al., 2012; Koolschijn et al., 2009; Lai, 2013), as well as the relationship between GMV and treatment response (Chen et al., 2007; Fu et al., 2013; Li et al., 2010; Liu et al., 2012; Soriano-Mas et al., 2011). However, the interpretation of GMV is somewhat ambiguous (Liu et al., 2014). GMV is a mixture of folding, surface area and cortical thickness (Fischl & Dale, 2000), and authors have argued that these measures not only possess different underlying mechanisms, but are also governed by different genetic influences (Rakic, 1995; Winkler et al., 2010). The lack of specificity in this context means that it is yet unclear as to what exactly GMV differences represent in depressed patients. Indeed, Lui and colleagues (2012) stated that their null results might be explained by the lack of sensitivity in traditional methods of measuring GMV (Salvadore et al., 2011) to detect subtle variations in brain structure between treatment-sensitive and treatment-resistant depression. Conflicting results between studies therefore, could be due to the differential mechanisms of action of the disease on folding, thickness and surface-area independently.

However, recent developments in surface-based methods have allowed submillimeter differences to be detected that can differentiate between cortical thickness, surface area and folding (Fischl & Dale, 2000). This allows for a higher sensitivity to more subtle variations that may exist in psychiatric illnesses (Kuperberg et al., 2003), and a more detailed explanation of what reductions in structural brain parameters represent. For instance, histological studies in MDD found evidence of cortical thinning, and have further related this thinning to reductions in neuronal size, and neuronal and glial cell densities (Rajkowska et al., 1999). Consequently, we favoured a surface-based approach to our imaging analysis, to allow us to distinguish between these parameters of GMV and produce a more interpretable result. However, while cortical surface-based methods provide a more precise definition of brain alterations, it is important to note that the relationship between thinning and glial cell density is not straightforward. One study found that the density of glutamate receptor agonist cells was actually increased in depressed patients (Steiner et al., 2011), which has further been linked to an abolished modulatory effect of glutamate in depressed cases (Li et al., 2014). Consequently, thinning is not necessarily indicative of cell loss and may relate to other types of glial cells within the cortex. One must therefore, be careful not to extrapolate our findings to suggest a lack of group differences at a cellular level.

A final a priori decision of this current work was to focus our investigations to specific regions of the cortex, due to limited sample size and hence power. However, there is strong evidence for the role of a number of subcortical regions in depression and treatment response. These include the amygdala (Whittle et al., 2014), thalamus (Hagan et al., 2015), striatum (Forbes et al., 2010) and, most importantly, the hippocampus (Fu et al., 2013; Schmaal et al., 2016; Videbech & Ravnkilde, 2004; Whittle et al., 2014). Hippocampal abnormalities in MDD are one of the most replicated findings across structural imaging (Arnone et al., 2012; Koolschijn et al., 2009). The ENIGMA-consortium in particular, found that the hippocampus was the only subcortical structure showing significant volume reductions (1.24%) in adult MDD compared to controls, of the 9 subcortical areas investigated. Moreover, meta-analytic findings have also supported the predictive value of the hippocampus in treatment outcome (Fu et al., 2013), and some studies have shown that these findings of reduced hippocampal volume associating with poorer response are robust even against differences in the definition of response (MacQueen, Yucel, Taylor, Macdonald, & Joffe, 2008), which is of particular relevance to this current thesis. The decision to focus on

cortical structure was taken to balance such a large investigation against our small sample; we prioritised the investigation of cortical regions because prior work suggested that their influence exceeds that of subcortical regions in whole-brain predictive models of depression and treatment response (Costafreda et al., 2009; Liu et al., 2012). Nevertheless, one cannot exclude the possibility that structural differences exist between classes in subcortical regions of the brain. An investigation of subcortical structures would be an important avenue to further this work.

We chose to conduct ROI analyses to identify predictors of class membership. However, we also felt it important to conduct an additional vertex-wise, whole-brain analysis. It is crucial to note that studies of predictive models of treatment response have almost unanimously highlighted the widespread nature of neurological differences. Prior studies have emphasised that frontal, temporal, parietal, occipital regions and the cerebellum all provided large contributions to the overall predictive model of treatment response (Foland-Ross et al., 2015; Gong et al., 2011; Liu et al., 2012). Consequently, their authors advocated that, due to the nature of their methodology, their results be interpreted as discriminative networks rather than specific regions (Gong et al., 2011; Liu et al., 2012). For this reason, our whole-brain analysis allowed us a conservative investigation of the presence of potential differences in other regions of the cortex, without unnecessary inflation of type 1 errors.

A further point on our work is that our analyses were at the group-level. This was due to the exploratory nature of this work. As was discussed in Chapter 2, a limitation of this approach is that the technique is unable to predict outcomes at the individual level. That is, we did not test the accuracy at which the thickness or area of a given region can categorise new patients. Advanced analytical methods exist that can achieve this individual-level categorisation with imaging data (Zeng et al., 2012). These multivariate pattern analysis (MVPA) techniques, such as support vector machine (SVM) learning, assess the contribution of multiple voxels (or vertices) simultaneously. Studies have employed these methods in voxel-based work for diagnosing depression, and reached 94.3% accuracy with resting-state data (Zeng et al., 2012). Further, SVM analysis for treatment responsiveness showed 89% accuracy of whole-brain GMV in predicting treatment responsiveness to fluoxetine (Costafreda et al., 2009). Liu and colleagues(2012) extended these findings by showing that grey and white matter density had comparable accuracy in elucidating medication-sensitive

and resistant subtypes (both 82.9%). However, a study conducted with a larger, yet still modest sample (n=61) reported a somewhat lower predictive accuracy of grey matter density on treatment outcome, at 69.57% (Gong et al., 2011). At present, the accuracy of these models in predicting treatment response classes is not sufficient for clinical utility. In addition, there is no treatment-response study known to the current author that incorporates such approaches with surface-based methods. While MVPA methods make a necessary step towards clinical practice for prognostic MRI research, it will be important for clinical application that these methods to further incorporate an assessment of prognostic error (Nouretdinov et al., 2011). Preliminary work has described methods where prediction algorithms can maintain the level of accuracy achieved with SVM methods, but add a valid measure of confidence (Nouretdinov et al., 2011). Future prognostic imaging should prioritise adopting such methods for imaging to be used beyond research purposes in this field.

Finally, our hypotheses were driven by the findings of prior literature, which suggested that the brain regions investigated here are important structures for differentiating between response-related outcomes in depression. However, abnormalities in these regions may not necessarily be specific to the depressive condition. Many other psychiatric conditions also implicate these regions as structurally or functionally maladaptive (Goodkind et al., 2015; Menon, 2011). For instance, a recent meta analysis of 193 studies showed that GMV reduction in the ACC and insula is observed across schizophrenia, bipolar disorder, depression, anxiety, obsessive-compulsive disorder and addiction (Goodkind et al., 2015). Moreover, the authors noted that very few diagnosis-specific abnormalities were found. The wider implication of the lack of specificity of these brain regions to depression is that: either neurological abnormalities are a more general symptom of mental illness, and consequently, neuroimaging is limited in its utility to advance our understanding of condition-specific differences; or that our current, symptom-driven diagnostic approach does not adequately translate into biological constructs (in this case, brain structure). Consequently, to really progress our understanding of mental illness, classifications must begin to incorporate our growing understanding of neuroscience and genetics (Insel et al., 2010). This topic is discussed in more detail in Chapter 5.

In conclusion, this current work suggests that baseline cortical structure is not

associated with trajectories of symptom change. However, it is possible that our sample size precluded the detection of significant differences, so it is imperative for these results to be replicated in larger samples. Future work should prioritise adopting analytical approaches that can predict outcome at an individual level, and produce a measure of confidence in this prediction for these techniques to be more applicable in clinic. The neuroscientific field as a whole is currently at a crossroads where it must decide whether our current diagnostic system is sufficient, given our advancing knowledge of the brain in health and disease.

**Appendix 3A**

*The effect of SSRI prescription on cortical thickness and surface area.*

Means and SDs for cortical thickness and surface area of each region of interest (ROI) of patients who were and were not taking SSRIs at the time of scanning are shown in Table 1. Results from each logistic regression (FDR corrected) are shown in Table 2.

*Table 1: Mean cortical thickness and surface area of each ROI in patients who were and were not taking SSRIs at the time of scanning.*

| | **Cortical Thickness (mm)** | | | | **Surface Area ($mm^2$)** | | | |
|---|---|---|---|---|---|---|---|---|
| | *No SSRIs* | | *SSRIs* | | *No SSRIs* | | *SSRIs* | |
| *Region* | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* |
| *Medial OFC* | | | | | | | | |
| *Left* | 2.58 | 0.15 | 2.59 | 0.15 | 1871.65 | 255.56 | 2017.16 | 292.43 |
| *Right* | 2.60 | 0.17 | 2.52 | 0.16 | 1803.10 | 179.71 | 1907.00 | 231.68 |
| *Lateral OFC* | | | | | | | | |
| *Left* | 2.86 | 0.15 | 2.82 | 0.14 | 2744.61 | 215.43 | 2858.38 | 283.73 |
| *Right* | 2.81 | 0.15 | 2.79 | 0.14 | 2559.11 | 227.36 | 2722.92 | 292.30 |
| *SFG* | | | | | | | | |
| *Left* | 2.91 | 0.14 | 2.91 | 0.14 | 7334.75 | 786.84 | 7625.81 | 1004.50 |
| *Right* | 2.80 | 0.14 | 2.81 | 0.13 | 7021.42 | 677.65 | 7551.35 | 867.52 |
| *rMFG* | | | | | | | | |
| *Left* | 2.45 | 0.15 | 2.45 | 0.12 | 5887.99 | 667.51 | 6300.87 | 718.98 |
| *Right* | 2.36 | 0.13 | 2.37 | 0.12 | 6255.93 | 898.89 | 6364.08 | 814.10 |
| *cMFG* | | | | | | | | |
| *Left* | 2.67 | 0.15 | 2.69 | 0.12 | 2373.72 | 394.04 | 2378.27 | 444.10 |
| *Right* | 2.60 | 0.14 | 2.64 | 0.15 | 2192.89 | 401.00 | 2209.11 | 399.05 |
| *rACC* | | | | | | | | |
| *Left* | 3.07 | 0.23 | 3.11 | 0.24 | 864.31 | 175.02 | 907.89 | 188.33 |
| *Right* | 3.06 | 0.21 | 3.07 | 0.19 | 657.25 | 108.40 | 693.57 | 118.51 |
| *Insula* | | | | | | | | |
| *Left* | 3.27 | 0.12 | 3.28 | 0.13 | 2164.75 | 207.21 | 2282.03 | 291.71 |
| *Right* | 3.27 | 0.11 | 3.28 | 0.13 | 2214.24 | 271.49 | 2251.60 | 232.09 |

*OFC: Orbitofrontal Cortex; SFG: Superior Frontal Gyrus; rMFG: Rostral Middle Frontal Gyrus; cMFG: Caudal Middle Frontal Gyrus; rACC: Rostral Anterior Cingulate Cortex*

Linear regressions were conducted to investigate whether SSRI prescription at baseline predicts cortical thickness or surface area, controlling for age, gender and white matter total surface area. Regressions were conducted using the *lm()* function in the core *stats* package in R, version version 3.3.3 ("R: The R Project for Statistical Computing," n.d.). Both cortical thickness and surface area were investigated separately as variables of interest.

Results from the linear regressions are shown in Table 2 and 3 of this Appendix below. No region of interest raised concerns of multicollinearity for either variable of brain structure. All ROIs also met the assumption of independent errors and normality for cortical thickness, supporting the use of linear regressions with these data. While five regions (left SFG, left cMFG, left insula and bilateral rMFG) failed the Shapiro-Wilks test for normality for surface area, and the left SFG additionally failed the Durbin-Watson test for independent errors, permutation testing results did not substantially differ from those reported and directionality of significance remained the same (Table 4).

*Table 2: Multiple regression results for SSRI prescription on cortical thickness, adjusted for age, gender and total white matter surface area.*

| ***Region*** | ***$R^2$*** | ***B*** | ***SE B*** | ***β*** | ***p*** | ***$p_{(FDR)}$*** |
|---|---|---|---|---|---|---|
| *Medial OFC* | | | | | | |
| *Left* | 0.140 | 0.016 | 0.029 | 0.059 | .594 | .756 |
| *Right* | 0.161 | -0.064 | 0.032 | -0.182 | .050 | .422 |
| *Lateral OFC* | | | | | | |
| *Left* | 0.084 | -0.032 | 0.030 | -0.104 | .282 | .563 |
| *Right* | 0.080 | -0.001 | 0.030 | -0.004 | .970 | .993 |
| *SFG* | | | | | | |
| *Left* | 0.198 | -0.003 | 0.027 | 0.106 | .240 | .560 |
| *Right* | 0.136 | 0.037 | 0.026 | 0.132 | .160 | .492 |
| *rMFG* | | | | | | |
| *Left* | 0.133 | 0.024 | 0.027 | 0.082 | .387 | .566 |
| *Right* | 0.122 | 0.023 | 0.025 | 0.088 | .354 | .566 |
| *cMFG* | | | | | | |
| *Left* | 0.111 | 0.044 | 0.029 | 0.145 | .129 | .492 |
| *Right* | 0.059 | 0.057 | 0.030 | 0.185 | .060 | .422 |
| *rACC* | | | | | | |
| *Left* | 0.146 | 0.061 | 0.045 | 0.127 | 0.176 | .492 |
| *Right* | 0.057 | 0.035 | 0.042 | 0.082 | .404 | .566 |
| *Insula* | | | | | | |
| *Left* | 0.022 | <0.001 | 0.026 | <.001 | .993 | .993 |
| *Right* | 0.028 | 0.010 | 0.024 | 0.041 | .681 | .795 |

*OFC: Orbitofrontal Cortex; SFG: Superior Frontal Gyrus; rMFG: Rostral Middle Frontal Gyrus; cMFG: Caudal Middle Frontal Gyrus; rACC: Rostral Anterior Cingulate Cortex*

*Table 3: Multiple regression results for SSRI prescription on surface area, adjusted for age, gender and total white matter surface area.*

| ***Region*** | ***$R^2$*** | ***B*** | ***SE B*** | ***β*** | ***p*** | ***$p_{(FDR)}$*** |
|---|---|---|---|---|---|---|
| *Medial OFC* | | | | | | |
| *Left* | 0.456 | 63.36 | 43.12 | 0.109 | .145 | .337 |
| *Right* | 0.462 | 41.49 | 31.65 | 0.097 | .193 | .337 |
| *Lateral OFC* | | | | | | |
| *Left* | 0.439 | 42.977 | 38.926 | 0.083 | .272 | .346 |
| *Right* | 0.499 | 86.318 | 39.201 | 0.157 | .030 | .209 |
| *SFG* | | | | | | |
| *Left* | 0.657 | -26.253 | 108.233 | -0.014 | .809 | .857 |
| *Right* | 0.550 | 296.8 | 111.4 | 0.180 | .009 | .125 |
| *rMFG* | | | | | | |
| *Left* | 0.554 | 178.1 | 100.3 | 0.119 | .079 | .220 |
| *Right* | 0.588 | -210.0 | 118.0 | -0.115 | .078 | .220 |
| *cMFG* | | | | | | |
| *Left* | 0.455 | -116.8 | 64.02 | -0.136 | .071 | .220 |
| *Right* | 0.420 | -78.55 | 64.21 | -0.094 | .224 | .346 |
| *rACC* | | | | | | |
| *Left* | 0.489 | -10.99 | 27.24 | -0.029 | .687 | .802 |
| *Right* | 0.411 | 3.308 | 18.31 | 0.014 | .857 | .857 |
| *Insula* | | | | | | |
| *Left* | 0.488 | 42.93 | 37.01 | 0.084 | .249 | .346 |
| *Right* | 0.558 | -47.857 | 36.355 | -0.088 | .191 | .337 |

*OFC: Orbitofrontal Cortex; SFG: Superior Frontal Gyrus; rMFG: Rostral Middle Frontal Gyrus; cMFG: Caudal Middle Frontal Gyrus; rACC: Rostral Anterior Cingulate Cortex*

Standardised residuals of the final model were inspected to assess for whether the model contained any outliers of concern. For the majority of regions, the percentage of the total sample of cases showing residuals greater than ±2.58 was less than 1%, and in exceptions, never exceeded 2%. Similarly, for the majority of regions, the percentage of the total sample of cases showing residuals greater than ±1.96 was less than 5%, and in exceptions, never exceeded 7% for either structural variable. Therefore the models can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases were exerting undue influence over the models. While there were a number of cases per regions and structure variables that suggested a potential problem with 2 or 3 times greater than average leverage, the Cook's distance never exceeded 1 (Cook & Weisberg, 1982), indicating that the fit would not significantly change upon removal of these cases.

There were no main effects of SSRI prescription on any ROI for both structural variables of interest. As such patients taking SSRIs at the time of scanning remained in the main analysis (n=109).

*Permutation testing for SSRI predictor of surface area.*

Linear regressions with permutation testing were conducted using the *lmp()* function in the core *lmPerm* package in R version 3.3.3. Results from the permutation tests, for regions that violated assumptions are shown in Table 2.

Table 4: *Permutation testing results for SSRI predictor of surface area.*

| ***Region*** | ***$R^2$*** | ***B*** | ***SE B*** | ***β*** | ***P(permuted)*** |
|---|---|---|---|---|---|
| *Superior Frontal Gyrus* | | | | | |
| *Left* | 0.657 | -26.253 | 108.233 | -0.014 | .824 |
| *Rostral Middle Frontal Gyrus* | | | | | |
| *Left* | 0.554 | 178.1 | 100.3 | 0.119 | .322 |
| *Right* | 0.588 | -210.0 | 118.0 | -0.115 | .106 |
| *Caudal Middle Frontal Gyrus* | | | | | |
| *Left* | 0.455 | -116.8 | 64.02 | -0.136 | .085 |
| *Insula* | | | | | |
| *Left* | 0.488 | 42.93 | 37.01 | 0.084 | .540 |

**Appendix 3B**

*Permutation testing for surface area ROI predictors of class membership.*

Results from the permutation tests, for regions that violated assumptions are shown in Table 1.

*Table 1: Permutation test results for surface area ROI predictors of class membership.*

| | **Surface area ($mm^2$)** | | |
|---|---|---|---|
| *Region* | *OR* | *95%CI* | *P(permuted)* |
| *Right Rostral Middle Frontal Gyrus* | 1.00 | 0.99-1.00 | .235 |
| *Right Caudal Middle Frontal Gyrus* | 1.00 | 0.99-1.00 | .369 |

## Chapter 4: Physiological predictors of trajectory classes: the role of cortisol

### Introduction

Cortisol is an endogenous steroid hormone and has a multitude of functions including being a critical component of the body's stress response. It increases blood sugar levels, stimulates the metabolism of fats, carbohydrates and protein, and also acts as an immunosuppressant with anti-inflammatory properties (Seifter, Sloane, & Ratner, 2005). The cortisol release at times of extrinsic or intrinsic demands on metabolism enables the body to respond adaptively to environmental challenge. The depressed patient however may exhibit behaviours indicative of maladaptive stress responses. For instance, depressed medical students have reported higher incidences of frustration and emotional reactions to stressors than non-depressed students (Saravanan & Wilks, 2014), and there is substantial overlap between symptoms of depression and chronic stress (APA, 2013). Consequently, it is unsurprising that elevated circulating cortisol is one of the most consistently reported physiological abnormalities in depressed cases (Ehlert, Gaab, & Heinrichs, 2001a; Guerry & Hastings, 2011; Pruessner, Hellhammer, Pruessner, & Lupien, 2003). A meta-analysis of 40 years of research showed that as many as 73% of patients displayed cortisol values greater than the median of non-depressed controls (Stetler & Miller, 2011). However, the release of cortisol occurs through a complex series of biological actions governed by the hypothalamic-pituitary-adrenal (HPA) axis (Pariante & Lightman, 2008). Currently however it is not entirely clear exactly where specific abnormalities in the HPA system might reside in depressed patients. Nevertheless, it is clear that at least for some patients (Rush et al., 1996; Cinnamon Stetler & Miller, 2011), dysfunction in this endocrine system may play a contributory role in the development of their illness (Ehlert, Gaab, & Heinrichs, 2001b; Varghese & Brown, 2001) and warrants further investigation.

#### *The HPA-axis stress response*

Briefly, the cortisol response begins at the hypothalamus, which receives stress information from a number of neuronal inputs, such as the locus coeruleus-norepinephrine system (which promotes immediate action to environmental challenges (Cunningham & Sawchenko, 1988)), as well as the amygdala (Gray, Carney, & Magnuson, 1989) and hippocampal (Herman & Cullinan, 1997) regions of the brain. Projections from these systems travel to the hypothalamus, where they stimulate the secretion of corticotrophin-releasing hormone

(CRH) from the parvocellular neurons in the paraventricular nucleus. CRH then triggers the anterior pituitary to secrete adrenocorticotropic hormone (ACTH) into the blood, which upon reaching the adrenal gland, initiates the synthesis and secretion of cortisol from the zona fasciculata layer of the adrenal cortex (Nicolson, 2008). Once cortisol enters the circulatory system, it acts on target tissues by binding to corticosteroid receptors that are present throughout the body and in many parts of the brain, with various consequential outcomes (Gold, 2014; Nicolson, 2008).

*Cortisol and depression*

There is a corpus of evidence showing that depressed patients express abnormalities at many of the HPA-axis stages, which could all contribute to elevated cortisol seen in depression (Ehlert et al., 2001a; Guerry & Hastings, 2011; Pruessner et al., 2003). Firstly, the amygdala and hippocampus govern many behaviours associated with depression, such as fear conditioning (Knight, Smith, Cheng, Stein, & Helmstetter, 2004) and episodic memory (Burgess, Maguire, & O'Keefe, 2002). Both have been shown to exhibit structural deficits, as well as abnormal functional activity, in depressed patients (Fu et al., 2013; Schmaal et al., 2016; Videbech & Ravnkilde, 2004; Whittle et al., 2014). This could imbalance this system at the very start of the chain. The hypothalamus itself has also demonstrated a number of anatomical and functional abnormalities in depressed cases. Firstly, depressed patients have been shown to exhibit a four-fold increase in the number of CRH-expressing neurons in the hypothalamus, which is known to correlate positively with CRH secretory activity (Raadsheer, Hoogendijk, Stam, Tilders, & Swaab, 1994). Basal plasma levels of CRH were found to be significantly higher in depressed cases (Galard, Catalán, Castellanos, & Gallart, 2002), suggesting that hypothalamic hyper-activity might significantly contribute to elevated cortisol. Perhaps as a consequence, elevated levels of circulating ACTH have been documented (Cinnamon Stetler & Miller, 2011), which in itself could lead to higher levels of cortisol secretion. However, many studies actually report blunted ACTH-response to CRH (Gold et al., 1984; Holsboer, Bardeleben, Gerken, Stella, & Muller, 1984; Schüle, Baghai, Eser, & Rupprecht, 2009), which is argued a result of chronically elevated levels of CRH causing a down-regulation of CRH-receptors in the anterior pituitary (Schüle et al., 2009). Indeed, reduced CRH1-receptor binding has been illustrated in cases of suicide (Nemeroff, Owens, Bissette, Andorn, & Stanley, 1988). Excess plasma cortisol, therefore, may be a potential consequence of increased sensitivity of the adrenal cortex to ACTH in depressed cases

(Jaeckle, Kathol, Lopez, Meller, & Krummel, 1987), as a compensation for this blunted ACTH response earlier in the system. The adrenal glands themselves have also shown up to a 70% increase in volume during depressive episodes (Kahl et al., 2015; Nemeroff et al., 1992; Rubin, Phillips, Sadow, & McCracken, 1995). Interestingly, this appears to return to normal in remission (Rubin et al., 1995). However, while increased adrenal volume has been associated with increased cortisol secretion (Golden et al., 2007), it is not a consistent finding (Nemeroff et al., 1992; Rubin et al., 1995).

Taken together, there are many aspects of the HPA-axis functioning that may be dysfunctional in depression, resulting in hypercortisolaemia. Hypercortisolaemia can have detrimental effects on the body, including impaired cardiovascular and immune functioning (Jyotsna, Naseer, Sreenivas, Gupta, & Deepak, 2011; Ntali, Grossman, & Karavitaki, 2015), as well as neural atrophy (Resmini, Santos, & Webb, 2016). Moreover, hypercortisolaemia has been found to have a causal relationship with symptoms that resemble MDD (Sher, 2004). These include tiredness (Melamed et al., 1999), anhedonia (Putnam, Pizzagalli, Gooding, Kalin, & Davidson, 2008), irritability (Melamed et al., 1999), and impaired cognitive and decision-making abilities (Dias-Ferreira et al., 2009).

Another distinctive, yet critical, characteristic of the HPA-axis is its ability to self-regulate. A negative feedback loop exists, such that once an appropriate dose of cortisol has reached the target tissues, further cortisol inhibits the secretion of CRH by the hypothalamus and reduces pituitary sensitivity to CRH (Nicolson, 2008). Self-regulation of the HPA-axis primarily functions through the action of cortisol binding to glucocorticoid (GR) receptors (as opposed to mineralocorticoid (MR) receptors) in target tissues. GR receptors only activate under higher concentrations of cortisol, to restore homeostasis following an acute imbalance (Nicolson, 2008; Reul & De KLoet, 1985). However, impaired negative feedback is a widely supported theory of abnormal cortisol levels in depression (Lopez-Duran, Kovacs, & George, 2009; Mokhtari, Arfken, & Boutros, 2013; Nicolson, 2008). Firstly, depressed cases reliably exhibit reduced functionality of GR receptors (Pariante & Miller, 2001; Rohleder, Wolf, & Wolf, 2010), as well as a reduction in GR receptor numbers in some (Sallee, Nesbitt, Dougherty, & Hilal, 1995), but not all, studies (Rupprecht et al., 1991). Authors have also hypothesised that the strength of negative feedback elicited from GR receptors could be related to the degree of heritability in the cortisol response (Bartels, de Geus, Kirschbaum,

Sluyter, & Boomsma, 2003). Furthermore, animal models have shown that antidepressant medications have a positive and direct effect on GR receptors, through restoring functionality and expression (Pariante & Miller, 2001). These antidepressant effects on GR receptors have subsequently been suggested to underlie the induced neurogenesis observed in the human hippocampus following treatment (Anacker et al., 2011). All of these contributory factors fundamentally suggest impairments in depression exist in the system's ability to self-regulate, resulting in hypercortisolaemia.

For clinical research, the self-regulatory mechanism of cortisol is often tested using the dexamethasone (DST) suppression test (Rush et al., 1996), which involves the administration of a synthetic glucocorticoid, dexamethasone, before sleep. This drug acts to suppress CRH release, akin to normal endogenous cortisol resulting in a post-administration decrease in morning cortisol levels under normative HPA-functionality (i.e. a suppression effect on cortisol) (Nicolson, 2008). This test has now been superseded by the dexamethasone/corticotrophin-releasing-hormone (DEX-CRH) test, which additionally investigates the functionality of CRH on ACTH and cortisol, under the influence of dexamethasone (Mokhtari et al., 2013). Under normative function, CRH would counteract dexamethasone, and reinstate secretion of ACTH and consequently, cortisol (Mokhtari et al., 2013). The DEX-CRH test therefore has an added benefit for delineating at which point in the pathway an abnormality lies (Mokhtari et al., 2013). A consistent finding in depression is that depressed cases exhibit non-suppression in the DST test(Rush et al., 1996), and cortisol hypersecretion with the DEX-CRH test (Mokhtari et al., 2013). Both findings are suggestive of hypercortisolaemia, via reduced sensitivity to the negative feedback response to high levels of corticosteroids (Lopez-Duran et al., 2009; Mokhtari et al., 2013).

It is important to note that another relevant, yet complicated, characteristic of the HPA axis is that it expresses both a basal activity pattern of cortisol release and a phasic activity pattern, and it is possible that they have influential roles on one-another (Dickstein et al., 1991; Young, Abelson, & Lightman, 2004). Alterations in a diurnal pattern of cortisol release may affect phasic responses to external stressors and conversely, should an acute stressor become more chronic, these phasic responses could affect diurnal rhythms (Young, Abelson, et al., 2004). During a typical circadian cycle, the HPA-axis' basal activity follows a diurnal pattern of cortisol release. Levels begin to rise towards waking, and reach a peak usually

between 30-40 minutes after waking (Lovallo & Thomas, 2000). This is known as the Cortisol Awakening Response (CAR). Following this, cortisol levels gradually decline through the day to reach a low point before sleep (Lovallo & Thomas, 2000). This diurnal pattern may help the body to transition appropriately between activity levels of sleep and waking (Edwards, Evans, Hucklebridge, & Clow, 2001).

A number of different methodologies have been used to investigate basal levels of cortisol in depressed cases, which makes gaining a coherent narrative challenging (Guerry & Hastings, 2011). This includes variation in the biological type of assay (blood, saliva, urine) and the cortisol index itself (waking, evening, 24hour, CAR) (Fischer, Macare, & Cleare, 2017; Fischer, Strawbridge, Vives, & Cleare, 2017; Guerry & Hastings, 2011). However despite methodological heterogeneity, there is a general convergence that disturbances in this rhythm are both a risk factor for, and feature of current depression (Herbert, 2013; Lopez-Duran et al., 2009). There is also tentative evidence that morning and evening cortisol may be driven by separate mechanisms. Morning cortisol levels have demonstrated a significant degree of heritability, which was not observed for evening measures (Bartels et al., 2003; Wüst, Federenko, Hellhammer, & Kirschbaum, 2000). Indeed, the authors postulated that variation in evening cortisol may be more strongly influenced by environmental factors than genetic components (Wüst et al., 2000). In line with these findings, elevated morning cortisol has more consistently been described as a vulnerability marker for depression onset (Goodyer, 2000; Harris et al., 2000; Owens et al., 2014), rather than a characteristic of a current pathological state. However, this assertion may be more nuanced. Owens and colleagues (2014) used a large population based adolescent cohort (n=~1800) to show that morning (but not evening) cortisol in the adolescent population was associated with a 14 fold increase in the presence of clinical depression only in males with higher levels of pre-existing depressive symptoms. Replication of these findings remains to be firmly established. In contrast, elevated evening cortisol has been reported in patients across the life-course with current depression (Dahl et al., 1991; Van den Bergh & Van Calster, 2009; Vreeburg et al., 2009), along with blunted CAR response (Huber, Issa, Schik, & Wolf, 2006; C Stetler & Miller, 2005), such that the overall diurnal rhythm is flatter (Van den Bergh et al., 2009; Van den Bergh, Van Calster, Pinna Puissant, & Van Huffel, 2008). Interestingly however, this flattening was found to be predominantly driven by elevated evening cortisol (Van den Bergh & Van Calster, 2009). Similar patterns of a reduced CAR and elevated evening cortisol

have also been found in people with high ruminative tendencies (Cropley, Rydstedt, Devereux, & Middleton, 2015), a common feature of depression (Nolen-Hoeksema, 1991). However, not all studies found a blunted CAR to be present in depressed cases (Schmidt, Laessle, & Hellhammer, 2013; Vreeburg et al., 2009).

On top of this basal activity, phasic cortisol secretions function to respond appropriately to external stressors (Smyth et al., 1998). In depression, there has been a wealth of research highlighting the potentially causal relationship between stressful life events and illness onset (Guerry & Hastings, 2011; Hammen, 2005; Kendler, Karkowski, & Prescott, 1999). Mazure's review(1998) of the literature estimated that a severe life event preceded a depressive episode in up to 80% of cases in community samples. The stress caused from daily hassles is also often noted as a significant contributor to the development and maintenance of depressive illness (Sher, 2004). Furthermore, and more so than the nature of the event itself, individual differences in response to these stressors, both physiologically and psychologically, can impact the risk of developing depression (Guerry & Hastings, 2011). Physiologically, meta-analytic findings have shown that depressed individuals exhibit an abnormal rise in cortisol in response to psychological stress, but also a slower recovery to baseline (Burke, Davis, Otte, & Mohr, 2005). Taken together, this suggests that phasic cortisol responses are also disturbed in people with depression.

Overall, it is clear that many functions of the HPA-axis may be disturbed in depression. Whether the aspects are part of the generation of the cortisol response, its appropriate suppression, or a dynamic interplay between basal and phasic responses remains unclear. It is quite likely that they all play some role to varying degrees, and teasing these mechanisms will be a major work of this field. However, irrespective of mechanism, elevated circulatory cortisol appears a significant characteristic for some patients, and may itself provide a viable, convenient marker to investigate variations in response to treatment within a depressed cohort.

*Cortisol and treatment response in depression*

Studies of disrupted HPA mechanisms in depression have noted that there is a great deal of variability between patients (Chida & Steptoe, 2009; Dedovic & Ngiam, 2015; Stetler & Miller, 2011). Some authors have argued that this heterogeneity might be a consequence of

differing levels of symptom severity in patient samples (Chida & Steptoe, 2009). Indeed, it is a common finding that cortisol levels positively correlate with the severity of depressive symptoms, as well as perception of stress (Pruessner et al., 2003). In one meta-analysis, the observed effect sizes of elevated cortisol markedly decreased when studies of hospitalised patients were removed (Stetler & Miller, 2011). Furthermore, comorbidity (an index often associated with clinical severity (Shelli Avenevoli et al., 2015)) has been associated with elevated waking cortisol (Vreeburg et al., 2009). Hypercortisolaemia therefore, might be a specific feature of severe, or clinically complex, cases of depression. These in turn, as previously stated, are the most consistent behavioural predictors of non-response (Curry et al., 2006; Goodyer et al., 1997; Wilkinson et al., 2009). Indeed, our previous work outlined in Chapter 2 illustrated the significance of comorbidity and baseline severity in both the characterisation and prediction of unfavourable symptom trajectories following treatment. Consequently, the utility of cortisol as a biological marker of response has gained growing attention over recent years (Fischer & Cleare, 2017; Fischer, Macare, et al., 2017; Fischer, Strawbridge, et al., 2017; Horstmann & Binder, 2011).

There is a greater amount of work in the field of treatment response for pharmacological interventions, as theories of the interactions between antidepressants and the HPA-axis are more developed (Pariante et al., 2001). For instance, it is known that a dynamic interplay between the HPA-axis and serotonergic (5-HT) systems exists (the serotonin system is the primary target for the most common antidepressants, SSRIs) (Diaz et al., 2012; Wong & Licinio, 2004). CRH and 5-HT receptors occupy similar areas in the brain (Hammack et al., 2003; Joëls, Hesen, & de Kloet, 1991), and under stress, CRH, as well as MR/GR expression, can have modulatory effects on 5-HT function in these areas (Beck, Choi, List, Okuhara, & Birnstiel, 1996; Joëls et al., 1991; Valentino & Commons, 2005). Consequently, some argue that antidepressants exert their therapeutic effect, at least in part, on restoring HPA-axis regulation (Barden, Reul, & Holsboer, 1995). Indeed, studies have found that antidepressant medications lower the CAR, measured through salivary samples (Ruhé et al., 2015). It is also hypothesised that this process occurs through up-regulation of MR and GR receptor expression by antidepressants, restoring their sensitivity to cortisol and the consequent negative feedback functionality (Raison & Miller, 2003; Schüle, 2007). Ising and colleagues (2007) found that an early normalisation of elevated cortisol response to the DEX-CRH test within the first two weeks of treatment was associated with increased remission rates at 5

weeks. Taken together, this suggests that standard antidepressants might facilitate HPA-axis rebalancing and thus, response to treatment (Schüle, 2007).

From the perspective outlined above, it is plausible that patients with the greatest deviations in baseline cortisol might struggle to rebalance their HPA-axis dysfunction sufficiently to the extent necessary for therapeutic effect, and consequently, would demonstrate resistance to antidepressants (Horstmann & Binder, 2011; Young, Altemus, et al., 2004). Indeed, a number of studies that employed the DST or DEX-CRH test have found that patients who demonstrated DST non-suppression before admission, had significantly poorer clinical outcomes than those who elicited a more normative response to DST (Juruena et al., 2009; McKnight, Nelson-Gray, & Barnhill, 1992).The predictive value of cortisol non-suppression was even evident in patients who were already classified as resistant to two or more antidepressants (Juruena et al., 2009). Abnormal HPA-axis functioning through DST testing has also been associated with risk of relapse in a number of studies (Schüle, 2007; Zobel et al., 2001, 1999). For instance, patients demonstrating cortisol non-suppression to the DEX-CRH test at either pre or post-treatment were more likely to relapse within the next 6 months (Zobel et al., 1999), and moreover, sustained cortisol elevation in response to DEX-CRH in remitted patients was found to have an estimated 4 to 6 fold increased risk of relapse at 6 months (Zobel et al., 2001). A meta-analytic study concluded that in high quality studies, DST non-suppression associated with non-response to antidepressants (Fischer, Macare, et al., 2017), however, not all studies support this direction of effect (Binder et al., 2009).

The relationship between basal levels of cortisol and non-response to pharmacological interventions have been less well-studied, and consequently, shows a lack of consensus. Elevated pre-treatment basal cortisol levels have been found in patients who had relapsed, measured by evening blood sampling, and this identified patients that required continued treatment to maintain remission (O'Toole, Sekula, & Rubin, 1997). Hypercortisolaemia may therefore exist across various methodologies of cortisol measurement, and act as a robust indicator of a resistant subtype of depression. However, inconsistencies exist in the literature (Klimes-Dougan et al., 2018) and meta-analytic findings call for more controlled research to be conducted in basal measures (Fischer, Macare, et al., 2017). While the authors of this review found that high levels of basal cortisol reported from non-invasive

measures (saliva or urine) discriminated non-responders from responders to antidepressant medication, unadjusted statistical analyses were associated with these positive results (Fischer, Macare, et al., 2017). Moreover, some authors suggest that the prognostic utility of cortisol may be gender specific (Binder et al., 2009).

Psychotherapies are the most commonly prescribed treatments for depression (NICE, 2015, NICE, 2016), and there is a large evidence base supporting their effectiveness (Compton et al., 2004; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012; Midgley & Kennedy, 2011). These treatments often report similar efficacy to pharmacological therapies (Cuijpers, Berking, et al., 2013; Khan et al., 2012), and some authors have suggested models of SSRI action that reflect similar mechanisms to psychological therapies (Harmer & Cowen, 2013). Consequently, hypotheses regarding cortisol dysfunction from pharmacological studies could feasibly extend to those investigating response to psychological treatments. Indeed, McKnight and colleagues (1992) concluded that abnormal HPA function measured by the DST test predicted poorer response to both types of treatment, but failed to find any differential treatment effects. Furthermore, they found that these abnormalities normalised following treatment with both psychological and pharmacological modalities, supporting the hypotheses that medications and psychotherapies target similar mechanisms, and that abnormal cortisol levels influence their effectiveness.

The field of research investigating cortisol as a predictor of response to psychological therapies is small, and complicated by a number of factors. A recent meta-analysis only identified 8 articles meeting selection criteria (Fischer, Strawbridge, et al., 2017), and as well as investigating a range of cortisol methodologies (Robbins, Alessi, & Colfer, 1989; Thase et al., 1996; Thase, Simons, & Reynolds, 1993), these studies investigated a range of psychological therapies, including psychosocial (Robbins et al., 1989), CBT (Thase et al., 1996, 1993) and interpersonal therapy (Gunlicks-Stoessel, Mufson, Cullen, & Klimes-Dougan, 2013). However, in line with the findings from the pharmacological literature, the results of the meta-analysis found that overall, elevated pre-treatment cortisol levels were associated with unfavourable treatment outcome (Fischer, Strawbridge, et al., 2017). The effect was apparent for both basal cortisol levels (Thase et al., 1996, 1993), and cortisol levels in response to a stressor or challenge tests (Robbins et al., 1989), although the evidence for basal measures were less assertive (Fischer, Strawbridge, et al., 2017). Furthermore, not all

studies in this review fully supported the direction of effect (Gunlicks-Stoessel et al., 2013). Gunlicks-Stoessel and colleagues (2013) for example, reported that higher salivary cortisol levels in response to a stressful conflict-negotiation task were predictive of a better clinical outcome to interpersonal psychotherapy. Interpersonal therapy has been shown to be particularly beneficial for patients whose sense of self is reliant on their social relationships (Horowitz, Garber, Ciesla, Young, & Mufson, 2007). The authors suggested that elevated cortisol to tasks that target this specific preference could identify those patients who would benefit from treatment that specifically targets interpersonal relationships (Gunlicks-Stoessel et al., 2013). However, when Fischer and colleagues (2017) adjusted for the small sample size used in this study (n=15) in their meta-analysis, the effect-size was 0.

Another factor complicating the investigation of psychological treatment response is that different psychosocial factors have been found to associate with differential cortisol alterations (Chida & Steptoe, 2009). Some stressors relate to increased cortisol levels, while others relate to decreased levels (Chida & Steptoe, 2009). Considering the wide variety of psychotherapies that exist, it is possible, therefore, that depressed patients with elevated cortisol may respond positively to one type of therapy and negatively to another, depending on the target of the treatment modality in the psychological work. For cognitive-based therapies, elevated cortisol might hinder patients engaging in therapeutic tasks, due to the negative effect of cortisol on cognitive abilities, such as concentration and memory (Dias-Ferreira et al., 2009; Schlosser, Wolf, & Wingenfeld, 2011). It would follow that these patients would exhibit a greater degree of this cognitive impairment and thus find this type of therapy more challenging (Fischer, Strawbridge, et al., 2017; Thase et al., 1996). Indeed, when the treatment modality places a strong emphasis on cognition (such as CBT), studies have indicated that those with abnormal cortisol functioning often show a poorer response to treatment (Bockting et al., 2006; Holland, Schatzberg, O'Hara, Marquett, & Gallagher-Thompson, 2013). This effect was observed in both elevated morning salivary cortisol (Bockting et al., 2006), and in flatter diurnal curves; whereby morning salivary cortisol were elevated and additionally, failed to lower sufficiently by evening (Holland et al., 2013). CBT alone has also been found to have a poorer response rate in those with elevated urinary free cortisol (Thase et al., 1996, 1993) and elevated levels following DST challenge (Robbins et al., 1989). Treatment often requires augmentation with medications in these patients (Robbins et al., 1989), suggesting that elevated cortisol might index a particularly resistant subgroup

of depressed cases to cognitive therapies that require a higher intensity of treatment. However, it is worth noting that one study had a number of methodological limitations, including only defining a favourable response based on improvement in mood and anhedonia alone (Robbins et al., 1989). Considering the aforementioned associations between hypercortisolaemic conditions and cognitive impairment in depression (Dias-Ferreira et al., 2009; Schlosser et al., 2011), it would have been beneficial for this aspect of depressive symptomatology to be included in assessment of response. Moreover, Thase and colleagues (1993) highlighted that other factors, such as gender, comorbidity and anomalous sleep profiles also associated with poorer response. All these factors have also been associated with differential cortisol functioning (Horstmann & Binder, 2011; Oquendo et al., 2003; Rodenbeck, Huether, Rüther, & Hajak, 2002; Verma, Balhara, & Gupta, 2011; Zorn et al., 2017).

*Cortisol in adolescent depression*

Considering the population of my sample of investigation, it is important to consider the impact of the development of HPA-axis function on depression aetiology. It would be premature to assume a simple extension from the above literature, which has been predominantly adult studies, would apply to adolescent depression. Adolescence is characterised by a significance rise in basal cortisol levels from childhood, and is also a developmental stage where HPA-reactivity to stressors markedly increases (Gunnar & Quevedo, 2007; Gunnar, Wewerka, Frenn, Long, & Griggs, 2009). Indeed, stressors of objectively similar magnitude appear to affect the HPA-axis to a greater degree in adolescence than adults (Lupien, McEwen, Gunnar, & Heim, 2009). Adolescence is also known as a period of development where there is a notable increase in frequency of stressful life events (Larson & Ham, 1993). These events are often interpersonal in nature (Rudolph & Hammen, 1999), and the extent of such stressors, like peer victimisation, positively associates with HPA-axis dysfunction (Knack, Jensen-Campbell, & Baum, 2011). Together, this suggests that adolescence is a period where the HPA-axis may be particularly vulnerable to pathology. Simultaneously, the risk for depression also markedly increases during pubertal age (Avenevoli et al., 2008).

Some authors have advocated that the theory of abnormal HPA-axis function in *pre-pubertal* depression is unconvincing (Birmaher et al., 1996), as this stage of development is

characterised by very low cortisol reactivity to stressors (Gunnar & Quevedo, 2007). However, the relationship between cortisol and depression strengthens with age (Guerry & Hastings, 2011), and for *adolescent* depression, the developmental physiology and environmental interplay noted above suggests that cortisol dysregulation in adolescence should be further investigated. There is evidence across methodologies (Adam et al., 2010; Goodyer, Herbert, & Tamplin, 2003; Halligan, Herbert, Goodyer, & Murray, 2007; Lopez-Duran et al., 2009; Rao, Hammen, Ortiz, Chen, & Poland, 2008), particularly from studies of more vigorously controlled basal measures (Adam et al., 2010; Goodyer et al., 2003; Mathew et al., 2003), suggesting that cortisol dysregulation is a presenting characteristic for at least some adolescent patients. A meta-analytic review supported these findings, stating that depressed youths display symptoms of hypercortisolemia across both DST challenge test and measures of basal cortisol (Lopez-Duran et al., 2009). The authors concluded that cortisol dysregulation in youth appears to relate specifically to negative feedback mechanisms and may affect the entire diurnal pattern (Lopez-Duran et al., 2009).

Two of the most recent meta-analyses investigating the relationship between cortisol and treatment response in depression highlighted the scarcity of research in adolescent populations (Fischer, Macare, et al., 2017; Fischer, Strawbridge, et al., 2017). Of the studies that do exist, all are limited by small samples (Gunlicks-Stoessel et al., 2013; Klimes-Dougan et al., 2018; Robbins et al., 1989). Only two (Gunlicks-Stoessel et al., 2013; Robbins et al., 1989) of the eight studies included in Fischer and colleagues' (2017) meta-analysis investigating response to psychological therapies were conducted in adolescent samples. Furthermore, no pharmacological study included in the meta-analysis focused specifically on this age bracket (Fischer, Macare, et al., 2017). However, for pharmacological treatments, the authors found that studies that failed to control for age were more likely to report significant differences in cortisol between responders and non-responders. This suggests that age (and its underlying maturational components) may be an important moderator in the relationship between cortisol and antidepressant response, particularly as independent relationships exist between older age and elevated cortisol (Guerry & Hastings, 2011), and older age and poorer treatment outcome (Corey-Lisle, Nash, Stang, & Swindle, 2004; Uher et al., 2010).

Only one study to date, published after the meta-analysis (Fischer, Macare, et al., 2017), investigated neuroendocrine predictors of response to SSRIs in a specifically adolescent sample (Klimes-Dougan et al., 2018). This study, while limited by sample size, found that elevated cortisol levels at the peak of the stressor-induced cortisol response function, was actually associated with a more favourable outcome. This finding was supported by a study investigating cortisol levels as predictors of response to interpersonal therapy in adolescence (Gunlicks-Stoessel et al., 2013), and conflicts with the direction of effect reported in adult studies (Fischer, Macare, et al., 2017; Fischer, Strawbridge, et al., 2017). However, in another study, DST non-suppression, which is suggestive of elevated cortisol, was linked to poorer response to psychosocial treatment (Robbins et al., 1989). It remains unclear whether these conflicting findings between the adult and adolescent literature are a reflection of the developmental age of the sample, or of unique and differential effects of the different types of treatment, as discussed previously (Gunlicks-Stoessel et al., 2013). Coupled with the small number of studies available to date, the field is currently in a state of ambiguity. More research is necessary to understand the prognostic value of cortisol in the treatment of adolescent depression.

*Evening cortisol*

Some studies have suggested that the period before sleep onset is of particular importance for mood disorders in adolescence (Dahl et al., 1991; Forbes et al., 2006; Van den Bergh & Van Calster, 2009). For instance, adolescents with depression demonstrated higher plasma cortisol levels prior to sleep onset than controls (Forbes et al., 2006). Van den Bergh and colleagues (2009) also found that high depressive symptoms were related to a flatter diurnal rhythm and elevated evening cortisol. Moreover, they concluded that only elevated evening cortisol showed predictive utility for identifying symptom severity. In line with these findings, Dahl and colleagues (1991) found that plasma cortisol levels were similar between depressed adolescents and controls across the 24 hour period, except for just prior to sleep onset. Here, elevated cortisol was observed in the depressed group. Interestingly, this effect was driven by the most severe cases. There is currently a lack of studies investigating how cortisol levels at this particular time of day might relate to treatment response in adolescence, but case-control data would suggest that elevated evening cortisol may be of particular importance in differentiating the more severe cases of depression in adolescence, which in turn, may link to treatment non-response.

A final important observation of this field to date is that all studies have defined response a priori. A number of studies have taken remission (O'Toole et al., 1997), or a percentage decrease in symptoms (Holland et al., 2013; Juruena et al., 2009) to identify a favourable end-of-treatment outcome, while others have investigated response on a continuous scale (Gunlicks-Stoessel et al., 2013; Thase et al., 1996). Chapter 1 of this thesis details the limitations of such approaches. No study has yet investigated associations with cortisol and response in empirically-derived trajectory classes. GMM and latent class studies to date limit themselves to the investigation of behavioural measures and have not extended their work into biological variables. Doing so would mark a significant step forward in both fields of research.

*Objectives and Hypotheses*

The objective of this next piece of work was to conduct a secondary analysis of another subsample of adolescents enrolled in the IMPACT trial (Goodyer et al., 2011). This subsample was recruited into the hormonal study associated with IMPACT (IMPACT-Genes and Hormones (IMPACT-GH)), providing saliva samples to assess cortisol levels. The specific aim of this piece of work was to investigate the relationship between empirically derived trajectory classes, as defined in Chapter 1, and basal cortisol levels, as measured in salivary samples. Prior literature would suggest that elevated evening cortisol in this adolescent cohort might associate with the unfavourable trajectory class.

**Materials and Methods**

*Study Design*

This study was a re-analysis of the IMPACT-GH trial. IMPACT-GH consisted of a subgroup of patients who, in addition to the data of the IMPACT trial, also provided salivary cortisol data. All patients enrolled into the IMPACT study were asked to take part in IMPACT-GH. Consequently, details of the study design, setting, bias and participant criteria can be found in Chapter 1. In total, 300 patients of the 465 IMPACT sample supplied cortisol data, and were thus recruited into IMPACT-GH.

*Study Size*

Of the 300 patients recruited to IMPACT-GH, 21 patients were excluded due to medications that would alter cortisol mechanisms (19 corticosteroids, and 2 antipsychotics). Thus, 279 patients provided at least one cortisol variable that were available for analysis. These were subdivided into two sub-samples of interest for the present study; one of 112 patients providing data on peak cortisol, and another of 166 patients providing data on evening cortisol for both days (see description below). Preliminary analyses showed that these sub-samples were representative of the full IMPACT sample (see Appendix 4A), and thus patients were allocated to trajectory classes following the analysis of the full IMPACT sample outlined in Chapter 1.

*Cortisol Measurement*

Patients were asked to collect 3 saliva samples on two consecutive school or working days. Participants were instructed to provide the first sample as soon as they woke up, which acted as their “waking” cortisol measure. The second sample was collected 30 minutes after waking to capture the cortisol waking response. The final “evening” sample was provided at 10pm. Participants were asked to note the exact time cortisol measures were provided, in relation to them waking. This allowed us to assess compliance with study protocol, in addition to increasing the accuracy of finding the peak cortisol value of the CAR curve for patients. Patients also noted any medication taken in the previous week. Participants were asked to avoid collecting saliva samples within 20 minutes of eating, smoking or brushing teeth, and to avoid collecting samples within 8 hours of alcohol consumption or the use of recreational drugs. This study was primarily interested in peak-morning and evening cortisol measures.

*Laboratory Measurements*

Saliva was collected by a Salimetrics SOS cotton swab, which was left under the tongue for at least one minute. Salivary cortisol levels were measured by competitive enzyme immunoassay (EIA) using a Salimetrics Europe Ltd kit. Cortisol was measured in micrograms per deciliter. Intra-assay precision was 9.6% at 0.106 µg/dL and 9.8% at 1.058 µg/dL. Inter-assay precision was 3.7% at 0.097 µg/dL and 3.4% at 0.999 µg/dL.

*Inclusion criteria*

For inclusion in analyses, a number of steps were taken to ensure the cortisol data after collection was accurate, in addition to the participant instructions noted above. Firstly, a number of patients had provided cortisol samples following the instructions above, but only for a single day. Thirty-seven patients were found to have provided peak cortisol samples on one of two days only, and 13 provided evening cortisol samples on one of two days only (see Table 1). Basal cortisol measurements are known to show significant within-subject fluctuation such that it is standard practice to derive an average cortisol value for each individual over a number of time-points (Harris et al., 2000; Pruessner et al., 1997). We therefore investigated the relationship between day 1 and day 2 scores, to determine the reliability of a single time point measurement in our data (Appendix 4B). We decided that there was insufficient agreement in cortisol scores between the two days to warrant inclusion of patients with cortisol for only one day. Therefore, for any given cortisol variable, patients must have provided a cortisol sample on both days to be included in our analyses. The means, standard deviations (SD), medians and inter-quartile ranges across day 1 and 2 are presented in Table 1, as well as the two-day average.

*Table 1: Descriptive statistics for cortisol data present on day 1 and day 2*

| | N | Mean | SD | Median | IQR |
|---|---|---|---|---|---|
| ***Peak morning*** | | | | | |
| *Day 1* | 129 | 0.64 | 0.30 | 0.63 | 0.37 |
| *Day 2* | 132 | 0.60 | 0.24 | 0.59 | 0.31 |
| *Two-day average* | 112 | 0.61 | 0.24 | 0.59 | 0.31 |
| ***Evening*** | | | | | |
| *Day 1* | 172 | 0.12 | 0.24 | 0.05 | 0.06 |
| *Day 2* | 173 | 0.11 | 0.22 | 0.06 | 0.06 |
| *Two-day average* | 166 | 0.11 | 0.21 | 0.07 | 0.07 |

In addition, both peak-morning and evening cortisol variables had their own inclusion criteria. In most cases, peak cortisol values were taken as their 30-minute morning sample. This was valid if across both days, saliva was collected between 20 and 40 minutes from their documented waking time, and between the hours of 0400 and 1159. However, in cases where patients provided a second morning sample (between the hours of 0400 and 1159) that was higher than their 30-minute value, this value was recorded as their peak cortisol value. Evening cortisol values were considered valid if across both days, saliva was collected between 1900 and 0100. The average of the two days was taken for both peak-morning and evening cortisol values.

*Statistical Analyses*

Univariate analyses of associations between classes were conducted on each of the two subsamples, to explore differences between the trajectory classes on demographic and clinical characteristics. Independent sample t-tests were performed on all continuous demographic variables. Chi-square tests were performed on all categorical variables. Mann-Whitney-Wilcoxon tests were performed where data were non-normal.

Two logistic regressions were conducted to investigate whether specific cortisol levels predict class membership, controlling for age and gender (see reasons below) and bonferroni corrected. Peak-morning and evening cortisol measures were investigated separately. Regressions were conducted using the *glm()* function in the core *stats* package in R, version 3.3.3(“R: The R Project for Statistical Computing,” n.d.).

*Sensitivity analyses*

A number of considerations were necessary prior to the main analysis detailed above. Firstly, cortisol levels are known to increase with age, particularly through adolescence (Guerry & Hastings, 2011). In addition, failing to control for age has shown an increased likelihood of positive results (Fischer, Macare, et al., 2017). Therefore, to control for type 1 errors, age was included as a fixed covariate in regression models.

Secondly, marked gender differences exist in HPA-activity between males and females (Horstmann & Binder, 2011; Verma et al., 2011). While less consistently reported in adolescents (Bouma, Riese, Ormel, Verhulst, & Oldehinkel, 2009), gender differences have been observed in stress-induced plasma cortisol levels (Uhart, Chong, Oswald, Lin, & Wand, 2006), females have greater sensitivity to ACTH than males (Roelfsema et al., 1993), and greater HPA-axis non-suppression (Heuser et al., 1994). Indeed, we observed a gender difference in our data for peak-morning cortisol at the univariate level (see Table 2). Consequently, we controlled for gender in our regression models. However due to the small number of males present in class 1 for each regression, we also repeated analyses on females only.

*Table 2: Gender differences in cortisol levels*

| **Cortisol measures** | **Females** | **Males** | **Statistics** | |
|---|---|---|---|---|
| | *Median* | *Median* | *W* | *p* |
| *Peak-morning* | 0.65 | 0.49 | 1752 | <.001 |
| *Evening* | 0.07 | 0.06 | 3007 | .096 |

Related to gender, studies have shown that women taking oral contraceptives actually display blunted CAR measured by saliva, in both basal (Pruessner et al., 1997) and stress-induced cortisol responses (Bouma et al., 2009; Kirschbaum, Pirke, & Hellhammer, 1995) than women not taking contraceptives. Consequently, to determine if this was a necessary variable to include in our regression models, we initially investigated the effect of oral contraceptives on each cortisol variable in females, controlling for age and baseline MFQ score. No significant effects of oral contraceptives were found for any cortisol variable (peak-morning: $\beta$=-0.000, $p$=>.999, evening: $\beta$=-0.037, $p$=.612, Bonferroni corrected, details found in Appendix 4C). Therefore, oral contraceptive prescription was not entered as an additional covariate in the final regression model.

Finally, we conducted a similar preliminary analysis for SSRI prescription. As outlined in the introduction, SSRIs have been shown to reduce salivary cortisol levels (Ruhé et al., 2015). Therefore, to determine if this was a necessary variable to include in our regression models, we investigated the effect of baseline SSRIs prescription on each cortisol variable, controlling for age, gender and baseline MFQ score. No significant effects of SSRIs were found for any cortisol variable (peak-morning: $\beta$=-0.023, $p$=>.999, evening: $\beta$=0.050, $p$=.477, Bonferroni corrected, details found in Appendix 4D). Consequently, SSRI prescription was not entered as an additional covariate in the final regression model. Medication data for both oral contraceptives and SSRIs were available for all patients (n=279).

While baseline depression severity is known to correlate with cortisol (Pruessner et al., 2003), baseline MFQ was used to derive the trajectory classes. Therefore, baseline MFQ was not entered as a covariate in the model.

**Results**

*Patient Characteristics*

Of the 279 patients, 112 patients provided peak-morning cortisol samples that met requirements on two days (Table 1). Eighteen (16.0%) were allocated to class 1 and 94 patients were allocated to class 2 (84.0%). Classes contained a similar proportion of patients to the full IMPACT sample (15.9% and 84.1% respectively) and these two samples did not significantly differ in class membership proportions ($X^2(1)< 0.001$, $p$=.958).

Of the 279, 166 patients provided evening cortisol samples that met requirements on two days (Table 1). Twenty-nine (17.5%) were allocated to class 1 and 137 patients were allocated to class 2 (82.5%). While class 1 contained a slightly higher proportion of patients and class 2 a slightly lower proportion of patients than the full IMPACT sample (15.9% and 84.1% respectively), this difference was not significant ($X^2(1)= 0.467$, $p$=.494).

*Class Characteristics*

Table 3 presents the descriptive statistics and univariate tests for each cortisol variable. No significant differences were present between the classes on peak cortisol levels. However, class 1 on average showed significantly higher evening cortisol values than class 2 at baseline.

Tables 4 presents the demographic and clinical data of each trajectory class for the peak-morning cortisol subsample. In contrast to the full IMPACT sample, no demographic or clinical characteristics showed significant class differences, with the exception of region. This is likely a chance finding due to the small sample size in this cohort.

Table 5 presents the demographic and clinical data of each trajectory class for the evening cortisol subsample. For demographic characteristics, and in line with the findings from the full IMPACT sample, the evening cortisol subsample found that the distribution of males and females was significantly different between classes. 93.1% of class 1 were female, whereas only 71.5% of class 2 were female. No other demographic characteristics were significantly different between classes in the evening cortisol subsample. In terms of clinical characteristics, and in line with the findings from the full IMPACT sample, the evening cortisol subsample found that class 1 on average, showed higher LOI and HoNOSCA scores

indicating more obsessional traits and more severe psychiatric symptomatology and function than class 2 at baseline. No other clinical characteristics across cortisol variables were significantly different between classes.

Finally, no differences were found between classes in either cortisol sub-sample, on treatment arm allocation or SSRI prescription at baseline.

*Table 3: Descriptive statistics for each cortisol variable per trajectory class.*

| **Cortisol** | **Class 1: Halted-improvers** | | | | | **Class 2: Continued-improvers** | | | | | **Comparison** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | *Mean* | *SD* | *Median* | *IQR* | *N* | *Mean* | *SD* | *Median* | *IQR* | *W* | *p* |
| *Peak-morning* | 18 | 0.647 | 0.219 | 0.655 | 0.331 | 94 | 0.606 | 0.244 | 0.568 | 0.300 | 751 | .454 |
| *Evening* | 29 | 0.227 | 0.444 | 0.090 | 0.100 | 137 | 0.089 | 0.095 | 0.060 | 0.060 | 1466 | .027 |

*Table 4: Demographic and clinical characteristics of trajectory classes in IMPACT-GH with peak-morning cortisol data*

| | Class 1: Halted-improvers (n=18) | | Class 2: Continued-improvers (n=94) | | Comparison | |
|---|---|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** | ***$X^2$/OR*** | ***p*** |
| *Female*[%] | 16 | 88.9 | 67 | 71.3 | 0.31[%] | .149 |
| *Region*[%] | - | - | - | - | N/A[%] | .017 |
| *East Anglia* | 3 | 16.7 | 49 | 52.1 | - | - |
| *North London* | 7 | 38.9 | 20 | 21.3 | - | - |
| *North-West England* | 8 | 44.4 | 25 | 26.6 | - | - |
| *Ethnicity(white)*[%] | 18 | 100.0 | 85 | 90.0 | N/A[%] | .350 |
| *Suicidal attempts* | 7 | 38.9 | 29 | 30.9 | 0.45 | .504 |
| *Suicidal thoughts*[%] | 17 | 94.4 | 81 | 86.2 | 2.71[%] | .461 |
| *NSSI* | 9 | 50.0 | 57 | 60.6 | 0.71 | .401 |
| *Comorbidity** | - | - | - | - | 1.24 | .266 |
| *1* | 5 | 27.8 | 23 | 24.5 | - | - |
| *2* | 5 | 27.8 | 20 | 21.3 | - | - |
| *3* | 1 | 5.6 | 1 | 1.1 | - | - |
| *Treatment arm*[%] | - | - | - | - | N/A[%] | .553 |
| *BPI* | 7 | 38.9 | 24 | 25.6 | - | - |
| *CBT* | 5 | 27.8 | 35 | 37.2 | - | - |
| *STPP* | 6 | 33.3 | 35 | 37.2 | - | - |
| *Baseline SSRI*[%] | 2 | 11.1 | 18 | 19.1 | 0.53[%] | .521 |
| *Oestrogen medication*[% +] | 1 | 6.3 | 4 | 6.0 | 1.05[%] | >.999 |
| | ***Mean*** | ***S.D*** | ***Mean*** | ***S.D*** | ***t*** | ***p*** |
| *Age* | 15.8 | 1.7 | 15.5 | 1.5 | -0.51 | .617 |
| *RCMAS* | 40.9 | 7.9 | 39.6 | 6.6 | -0.65 | .521 |
| *LOI* | 11.3 | 5.8 | 9.5 | 4.7 | -1.25 | .226 |
| *HONOSCA (available for 17 in class 1 and 90 in class 2)* | 20.2 | 6.1 | 18.0 | 5.7 | -1.36 | .187 |
| | ***Median*** | ***IQR*** | ***Median*** | ***IQR*** | ***W*** | ***p*** |
| *IMD* | 21.0 | 17.6 | 19.4 | 26.1 | 803.5 | .765 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

*+ conducted on females only.*

*[%] Fishers exact test conducted on variables with insufficient cell size for chi-square test.*

**Variable was recoded as binary to meet assumptions of chi-square test.*

*Table 5: Demographic and clinical characteristics of trajectory classes in IMPACT-GH with evening cortisol data*

| | **Class 1: Halted-improvers (n=29)** | | **Class 2: Continued-improvers (n=137)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | *N* | *%* | *N* | *%* | *$X^2$OR* | *p* |
| *Female*[%] | 27 | 93.1 | 98 | 71.5 | 0.19[%] | .014 |
| *Region* | - | - | - | - | 4.97 | .083 |
| *East Anglia* | 10 | 34.5 | 63 | 46.0 | - | - |
| *North London* | 11 | 37.9 | 26 | 19.0 | - | - |
| *North-West England* | 8 | 27.6 | 48 | 35.0 | - | - |
| *Ethnicity(white)*[%] | 27 | 93.1 | 122 | 89.1 | 1.66[%] | .740 |
| *Suicidal attempts* | 14 | 33.6 | 46 | 48.3 | 2.24 | .134 |
| *Suicidal thoughts*[%] | 25 | 86.2 | 119 | 86.7 | 0.95 | >.999 |
| *NSSI* | 20 | 69.0 | 81 | 59.1 | 0.97 | .324 |
| *Comorbidity** | - | - | - | - | 2.21 | .137 |
| *1* | 10 | 34.5 | 43 | 31.4 | - | - |
| *2* | 8 | 27.6 | 25 | 18.2 | - | - |
| *3* | 1 | 3.4 | 1 | 0.7 | - | - |
| *Treatment arm:* | - | - | - | - | 0.68 | .714 |
| *BPI* | 11 | 37.9 | 44 | 32.1 | - | - |
| *CBT* | 11 | 37.9 | 50 | 36.5 | - | - |
| *STPP* | 7 | 24.2 | 43 | 31.4 | - | - |
| *Baseline SSRI*[%] | 2 | 6.9 | 21 | 15.3 | 0.41[%] | .374 |
| *Oestrogen medication*[%+] | 2 | 7.4 | 7 | 7.1 | 1.04[%] | >.999 |
| | *Mean* | *S.D* | *Mean* | *S.D* | *t* | *p* |
| *Age* | 15.9 | 1.3 | 15.5 | 1.4 | -1.53 | .134 |
| *RCMAS* | 40.5 | 7.9 | 39.9 | 7.1 | -0.34 | .737 |
| *LOI* | 11.5 | 5.3 | 9.3 | 4.8 | -2.11 | .042 |
| *HONOSCA (available for 26 in class 1 and 131 in class 2)* | 20.8 | 6.3 | 17.7 | 6.0 | -2.29 | .028 |
| | *Median* | *IQR* | *Median* | *IQR* | *W* | *p* |
| *IMD* | 21.8 | 20.9 | 19.9 | 27.3 | 2024 | .875 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

*+ conducted on females only.*

*[%] Fishers exact test conducted on variables with insufficient cell size for chi-square test.*

**Variable was recoded as binary to meet assumptions of chi-square test.*

*Predictors of trajectory class membership*

Investigation of the correlations between cortisol measures and variables entered into the regression models revealed the strength of collinearity between variables. Results are shown in Table 6 and 7. Variance inflation factor scores however, indicated that multicollinearity for these data was not a concern (all VIF values <10, all tolerance values >.2, See Table 1, Appendix 4E). Further, both models also met the assumption of independent errors (Tables 2, Appendix 4E), and linearity of logit using the Box-Tidwell test, supporting the use of logistic regressions with these data.

*Table 6: Correlation matrix for relationships between variables in peak morning cortisol model*

| | **Peak morning** | **MFQ** | **Age** | **Gender** |
|---|---|---|---|---|
| *Peak morning* | - | 0.03 | 0.01 | -0.36* |
| *MFQ* | - | - | 0.20* | -0.17 |
| *Age* | - | - | - | -0.36* |
| *Gender* | - | - | - | - |

*MFQ; Mood and Feelings Questionnaire*

**p<.05*

*Table 7: Correlation matrix for relationships between variables in evening cortisol model*

| | **Evening** | **MFQ** | **Age** | **Gender** |
|---|---|---|---|---|
| *Evening* | - | 0.02 | 0.14 | -0.10 |
| *MFQ* | - | - | 0.18* | -0.24* |
| *Age* | - | - | - | -0.03 |
| *Gender* | - | - | - | - |

*MFQ; Moode and Feelings Questionnaire*

**p<.05*

Results from each logistic regression are shown in Table 8. Gender emerged as the only significant predictor of class membership when age and cortisol levels were controlled for, in the evening cortisol regression model. However, this did not survive Bonferroni correction for multiple comparisons. Removal of males from the analyses did not alter the significance of other variables in the model (See Appendix 4F).

Standardised residuals of the models were inspected to assess for whether the models contained any outliers of concern. There were no cases in any cortisol sample of residuals

larger than ±2.58, and less than 5% of the total sample for each model showed residuals greater than ±1.96. Therefore, the models can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases are exerting undue influence over the models. It has been recommended to investigate cases where leverage values are greater than twice (Hoaglin & Welsch, 1978) or three times the average (Stevens, 2009). While there were a number of cases in each regression model that suggested a potential problem with 2 or 3 times greater than average leverage, the Cook's distance never exceeded 1 (Cook & Weisberg, 1982), indicating that the fit would not significantly change upon removal of these cases.

*Table 8: Cortisol predictors of trajectory class membership, adjusted for gender and age.*

| | ***OR*** | ***95%CI*** | ***p*** | ***p*** *(bonferroni)* |
|---|---|---|---|---|
| ***Peak morning cortisol*** | | | | |
| *Gender* | 0.318 | 0.046-1.323 | .161 | .483 |
| *Age* | 1.109 | 0.776-1.620 | .576 | 1.000 |
| *Peak morning cortisol* | 1.129 | 0.109-10.712 | .916 | 1.000 |
| ***Evening cortisol*** | | | | |
| *Gender* | 0.206 | 0.032-0.748 | .039* | 0.117 |
| *Age* | 1.202 | 0.871-1.685 | .272 | 0.816 |
| *Evening cortisol* | 15.36 | 1.576-355.075 | .058 | .174 |

**Significant predictors*

## Discussion

### *Key Results and Interpretation*

This piece of work was the first to our knowledge to investigate the relationship between basal cortisol levels and empirically-derived trajectories of symptom change over time, in a cohort of depressed adolescents receiving treatment. Our hypothesis that elevated evening cortisol was predictive of class membership was not supported. While halted-improvers on average showed significantly higher levels of evening cortisol than the continued-improvers at a univariate level, when controlling for gender and age, this effect disappeared. Neither evening nor peak morning cortisol significantly predicted class membership in these models. This is contrary to prior work, which has suggested that variations in cortisol levels are associated with variation in symptom improvement (Fischer, Strawbridge, et al., 2017; Gunlicks-Stoessel et al., 2013; Klimes-Dougan et al., 2018; Robbins et al., 1989). However, it is important to note that the literature on basal cortisol as a predictor of treatment response to psychological therapies is in its infancy. Indeed, the sample size of this study alone exceeded that of all studies of basal measures combined in Fischer and colleague's (2017) meta-analytic work. Furthermore, the number of studies in favour of a significant effect of basal cortisol in this meta-analysis was small and the authors comment on the inconsistencies in the reporting of confounds across studies. It is therefore possible that study specific effects account for the inconclusive literature to date. With these limitations of prior work in mind, our findings may tentatively suggest that variation in basal cortisol may not have a prominent role in differentiating patients' symptom trajectories in adolescent depression.

A strength of the current analysis was that it demonstrated internal validity with regards to the full IMPACT sample. No significant differences were reported between the proportions of patients in trajectory classes between the full IMPACT sample and both IMPACT-GH subsamples. Furthermore, the demographic and clinical characteristics of each these classes within each IMPACT-GH subsample mirrored those of the main sample. Gender, obsessional traits and psychiatry symptomatology and function were the only reported differences at the univariate level in both subsamples, which were also observed in the full IMPACT sample. Consequently, we are confident that the findings here can generalise to the main IMPACT cohort.

We did observe significant gender differences in peak morning cortisol at a univariate level, which would be expected. Females on average exhibited higher cortisol levels at waking, which agrees with the previous literature that found increased stress hormone secretion in females (Goel, Workman, Lee, Innala, & Viau, 2014; Heuser et al., 1994). Conversely, no gender difference was observed in evening cortisol. There are a number of possible reasons for this. Evening cortisol has been favoured in meta-analyses as they are suggested to exhibit smaller inter-individual variability (Fischer, Macare, et al., 2017; Fischer, Strawbridge, et al., 2017). While providing greater reliability, it is possible that the effect size might be consequently smaller in evening measures. In addition, gender differences in cortisol in adolescence are not consistently reported, suggesting that they also may be of smaller magnitude (Bouma et al., 2009). Consequently, our small sample (discussed later) may have precluded us from observing an expectedly smaller effect size for gender differences in evening cortisol in adolescents. Indeed, gender did emerge as the only significant predictor of class membership in the evening cortisol regression models, although this disappeared upon correction. It may be that the interaction between gender, cortisol and an affinity to a particular symptom trajectory over time may be more complex, and not adequately represented in this particular sample. For instance, while the prevalence of depression itself shows a large gender bias towards females (Kuehner, 2003; Piccinelli & Wilkinson, 2000), and females are often associated with elevated cortisol in various measures (Goel et al., 2014), Owens and colleagues (2014) found that only in males did elevated morning cortisol associate with an increased likelihood of being clinically depressed. Studies have also found that different types of stressors elicit differential gender biases in cortisol responses; greater cortisol responses are seen in males to psychological stress, whereas greater cortisol responses are seen in females to pharmacological challenge (Uhart et al., 2006). Taken together, this suggests that the relationship between these three variables is complex, and likely a result of a dynamic interplay between modulatory roles of sex hormones on HPA-axis (Kirschbaum et al., 1996), and key neurotransmitter systems involved in depression (Goel et al., 2014). Our small sample (especially for males) precluded us from investigating this further, however given the gender differences observed in trajectory classes reported in Chapter 2, future work may benefit from specifically investigating the interaction between cortisol variations, gender and class membership.

*Confounds*

There are a number of possible confounds that may have precluded findings of significant associations. Firstly, subtypes within depression have reportedly exhibited differential characteristics of HPA-function. For instance, while symptoms of melancholic depression are associated with hypercortisolaemia, patients with atypical symptoms show hypocortisolaemia (Chrousos, 2009). By extension, it is possible that the lack of positive results may be in part due to the inclusion of patients with comorbidities. Approximately 50% of our sample presented with at least 1 comorbid mental illness meeting clinical threshold for diagnosis. Abnormalities in HPA-axis functioning is a common feature of a number of psychiatric illnesses including schizophrenia (Walker, Mittal, & Tessner, 2008), bipolar disorder (Streit et al., 2016), eating disorders (Luz Neto et al., 2018), alcoholism (Lovallo, Dickensheets, Myers, Thomas, & Nixon, 2000), Post Traumatic Stress Disorder (PTSD; (Meewisse, Reitsma, De Vries, Gersons, & Olff, 2007) and anxiety disorders (Vreeburg et al., 2010). As the direction of cortisol abnormalities differs across psychiatric conditions (Chrousos, 2009), hypotheses for depression samples that include patients with comorbidities are complicated. Comorbid anxiety in particular is an important consideration for our work, given that it is the most commonly reported comorbidity with depression (Avenevoli et al., 2015; Kessler et al., 2003). Indeed this was true for our study; 77% of those patients presenting with comorbidities in IMPACT had a comorbid diagnosis of anxiety. The relationship between HPA-axis abnormalities and anxiety, as well as anxiety-comorbid depression is currently unclear. For instance, one study found depressed patients with comorbid anxiety showed higher CAR than those with depression alone (Vreeburg et al., 2009); an effect that is suggested to be driven specifically by panic disorder and agoraphobia (Vreeburg et al., 2010). However, Van den Bergh and colleagues (2008) found that only trait anxiety, and not depressive symptoms provided a significant contribution to the association between emotional distress and elevated evening cortisol. Conversely, Cameron (2006) stated that elevated HPA-axis responses was only a feature of comorbid depression and anxiety; and not of each condition alone. The dynamic between depression and anxiety and cortisol abnormalities appears to be additionally influenced by pubertal age. Forbes and colleagues (2006) reported that elevated cortisol prior to sleep onset was only associated with anxiety in children, yet for adolescents, elevated cortisol prior to sleep onset was associated with depression. The age range of our sample likely crossed this important

developmental period, and thus developmental changes in our sample may have confounded our results.

The relationship between aberrant HPA-axis function and treatment response for other psychiatric conditions is also unclear (Fischer & Cleare, 2017; Luz Neto et al., 2018). Indeed, the prognostic utility of basal cortisol for treatment outcome for anxiety disorders has not yet gained adequate support (Fischer & Cleare, 2017). Furthermore, Fischer and colleagues (2017) found that only studies that excluded depressed patients with comorbidities reported an association between elevated cortisol after DST challenge and non-response to antidepressants. It is, therefore, possible that the large proportion of patients with comorbidities in our sample, particularly those with anxiety disorders, clouded results indicative of an association between elevated cortisol and an unfavourable trajectory class in depression alone. However, exclusion of such a large proportion of patients would strongly question the generalizability of these results to real cases of patients with depression. While inconclusive, prior work suggests that the consideration of comorbidity is an essential aspect for understanding HPA-axis dysregulation in depression and anxiety (Cameron, 2006); an association that our current work was not primarily designed to investigate.

Secondly, it is important to acknowledge the potential influence of sleep disturbances on the results of the current analysis. Sleep-wake cycles play a major role in the regulation of basal cortisol (Buckley & Schatzberg, 2005). While we ensured that samples were collected in reference to sleep and waking time, documenting and controlling for the *quality* of sleep was not possible. Good quality sleep during adolescence is known to be especially important for mental health and academic achievement (Fuligni, Arruda, Krull, & Gonzales, 2018). In addition, adolescence marks a period where sleep undergoes significant alterations between childhood and adulthood (Forbes et al., 2006). HPA-axis dysfunction has been shown to exacerbate disruptions during sleep (Buckley & Schatzberg, 2005) and in turn, chronic insomnia has been associated with disturbances in nocturnal pulsatile secretions of cortisol, as well as elevated cortisol, particularly during the evening (Rodenbeck et al., 2002). Sleep disturbance is a very common symptom of depression (Mezuk & Kendler, 2012), and indeed, sleep disturbance was reported as the most prevalent symptom in the IMPACT cohort, with 92% of the sample meeting criteria for significant impairment (Goodyer et al., 2017). Consequently, we cannot discount the possibility that the prevalence of sleep disturbance in

our sample may have precluded our observation of significant differences between classes in evening cortisol. A number of studies have supported the assertion that a dynamic and reciprocal relationship exists between cortisol, sleep quality, symptoms of depression and perception of stress in adolescence (Blake, Trinder, & Allen, 2018; Hsiao et al., 2010; Ly, McGrath, & Gouin, 2015). The exact nature and influential power of these relationships is unclear(Blake et al., 2018). Therefore, it is possible that the role of cortisol in treatment outcome may be indirect and act through the more physical symptoms like sleep disturbance observed in depression. The near ubiquitous presence of sleep disturbance in our adolescent cohort suggests that sleep itself may be an important therapeutic target for this demographic. Investigating cortisol abnormalities in this context may yield more fruitful results. Future work will need to focus on elucidating this interactive relationship between sleep, cortisol and treatment outcome more accurately than was possible in our limited sample.

Finally, it is possible that this cohort presented a sample with significantly large inter-individual variability, owing to the role that puberty plays in hormonal development and change at this age. The impact of puberty on our results is discussed in more detail in Chapter 5 as it pertains to the results entire thesis. However, investigating hormonal associations in a sample where individuals will be at differing pubertal stages is likely to introduce a significant amount of noise in our data, which may not have been adequately controlled by chronological age alone.

*Reflections on Methodology*

Our study focused on peak-morning and evening basal measures of cortisol. However, it is possible that the prognostic utility of cortisol as a biomarker for treatment response may not have been adequately represented in this current work. For instance, it may not be the absolute peak-morning cortisol that relates to response, but the size of the awakening response (CAR) that elicits between-class differences. CAR is a derived cortisol measure, and calculated by subtracting a participant's waking cortisol measure, from their peak-morning cortisol measure. The advantage is that it removes inter-individual variability in absolute values, and thus is able to focus solely on the size of response. A number of studies have related abnormal CAR to depression (Huber et al., 2006; Stetler & Miller, 2005). Some studies have also found that patients who respond to treatment experience changes in CAR

(Ruhé et al., 2015). It is therefore possible that our null results are due to an investigation of a biased variable. However, CAR requires both waking and 30 minute-post-waking measurement to be accurate, and is sensitive to deviations in these variables from study protocols. This is obviously difficult for home collection (see below). Nevertheless, we repeated our investigations using CAR (Appendix 4G), and the findings are in agreement with those of the peak-morning cortisol variable. As such, inter-individual variability in absolute values cannot fully explain our null results.

Furthermore, a number of studies have suggested that, rather than baseline levels of cortisol, it is a change in cortisol that is predictive of a favourable trajectory. Zobel and colleagues (2001) found that the change in cortisol and ACTH responses to DST test was predictive of outcome, yet patients did not differ on these variables upon admission. In addition, Ising and colleagues (2007) found that it was the normalisation of cortisol suppression to DST test within the first two weeks of treatment that was associated with increased remission rates at the end of hospitalisation (~15 weeks). A number of studies have also found that sustained elevated cortisol following remission is predictive of relapse (Schüle, 2007; Zobel et al., 2001, 1999). Assessment of whether early basal changes was indicative of increased likelihood membership to a favourable trajectory class was not possible in this current work, as we only had salivary data at baseline. However, this perspective would allow for further hypotheses about the mechanism of treatment action; that for patients who follow a favourable symptom trajectory, treatments work in part through altering their HPA-axis function. Assessment of early changes in cortisol is therefore an important avenue for future work, as it may help clinicians gain an earlier idea of prognosis.

Alternatively, the predictive value of cortisol measures may reside more strongly in other aspects of the HPA-axis functionality. For instance, a number of studies have investigated stress reactivity, either in response to pharmacological or psychological stressors, rather than basal measures. These studies have found significant associations with abnormal stress response and treatment outcome (Gunlicks-Stoessel et al., 2013; Klimes-Dougan et al., 2018; Robbins et al., 1989). A second alternative might hypothesise that cortisol abnormalities between responders and non-responders may not be observed in real-time measures at all, and the assessment of more long-term systemic levels of cortisol exposure might better

elicit predictive value. The assessment of hair samples for such systemic cortisol measures has become increasingly recognised (Russell, Koren, Rieder, & van Uum, 2012). It is not invasive, and as it is not a measure of daily cortisol, it does not rely on collection to be rigidly timed to sleep onset. It is also robust against the high degree of daily variability observed in blood and saliva samples, which are major limitations of real-time measures (Russell et al., 2012). Indeed, a small number of studies have associated abnormalities in hair cortisol concentrations with depression (Dettenborn et al., 2012; Hinkelmann et al., 2013; Staufenbiel, Penninx, Spijker, Elzinga, & Van Rossum, 2012; Steudte-Schmiedgen et al., 2017), however the direction of effect is debated, and effect sizes remain small (Staufenbiel et al., 2012). No study known to the current author has investigated hair cortisol concentrations in the context of a biomarker for response. Consequently, while it is possible that systemic cortisol concentrations may be the variable of interest in this field, more research is required (Fischer, Strawbridge, et al., 2017). We generated our specific hypotheses from the more established literature that supported the investigation of basal cortisol, and it is important to note that hypotheses regarding hair cortisol, while complimentary, are theoretically distinct from those that investigate real-time measures.

We also need to consider the type of cortisol assay that was taken. The IMPACT-GH study chose to collect salivary cortisol samples for a number of reasons. For challenging clinical populations, home salivary sampling for cortisol data is extremely useful; it is non-invasive, and easy to obtain. While salivary cortisol has shown to be a valid and reliable measure of unbound cortisol concentrations, it is important to note that due to specific enzymes present in saliva, the concentrations of cortisol are significantly lower than that found in blood samples, as blood samples measure both bound and unbound cortisol (Kirschbaum & Hellhammer, 1994). Only unbound cortisol can cross the blood-brain barrier and thus salivary cortisol is argued to better reflect the levels of cortisol that are likely to be present in cerebrospinal fluid and affect the brain (Herbert et al., 2006). However, it is possible that salivary measures do not provide the level of sensitivity necessary to detect small variations within a cohort of depressed patients. Blood samples may have provided more promising results, particularly as they have the additional advantage of a more controlled experimental setting at data collection. One must remember though that a disadvantage of blood samples is their invasive nature, which may be stress-inducing itself, and thus bias results of blood measures (Vitiello et al., 1996). Despite questions over sensitivity of salivary samples, it is

unlikely that this explanation fully accounts for our null results. High correlations between salivary and serum measures of cortisol via blood sampling have been repeatedly shown in a number of studies (Baghai et al., 2002; Kirschbaum & Hellhammer, 1994), and only increase when correlated with the corresponding unbound cortisol fraction in blood (Kirschbaum & Hellhammer, 1994).

Home collection was a further limitation of our study design. Home collection minimises inconveniences associated with travel and time commitments of clinic assessments and consequently, these factors often allow for more data to be collected. It also leads to a more valid measure of true cortisol levels in the normal environment, as opposed to the unusual and potentially stressful setting of a laboratory. However, home collection makes compliance rates and time interval between waking and sampling very difficult to verify, which is a common issue for all home-based cortisol collection studies. Various day-to-day factors can all affect cortisol parameters, such as whether the data are collected on working or non-working days (Kunz-Ebrecht, Kirschbaum, Marmot, & Steptoe, 2004), sleep patterns (Rodenbeck et al., 2002) and time of eating (Gibson et al., 1999). We provided patients with detailed instructions to minimise these confounds, and also impressed the importance of accurate recording of times of sampling. Furthermore, extensive data cleaning through our strict inclusion criteria further aided control of compliance through exclusion of patients reporting sampling outside our given time windows.

Finally, it is likely that this current work was underpowered to detect an association between basal cortisol levels and class membership. As previously mentioned, there are a number of factors, such as age, demographic and choice of biological assay (Fischer, Macare, et al., 2017; Fischer, Strawbridge, et al., 2017; Guerry & Hastings, 2011), that would suggest we should expect small effect sizes in our sample, compared to other work. Furthermore, as this was a secondary analysis from a treatment study with a significant assessment burden for participants, valid cortisol collection at all time points may have been a relatively low priority for the research team and participants, leading to a relatively low sample size. In addition, while the empirically-derived classes gained from growth mixture modelling offered a less arbitrary method of defining groups of symptom change over time, the resultant model produced imbalanced class sizes. Consequently, this meant that our smaller non-responder class only contained 18 and 29 patients for peak-morning and evening cortisol respectively

for regression models. This detail precluded our ability to better investigate potential interaction effects of gender, for which our results tentatively suggested may be of interest. The small sample size is arguably the biggest limitation of this work and therefore, the findings from this chapter should be interpreted as exploratory, and predominantly hypothesis-generating work.

In conclusion, this current work suggests that baseline levels of cortisol may not be associated with trajectories of symptom change in depressed adolescents, however it is imperative for these results to be replicated in larger samples before definitive conclusions can be made. It is possible that abnormalities in HPA-axis function may be more related to the development of depression, rather than its persistence following treatment (Vreeburg et al., 2009). Future work should investigate alternative measures of cortisol response, and the interactions between significant variables such as gender, comorbidity and sleep, in order to gain a fuller understanding of the role of cortisol in depression, and its relationship to treatment response in this patient group.

**Appendix 4A**

*Comparisons between IMPACT-GH sub-sample and the full IMPACT sample*

A total of 112 patients enrolled in the IMPACT trial provided valid peak-morning cortisol measures and a total of 166 patients have valid evening cortisol measures. Tables 1 and 2 summarize the demographic and baseline outcome measures of these individuals compared to those not providing cortisol measures.

Overall, there were very few significant differences between the full IMPACT sample and the IMPACT-GH subsample. Exceptions were ethnicity, where a higher proportion of white patients provided cortisol data than other ethnic groups (92.0 to 81.3% for peak-morning cortisol; and 89.8% to 80.6% for evening cortisol respectively). This is likely a chance finding due to imbalanced groupings.

*Table 1: Demographic and clinical characteristics of patients with and without peak cortisol data*

| | **With peak cortisol (n=112)** | | **Without peak cortisol (n=353)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** | ***$X^2$*** | ***p*** |
| *Female* | 83 | 74.1 | 265 | 75.0 | 0.04 | .838 |
| *Region* | - | - | - | - | 2.72 | .256 |
| *East Anglia* | 52 | 46.4 | 133 | 37.7 | - | - |
| *North London* | 27 | 24.1 | 100 | 28.3 | - | - |
| *North-West England* | 33 | 29.5 | 120 | 34.0 | - | - |
| *Ethnicity(white)* | 103 | 92.0 | 287 | 81.3 | 7.14 | .008 |
| *Suicidal attempts* | 36 | 32.1 | 123 | 34.8 | 0.28 | .600 |
| *Suicidal thoughts* | 98 | 87.5 | 316 | 89.5 | 0.35 | .551 |
| *NSSI* | 66 | 58.9 | 205 | 58.1 | 0.03 | .873 |
| *Comorbidity** | - | - | - | - | 0.06 | .808 |
| *1* | 28 | 19.0 | 119 | 33.7 | - | - |
| *2* | 25 | 32.5 | 52 | 14.7 | - | - |
| *3* | 2 | 25.0 | 6 | 1.7 | - | - |
| *4* | 0 | 0.0 | 1 | 0.3 | - | - |
| *Treatment arm* | - | - | - | - | 2.13 | .345 |
| *BPI* | 31 | 27.7 | 124 | 35.1 | - | - |
| *CBT* | 40 | 35.7 | 114 | 32.3 | - | - |
| *STPP* | 41 | 36.6 | 115 | 32.6 | - | - |
| *Baseline SSRI* | 20 | 17.9 | 61 | 17.3 | 0.02 | .888 |
| | ***Mean*** | ***S.D*** | ***Mean*** | ***S.D*** | ***t*** | ***p*** |
| *Age* | 15.6 | 1.5 | 15.6 | 1.4 | 0.27 | .785 |
| *RCMAS* | 39.8 | 6.8 | 41.3 | 7.4 | 1.94 | .053 |
| *LOI* | 9.8 | 4.9 | 10.1 | 5.3 | 0.53 | .596 |
| *HONOSCA ((available for 107 for those with cortisol and 328 for those without)* | 18.4 | 5.8 | 18.6 | 6.2 | 0.32 | .751 |
| | ***Median*** | ***IQR*** | ***Median*** | ***IQR*** | ***W*** | ***p*** |
| *IMD* | 19.4 | 25.6 | 25.3 | 28.0 | 22156 | .054 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

**Variable was recoded as binary to meet assumptions of chi-square test.*

*Table 2: Demographic and clinical characteristics of patients with and without evening cortisol data*

| | **With evening cortisol (n=166)** | | **Without evening cortisol (n=299)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** | ***$X^2$*** | ***p*** |
| *Female* | 125 | 75.3 | 223 | 74.6 | 0.03 | .864 |
| *Region* | - | - | - | - | 3.58 | .167 |
| *East Anglia* | 73 | 44.0 | 112 | 37.5 | - | - |
| *North London* | 37 | 22.3 | 90 | 30.1 | - | - |
| *North-West England* | 56 | 33.7 | 97 | 32.4 | - | - |
| *Ethnicity(white)* | 149 | 89.8 | 241 | 80.6 | 6.62 | .010 |
| *Suicidal attempts* | 60 | 36.1 | 99 | 33.1 | 0.44 | .509 |
| *Suicidal thoughts* | 144 | 86.7 | 270 | 90.3 | 1.04 | .308 |
| *NSSI* | 101 | 60.8 | 170 | 56.9 | 0.70 | .403 |
| *Comorbidity** | - | - | - | - | 0.87 | .351 |
| *1* | 53 | 31.9 | 94 | 31.4 | - | - |
| *2* | 33 | 19.9 | 44 | 14.7 | - | - |
| *3* | 2 | 1.2 | 6 | 2.0 | - | - |
| *4* | 0 | 0 | 1 | <1.0 | - | - |
| *Treatment arm* | - | - | - | - | 1.93 | .380 |
| *BPI* | 55 | 33.1 | 100 | 33.4 | - | - |
| *CBT* | 61 | 36.7 | 93 | 31.1 | - | - |
| *STPP* | 50 | 30.1 | 106 | 35.5 | - | - |
| *Baseline SSRI* | 23 | 13.8 | 58 | 19.4 | 2.28 | .131 |
| | ***Mean*** | ***S.D*** | ***Mean*** | ***S.D*** | ***t*** | ***p*** |
| *Age* | 15.6 | 1.4 | 15.6 | 1.4 | 0.20 | .842 |
| *RCMAS* | 40.0 | 7.2 | 41.4 | 7.3 | 1.98 | .049 |
| *LOI* | 9.7 | 4.9 | 10.2 | 5.4 | 0.97 | .334 |
| *HONOSCA (available for 154 for those with cortisol and 278 for those without)* | 18.3 | 6.2 | 18.7 | 6.0 | 0.70 | .483 |
| | ***Median*** | ***IQR*** | ***Median*** | ***IQR*** | ***W*** | ***p*** |
| *IMD* | 20.5 | 26.7 | 25.3 | 27.4 | 26602 | .199 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

**Variable was recoded as binary to meet assumptions of chi-square test.*

**Appendix 4B**

*The relationship between day 1 and day 2 cortisol measurements*

Due to significant skew present in the cortisol datasets, all cortisol data were log transformed before analyses were conducted. However, even after transformation, the models for both cortisol variables failed the Shapiro-Wilks test for normality. Consequently, permutation tests were conducted in place of linear regressions and results are reported in Table 1 below. Linear regressions with permutation testing were conducted using the *lmp()* function in the core *lmPerm* package in R, version 3.3.3.

*Table 1: Permutation results for day 2 scores on day 1 scores.*

| | $R^2$ | *B* | *Iterations* | *β* | *p* |
|---|---|---|---|---|---|
| *Peak(n=112)* | 0.312 | 0.657 | 5000 | 0.562 | <.001 |
| *Evening(n=166)* | 0.304 | 0.591 | 5000 | 0.551 | <.001 |

These data show that for every increase in one SD of day 1 peak-morning cortisol, day 2 waking cortisol increased by 0.562 SDs. Evening cortisol shows a weaker correspondence between day 1 and day 2 cortisol measures, in that for every increase in one SD of day 1 evening cortisol, day 2 evening cortisol increased by 0.551.

We conclude that these relationships are not sufficient to warrant inclusion of patients with only cortisol data present at one time-point.

**Appendix 4C**

*The effect of oestrogenic medication on cortisol measures.*

Due to significant skew present in the cortisol datasets, both cortisol variables were log transformed before analyses were conducted. However, even after transformation, the models failed the Shapiro-Wilks test for normality. Consequently, permutation tests were conducted in place of linear regressions and results are reported in Table 1 below. Linear regressions with permutation testing were conducted using the *lmp()* function in the core *lmPerm* package in R, version 3.3.3.

*Table 1: Permutation results for oestrogenic medication on cortisol levels, adjusted for age, gender and baseline MFQ.*

| ***Cortisol variable*** | ***$R^2$*** | ***B*** | ***Iterations*** | ***β*** | ***p*** | ***$p_{(bonferroni)}$*** |
|---|---|---|---|---|---|---|
| ***Peak*** | 0.010 | - | - | - | - | - |
| *Age* | - | 0.007 | 51 | 0.044 | .686 | >.999 |
| *MFQ* | - | -0.002 | 122 | -0.010 | .451 | >.999 |
| *Oestrogenic medication* | - | -0.000 | 51 | -0.000 | >.999 | >.999 |
| ***Evening*** | 0.033 | - | - | - | - | - |
| *Age* | - | 0.026 | 907 | 0.150 | .100 | .300 |
| *MFQ* | - | -0.000 | 51 | -0.008 | >.999 | >.999 |
| *Oestrogenic medication* | - | -0.033 | 393 | -0.037 | .204 | .612 |

*MFQ; Mood and Feelings Questionnaire*

**Appendix 4D**

*The effect of baseline SSRI prescription on cortisol measures.*

Due to significant skew present in the cortisol datasets, both cortisol variables were log transformed before analyses were conducted. However, even after transformation, the models failed the Shapiro-Wilks test for normality. Consequently, permutation tests were conducted in place of linear regressions and results are reported in Table 1 below. Linear regressions with permutation testing were conducted using the *lmp()* function in the core *lmPerm* package in R, version 3.3.3.

*Table 1: Permutation results for SSRI prescription on cortisol levels, adjusted for age, gender and baseline MFQ.*

| ***Cortisol variable*** | ***$R^2$*** | ***B*** | ***Iterations*** | ***β*** | ***p*** | ***$p_{(bonferroni)}$*** |
|---|---|---|---|---|---|---|
| ***Peak*** | 0.131 | - | - | - | - | - |
| *Gender* | - | 0.098 | 5000 | 0.180 | <.001* | .003 |
| *Age* | - | 0.002 | 51 | 0.011 | >.999 | >.999 |
| *MFQ* | - | -0.000 | 88 | -0.025 | .534 | >.999 |
| *SSRIs* | - | 0.015 | 51 | 0.023 | .863 | >.999 |
| ***Evening*** | 0.037 | - | - | - | - | - |
| *Gender* | - | 0.022 | 51 | 0.046 | >.999 | >.999 |
| *Age* | - | 0.023 | 1344 | 0.153 | .070 | .210 |
| *MFQ* | - | -0.000 | 51 | -0.008 | >.999 | >.999 |
| *SSRIs* | - | 0.030 | 534 | 0.050 | .159 | .477 |

*MFQ; Mood and Feelings Questionnaire, SSRI; Selective Serotonin Reuptake Inhibitor*

**p<.05*

**Appendix 4E**

*Table 1. Variance inflation factor and tolerance scores for logistic regression models.*

| | VIF | Tolerance |
|---|---|---|
| ***Peak-morning*** | | |
| *Gender* | 1.084 | 0.922 |
| *Age* | 1.000 | 0.999 |
| *Waking cortisol* | 1.084 | 0.922 |
| ***Evening*** | | |
| *Gender* | 1.000 | 0.999 |
| *Age* | 1.001 | 0.999 |
| *Evening cortisol* | 1.001 | 0.999 |

*Table 2. Durbin-Watson test for independent errors for logistic regression models.*

| | DW statistic | *p* |
|---|---|---|
| *Peak-morning* | 1.854 | .390 |
| *Evening* | 2.012 | .954 |

**Appendix 4F**

*Predictors of trajectory class membership for females only*

Investigation of the correlations between cortisol measures and variables entered into the regression models revealed the strength of collinearity between variables. Results are shown in Tables 1 and 2. Variance inflation factor scores however, indicated that multicollinearity for these data was not a concern (all VIF values <10, all tolerance values >.2). Further, all models also met the assumption of independent errors using the Durbin-Watson test, and linearity of logit using the Box-Tidwell test, supporting the use of logistic regressions with these data.

*Table 1: Correlation matrix for relationships between variables in peak-morning cortisol model*

| | **Peak cortisol** | **MFQ** | **Age** |
|---|---|---|---|
| *Peak cortisol* | - | -0.090 | 0.026 |
| *MFQ* | - | - | 0.191 |
| *Age* | - | - | - |

*MFQ; Mood and Feelings Questionnaire*

**p<.05*

*Table 2: Correlation matrix for relationships between variables in evening cortisol model*

| | **Evening cortisol** | **MFQ** | **Age** |
|---|---|---|---|
| *Evening cortisol* | - | 0.011 | 0.159 |
| *MFQ* | - | - | 0.146 |
| *Age* | - | - | - |

*MFQ; Mood and Feelings Questionnaire*

**p<.05*

Results from each logistic regression are shown in Table 3. With the removal of males from the analysis, no variable was shown to significantly predict class membership for either peak or evening cortisol.

Standardised residuals of the models were inspected to assess for whether the models contained any outliers of concern. There were no cases in any cortisol sample of residuals larger than ±2.58, and less than 5% of the total sample for each model showed residuals

greater than ±1.96. Therefore, the models can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases are exerting undue influence over the models. While there were a number of cases in each regression model that suggested a potential problem with 2 or 3 times greater than average leverage, the Cook's distance never exceeded 1 (Cook & Weisberg, 1982), for either cortisol variable. This indicates that the fit would not significantly change upon removal of these cases.

*Table 3: Cortisol predictors of trajectory class membership in females, adjusted for age.*

| | ***OR*** | ***95%CI*** | ***p*** | ***p*** *(bonferroni)* |
|---|---|---|---|---|
| *Peak cortisol (n=75)* | | | | |
| *Age* | 1.131 | 0.760-1.719 | .549 | 1.000 |
| *Peak cortisol* | 1.268 | 0.106-13.777 | .846 | 1.000 |
| *Evening cortisol (n=125)* | | | | |
| *Age* | 1.170 | 0.837-1.656 | .364 | 1.000 |
| *Evening cortisol* | 6.485 | 0.988-115.955 | .118 | .354 |

**Significant predictors*

**Appendix 4G**

*Cortisol Awaking Response (CAR) Analysis*

For patients to be included in the CAR analysis, they must have provided both waking and 30-minute samples that met specific inclusion criteria. Firstly, their waking cortisol value must have been collected less than 10 minutes from their documented waking time and between 0400 and 1159. Secondly, their 30 minute cortisol sample must have been collected between 20 and 40 minutes from their documented waking time, and again, between 0400 and 1159. CAR was then calculated by subtracting their waking value from their 30-minute value.

*The relationship between day 1 and day 2 cortisol measurements*

A number of patients met these criteria for CAR, but only for a single day. Thirty-one patients were found to have only provided CAR on one of two days (Table 1). We therefore investigated the relationship between day 1 and day 2 scores, to determine the reliability of a single time point measurement in our data (See below).

*Table 1: Descriptive statistics for CAR data present on day 1 and day 2*

| | N | Mean | SD | Median | IQR |
|---|---|---|---|---|---|
| ***CAR*** | | | | | |
| *Day 1* | 121 | 0.197 | 0.270 | 0.190 | 0.340 |
| *Day 2* | 120 | 0.225 | 0.275 | 0.200 | 0.313 |
| *Two-day average* | 101 | 0.216 | 0.215 | 0.190 | 0.285 |

*CAR; Cortisol Awakening Response*

Due to significant skew present in the cortisol datasets, and a number of patients recording negative CARs, permutation tests were conducted in place of linear regressions and results are reported in Table 2 below.

*Table 2: Permutation results for day 2 scores on day 1 scores for CAR.*

| | $R^2$ | *B* | *Iterations* | *β* | *p* |
|---|---|---|---|---|---|
| *CAR(n=101)* | 0.121 | 0.37 | 5000 | 0.348 | .001 |

*CAR; Cortisol Awakening Response*

These data show that for every increase in one SD of day 1 CAR, day 2 CAR increased by 0.348 SDs. It was decided that there was insufficient agreement in CAR scores between the two days to warrant inclusion of patients without an average score. Therefore, for the CAR variable, patients must have provided a cortisol sample on both days to be included in the analyses and a two-day average was taken as their CAR value. The means, standard deviations (SD), medians and inter-quartile ranges across day 1 and 2 are presented in Table 1, as well as the two-day average.

*Statistical and Sensitivity Analyses*

Univariate analyses of associations between classes were conducted for the CAR variable, to explore differences between the trajectory classes on demographic and clinical characteristics. A logistic regression was conducted to investigate whether specific cortisol levels predict class membership, controlling for age and gender (see reasons in main paper) and Bonferroni corrected. The methodology for these tests was identical to those of peak-morning and evening variables. Sensitivity analyses of gender, oral contraceptives and SSRIs did not reveal any significance for the CAR variable, and were thus not entered as additional covariates in these regression models.

*Patient characteristics*

Of the 279 patients, 101 patients provided CAR measurements that met requirements on two days (Table 1). Sixteen (15.8%) were allocated to class 1 and 85 patients were allocated to class 2 (84.2%). Classes contained a similar proportion of patients to the full IMPACT sample (15.9% and 84.1% respectively) and these two samples did not significantly differ in class membership proportions ($X^2(1)< 0.001$, $p$=.982) .

*Class Characteristics*

Table 3 presents the descriptive statistics and univariate tests for between class differences in CAR. No significant differences were present between the classes in CAR.

*Table 3: Descriptive Statistics for between class differences in CAR*

| **Cortisol** | **Class 1: Halted-improvers** | | | | | **Class 2: Continued-improvers** | | | | | **Comparison** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | *Mean* | *SD* | *Median* | *IQR* | *N* | *Mean* | *SD* | *Median* | *IQR* | *W* | *p* |
| *CAR* | 16 | 0.19 | 0.21 | 0.21 | 0.29 | 85 | 0.22 | 0.22 | 0.19 | 0.31 | 694 | .900 |

*CAR; Cortisol Awakening Response*

Table 4 presents the demographic and clinical data of each trajectory class in the CAR sub-sample. In contrast to the full IMPACT sample, no demographic or clinical characteristics showed significant class differences, with the exception of region. This finding was likely due to chance, given our small sample size. No differences were found between classes on treatment arm allocation or SSRI prescription at baseline.

*Table 4: Demographic and clinical characteristics of trajectory classes in IMPACT-GH with CAR data*

| | **Class 1: Halted-improvers (n=16)** | | **Class 2: Continued-improvers (n=85)** | | **Comparison** | |
|---|---|---|---|---|---|---|
| | ***N*** | ***%*** | ***N*** | ***%*** | ***$X^2$/OR*** | ***p*** |
| *Female*[%] | 14 | 87.5 | 61 | 71.7 | 0.37[%] | .229 |
| *Region*[%] | - | - | - | - | N/A[%] | .025 |
| *East Anglia* | 3 | 18.8 | 46 | 54.1 | - | - |
| *North London* | 6 | 37.5 | 17 | 20.0 | - | - |
| *North-West England* | 7 | 43.8 | 22 | 25.9 | - | - |
| *Ethnicity(white)*[%] | 16 | 100.0 | 78 | 91.8 | N/A[%] | .593 |
| *Suicidal attempts* | 7 | 43.8 | 28 | 32.9 | 0.70 | .405 |
| *Suicidal thoughts*[%] | 15 | 93.8 | 74 | 87.1 | 2.22[%] | .685 |
| *NSSI* | 8 | 50.0 | 54 | 63.5 | 1.04 | .308 |
| *Comorbidity** | - | - | - | - | 0.17 | .678 |
| *1* | 4 | 25.0 | 23 | 27.1 | - | - |
| *2* | 4 | 25.0 | 19 | 22.3 | - | - |
| *3* | 1 | 6.3 | 1 | 1.2 | - | - |
| *Treatment arm*[%] | - | - | - | - | N/A[%] | .552 |
| *BPI* | 6 | 37.5 | 23 | 27.1 | - | - |
| *CBT* | 4 | 25.0 | 32 | 37.6 | - | - |
| *STPP* | 6 | 37.5 | 30 | 35.3 | - | - |
| *Baseline SSRI*[%] | 2 | 12.5 | 18 | 21.2 | 0.53[%] | .732 |
| *Oestrogen medication*[%+] | 0 | 0 | 4 | 6.6 | 0[%] | >.999 |
| | ***Mean*** | ***S.D*** | ***Mean*** | ***S.D*** | ***t*** | ***p*** |
| *Age* | 15.9 | 1.7 | 15.5 | 1.5 | -0.944 | .357 |
| *RCMAS* | 41.9 | 7.1 | 39.8 | 6.6 | -1.085 | .291 |
| *LOI* | 11.3 | 5.7 | 9.4 | 4.8 | -1.274 | .218 |
| *HONOSCA (available for 15 in class 1 and 81 in class 2)* | 20.8 | 6.1 | 18.1 | 5.8 | -1.581 | .130 |
| | ***Median*** | ***IQR*** | ***Median*** | ***IQR*** | ***W*** | ***p*** |
| *IMD* | 19.4 | 16.1 | 18.4 | 23.6 | 678.5 | .993 |

*IMD; Index of Multiple Deprivation, RCMAS; Revised Children's Manifest Anxiety Scale, LOI; Leyton Obsessional Inventory, NSSI; Non-suicidal self-injury, HoNOSCA; Health of the Nation Outcome Scales for Children and Adolescents, BPI; Brief Psychological Intervention, CBT; Cognitive Behavioural Therapy, STPP; Short-Term Psychoanalytic Psychotherapy, SSRI; Selective Serotonin Reuptake Inhibitors*

*+ conducted on females only.*

*[%] Fishers exact test conducted on variables with insufficient cell size for chi-square test.*

**Variable was recoded as binary to meet assumptions of chi-square test.*

*Predictors of trajectory class membership*

Investigation of the correlations between cortisol measures and variables entered into the regression models revealed the strength of collinearity between variables. Results are shown in Table 5. Variance inflation factor scores however, indicated that multicollinearity for these data was not a concern (all VIF values <10, all tolerance values >.2). Further, both models also met the assumption of independent errors, and linearity of logit using the Box-Tidwell test, supporting the use of logistic regressions with these data.

*Table 5: Correlation matrix for relationships between variables in CAR model*

| | **CAR** | **MFQ** | **Age** | **Gender** |
|---|---|---|---|---|
| *CAR* | - | -0.12 | 0.05 | -0.08 |
| *MFQ* | - | - | 0.21* | -0.18 |
| *Age* | - | - | - | -0.23* |
| *Gender* | - | - | - | - |

*MFQ; Mood and Feelings Questionnaire*

**p<.05*

Results from each logistic regression are shown in Table 6. After controlling for gender and age, CAR was not a significant predictor of class membership. Removal of males from the analyses did not alter the significance of other variables in the model.

Standardised residuals of the models were inspected to assess for whether the models contained any outliers of concern. There were no cases of residuals larger than ±2.58, and less than 5% of the total sample for the model showed residuals greater than ±1.96. Therefore, the model can be viewed as a good representation of the actual data.

Leverage was investigated to assess whether any cases are exerting undue influence over the models. While there were a number of cases in the regression model that suggested a potential problem with 2 or 3 times greater than average leverage, the Cook's distance never exceeded 1 (Cook & Weisberg, 1982), indicating that the fit would not significantly change upon removal of these cases.

*Table 6: CAR as a predictor of trajectory class membership, adjusted for gender and age.*

| | ***OR*** | ***95%CI*** | ***p*** | $p_{(bonferroni)}$ |
|---|---|---|---|---|
| *CAR* | | | | |
| *Gender* | 0.328 | 0.048-1.340 | .168 | .504 |
| *Age* | 1.218 | 0.835-1.839 | .320 | .504 |
| *CAR* | 0.286 | 0.017-3.819 | .357 | .960 |

*CAR; Cortisol Awakening Response*

## Chapter 5: Depression and outcome following treatment: what have we learned?

*Summary of main findings*

This thesis has focused on individual differences in depressive symptom change in adolescents and predictors of such outcomes within a group of patients with DSM-IV clinical diagnosis of unipolar major depression. The findings of my first chapter highlighted 3 important and clinically relevant aspects to patients' symptom change over time:

1. Depressed adolescents do not all follow the same trajectory of symptom change over time during and after psychological treatment. Distinct subgroups exist which show dissimilar patterns of change. These categories differ to those generated by a priori approaches (specifically, percentage reduction of symptoms greater than an arbitrary threshold). Researchers should be aware that adopting cut-offs based on a priori thresholds could differentially and potentially erroneously categorise responsive patients as unresponsive, thus under-estimating treatment effects.
2. Symptom change over time is not a linear function and therefore researchers should be cautious of adopting definitions that assume such (linear) functions when assessing clinical response. A thorough investigation of trajectory shape in future work can help to build a more valid picture for clinicians of the likely journey to recovery for treatment-responsive patients. This could have beneficial effects on treatment adherence, through the management of expectations of patients, their families and, perhaps, clinicians themselves, which is one of the main challenges in this population of depressed patients (Stulz et al., 2010).
3. Through comparison with previous studies in adolescent depression (Scott et al., 2019), this work showed that long-term follow up is essential to reduce the risk of erroneous categorisation of patients with unfavourable long-term responses as rapid responders.

The work of chapters 2-4 then allowed us to investigate characteristics of the empirically derived classes, with the knowledge that these classes were defined in a less arbitrary way, and thus may elucidate more conclusive findings across predictors of interest. As expected, patients categorised to the unfavourable trajectory presented with baseline characteristics indicative of greater clinical complexity and severity than those in the favourable trajectory

class, and this was also evident in the investigations of the two subsamples defined by imaging and cortisol parameters respectively.

This work then extensively investigated a variety of predictors of class membership including clinical, neurological and physiological variables. The biological variables have not previously been investigated in empirical work. Interestingly, only comorbidity was found to independently predict class membership. The clinical model described in Chapter 2 only explained a small proportion of the variance associated with these trajectory classes, which warranted investigation of biological variables. However, the null results of the two further sub-studies suggested that perhaps those variables (cortical thickness, cortical surface area, and salivary cortisol), do not serve as strong predictors of symptom trajectory membership. These are tentative conclusions however as the investigations were limited by the small sample sizes of those sub-studies.

Taken together, the results presented in this thesis have provided novel contributions to understanding the trajectory of symptom change in adolescent depression, and the importance of trial design and analytical choice in influencing outcome data. Given that behavioural symptoms appear insufficient in predicting such trajectories, and our preliminary work suggests that some selected biological variables may not serve significant roles in these questions, this work has also opened lines of further inquiry into why these groups are responding differently, and what could distinguish them at outset.

*Reflections on null results*

As stated above, the first interpretation of the null results presented in this thesis is that they suggest that the variables investigated here do not serve an influential role in prognosis. However, I'd like to refer back to a topic I discussed in the introduction of Chapter 1: the issue of data reduction with categorical approaches (Altman & Royston, 2006; Cohen, 1983; Royston et al., 2006). Categorical approaches inevitably lose information when data are reduced to a small number of groups, or in this case, a binary outcome (Altman & Royston, 2006; Cohen, 1983; Royston et al., 2006). As mentioned in Chapter 1, it is theorised that up to one third of information about pattern variation is lost when dichotimising continuous outcomes in this way (Cohen, 1983). Consequently, our lack of positive results may be due to this choice of methodology reducing our power to detect significant effects.

Indeed, a number of studies have noted that their analyses were more successful when response is defined on a continuous scale (Emslie et al., 2002; Koenig et al., 2018), suggesting that differences within a depressed cohort are subtle, and thus analytical methods should consider these in their study design appropriately.

Moreover, many clinicians will advocate that categorical methods do not reflect clinical observations accurately because response to treatment is fundamentally continuous. Patients fall on a continuum with no clear characteristic that can separate individuals categorically to being well or unwell; it is a matter of degree (Malhi & Byrow, 2016). Clinical improvement turns into remission, which turns into recovery over time, the length of which differs per individual (Malhi & Byrow, 2016). This is the reason the majority of scales used to quantify depression are continuous scales of severity (Angold, et al 1961; Hamilton, 1960; Montgomery & Asberg, 1979). Moreover, there is a growing field of work advocating that psychopathology more generally (beyond only depression symptoms) needs to be interpreted in dimensional models rather than categorical boundaries (Kotov et al., 2011, 2017). For example, it has been shown that depressive symptomatology is common in many non-depressive disorders (Braun, Sunday, & Halmi, 1994; Buckley, Miller, Lehrer, & Castle, 2009; Garber & Weersing, 2010) and the presence of comorbidities across conditions is more common than their absence (Goodyer et al., 1997; Rohde et al., 1991). The findings of this thesis resonate with these statements as we found that over half our population presented with at least one comorbidity and this prevalence does not even consider patients who reached diagnostic threshold for some symptoms of other conditions, but did not reach full diagnosis of that condition. Furthermore, it was only comorbidity in this work that presented as an independent predictor of an unfavourable trajectory class. Consequently, it is possible that had our GMMs incorporated non-depressive symptoms at outset, patient groupings may have emerged that resulted in clearer links with our biological predictors and also provided an interesting perspective on questions of diagnoses. While this was not the focus of my research questions, it would be an interesting avenue for future research. The fact that the clinical profile of patients appears to predict their ability to respond to depression treatment highlights the importance of such trans-diagnostic research in understanding treatment response. What is clear is that the way in which psychopathological symptoms are grouped in diagnostic criteria is of critical importance in

research. Currently, it has a marked effect on the modality of treatment they receive, but it may also affect the success of that treatment.

While the data-related and conceptual issues around categorical approaches presented above suggest a strong argument that continuous approaches should be favoured in this field, there are two core reasons why categorical methods were chosen for this work. Firstly, this work aimed to investigate factors associated with those groups of patients experiencing unfavourable trajectories of change over time. This type of question is addressed most optimally with categorical approaches. Secondly, part of rationale behind this work was to more directly inform clinical decision-making, which, as stated in Chapter 1, is fundamentally categorical (Uher et al., 2010). While continuous outcomes can provide more detailed investigations of the subtleties affecting response, they are limited in their ability to inform clinical practice in this context. Moreover, if the effects of a variable are so subtle that they are undetectable in categorical approaches, the question remains: how influential can these variables truly be in affecting treatment outcome? Consequently, I believe that choosing a categorical approach, but one that takes a less arbitrary method to defining its groupings, was the optimal decision to address the research questions in a way that is clinically meaningful.

A second consideration regarding the null results in this work relates to a broader issue within psychopathological research that I touched upon in Chapter 3. That is, the extent to which differences in symptomatology can adequately relate to differences in underlying biological mechanisms (Insel et al., 2010). This is a poignant topic in the current literature. With the introduction of the Research Domain Criteria (RDoC) (Insel et al., 2010), there is a drive for neuroscientific and genetic research to be incorporated appropriately into classification schemes. This has emerged from the growing body of work in genetics and neuroscience that has highlighted the existence of biological associations with psychopathology, but a simultaneous failure of these associations to align with diagnostic criteria (Insel et al., 2010). An interesting consequence of this movement has been the investigation of “biotypes”: whether patients defined by biology provide a more coherent categorisation method of psychopathology (Clementz et al., 2016; Drysdale et al., 2017). The first study by Clementz and colleagues(2016) investigated biological profiles of psychosis; recruiting patients with schizophrenia, schizoaffective disorder and bipolar psychosis. The

authors made two important findings relevant to this current thesis. Firstly, they found that defining patients based on neuropsychological parameters (such as cognitive tasks and electrophysiological function) resulted in patients falling in 3 distinct biotypes (Clementz et al., 2016); however, the same analysis conducted on diagnostic criteria produced a single continuum of severity (ie- one group of patients, best defined dimensionally). This suggests that while symptoms may be better represented on a dimensional scale (and consequently present difficulties for the binary nature of clinical decision-making (Uher et al., 2010)), biological variables may be more discrete. It is possible that our reliance on clinical symptoms for diagnoses may be obscuring the distinction between neurobiologically discrete conditions with similar patterns of clinical symptoms: this would index aetiological as well as clinical heterogeneity. Similar findings have been reported in a study of depression, whereby differences in resting state fMRI produced 4 distinct biotypes of patients (Drysdale et al., 2017). These authors reported high sensitivity and specificity of the biotypes to classify patients correctly. Moreover, both of these studies showed that these biotypes corresponded poorly with diagnoses (Clementz et al., 2016), and could not be easily distinguished based on clinical presentation alone (Drysdale et al., 2017).

The second important finding of Clementz and colleagues (2016) was that biotypes displayed significantly different grey matter volume reductions in widespread areas of the brain, including cingulate and frontal regions. These findings were also replicated with grey matter density in this cohort (Ivleva et al., 2017), providing a strong validation of the proposed biotypes. Critically, grey matter volume was found to predict biotype class better than diagnostic class (Clementz et al., 2016). This is particularly relevant to this thesis as both chapters 3 and 4 highlighted that the abnormalities seen in neuroimaging data and cortisol variables are not unique to depression, but implicated in a great variety of other diagnoses, from anxiety to schizophrenia (Girshkin et al., 2016; Goodkind et al., 2015; Luz Neto et al., 2018). Perhaps, therefore, it is not that structural abnormalities in the brain are only indicative of general psychopathological illness, but that our categorisation approach is not optimally associating with these differences.

The work on biotypes suggests that differences in symptomatology do not easily propagate to observable differences in biology. This is especially likely in depression, a condition showing substantial heterogeneity. It is likely that the underlying biology would mimic this

heterogeneity. These interpretations suggest that a potential alternative explanation for why this current work produced the null results is that the relationship between depressive symptoms and biology is more complex, perhaps with currently unidentified moderators and mediators acting upon the causal relationship (Kazdin & Nock, 2003). Indeed, Drysdale and colleagues (2017) actually noted that core symptoms of depression (such as low mood and anhedonia) related to a core set of neurological indicators that were present in all patients, regardless of biotype. These included some of the key regions of interest investigated in this thesis, like the insula, OFC and ventral mPFC (Drysdale et al., 2017). Consequently, categorising patients based on a total symptom severity scale over time, as conducted in this current thesis, may not have been the criteria under which patients reflect differentiation in neurobiology, particularly in the regions investigated here. The null results may have been a consequence of the way in which the outcome groups were defined, rather than suggesting that these variables are insignificant in prognosis. Unfortunately, IMPACT was not designed as a biological investigation. Very few biological variables existed in this dataset and the sample size of those with biological data was too small to conduct similar biotype analyses. Future biological work could incorporate symptoms beyond depression in their models, to fully elucidate biological differences between patients that might associate with response.

*Limitations to interpretations*

One of the biggest limitations of this work has been our sample size. While the full IMPACT sample of 465 was sufficient for GMM (Nylund et al., 2007), it was modest, and as previously stated in Chapter 1 it is typical for GMM studies to possess a sample of 600 or more (Brière et al., 2016; Gueorguieva et al., 2011; Rhebergen et al., 2012; Thibodeau et al., 2015; Uher et al., 2010). Sample size was also a particular concern for the biological variables. I discuss these limitations more specifically in those chapters (3 and 4), however one of the biggest implications of this (particularly the low sample size of the biological samples) was that the studies may not have been powered enough to investigate more complex relationships between biological variables and clinical characteristics. As biotypes research has highlighted above (Clementz et al., 2016; Drysdale et al., 2017; Ivleva et al., 2017), the relationship between biology and symptomatology is not clear, and is not simple. Consequently, a larger sample with biological data would have allowed a more detailed investigation of the interactive relationships between biological and clinical profiles, but also between the

biological variables themselves. Investigating such questions would be an ideal extension of this current work.

A second difficulty in the interpretation of this work is the developmental age of our study participants. As stated in Chapter 3, adolescence is a highly dynamic period of rapid cortical development. Large multi-sample longitudinal studies have illustrated that widespread, nonlinear changes in cortical thickness and surface area occur across the age range of our cohort (Tamnes et al., 2017; Vijayakumar et al., 2016). These changes also vary in degree across regions of the brain, with regions investigated in this work showing greatest areas of change (Tamnes et al., 2017). Similarly, as stated in Chapter 4, HPA-reactivity sharply increases during adolescence, along with rises in basal cortisol (Gunnar & Quevedo, 2007; Gunnar, Wewerka, Frenn, Long, & Griggs, 2009). Mapping trajectories of symptom change, on top of trajectories of development is therefore a complicated task.

Compounding the above complication is the likelihood that a large degree of inter-individual variability exists within these cortical and hormonal changes, owing to differences in pubertal status. Pubertal status has been associated with changes in brain development beyond chronological age, and these changes may be sex-specific (Wierenga et al., 2018). Similarly, variability in cortisol levels is found to correlate with pubertal status (Kiess et al., 1995). Consequently, while we controlled for chronological age in both the neurological and cortisol analyses, additional variability may be present in our dataset relating to pubertal age that is not completely explained by chronological age. We were unable to control for pubertal status, as it was not part of the study protocol thus developmental status across the study participants remains an alternative explanation for null results, and a limitation of this work.

*Outstanding questions*

This work provided a thorough and arguably more meaningful investigation of predictors of symptom change over time. However, there are two important questions in this field that are not answered in this work. The first being the question of which treatment is best suited for which patient. The distinction between this question and that of this current work resides in the difference in the definition of predictors and moderators, which are often confused, owing to their simultaneous reporting in many papers of depression treatment

(Curry et al., 2006; Kraemer, Wilson, Fairburn, & Agras, 2002; Phillips et al., 2015). Both predictors and moderators are present before treatment initiation, however the difference is that predictors associate with general response regardless of treatment modality, whereas moderators investigate differential treatment response (Curry et al., 2006; Kraemer et al., 2002). That means that they investigate interactive effects of the variable with treatment modality. In terms of adolescent depression, the field of moderators of response is in its infancy (Weersing et al., 2017). There are many mixed results (Weersing et al., 2017), with most analyses of a secondary nature and therefore lacking optimal study design (Curry et al., 2006). However, a number of moderators to response have been suggested. For instance, the TADS trial found that severe cases of depression showed no additional benefit of CBT to fluoxetine, while mild and moderate cases benefitted from this combination treatment (Curry et al., 2006). This study also found that patients reporting higher cognitive distortions benefitted from the addition of CBT to fluoxetine treatment, whereas those with fewer distortions showed no benefit of the addition of CBT. A recent review of adolescent depression also found a positive effect of CBT in cases of comorbid anxiety, but a negative effect for patients who have significant life stress or previous trauma (Weersing et al., 2017). As such, it is possible that the variables investigated in this thesis may have played a more prominent role in identifying moderators of treatment, rather than prediction of general symptomatic change.

It is argued that investigations of moderators are more valuable to clinicians, as they provide an indicator for prescription (Curry et al., 2006; Phillips et al., 2015). Moreover, they could provide indicators of potential therapeutic weaknesses, which may aid questions of mechanism (see below) (Weersing et al., 2017). However, there is a great difficulty involved in detecting a moderation signal, with authors stating that the noise in standard clinical trials is too large to detect differences (Frank et al., 2011). Consequently, most variables emerge as non-specific predictors (Frank et al., 2011). Indeed, neuroimaging reviews have stated that there is an overriding focus on predictors in this field, and place the blame on the lack of sufficient sample sizes for moderator analyses (Phillips et al., 2015). Therefore, for this work it was more valuable to investigate questions of prediction, rather than attempt questions of differential response in sample sizes too small to give sufficient power. Indeed, in the TADS study, certain variables had to be collapsed to maintain a sample size adequate for moderator analyses (Curry et al., 2006). However, a more thorough investigation of

moderators would indeed provide clinicians with useful prescriptive information, provided the study was adequately powered. Similarly, we did not have sufficient power to investigate interactive effects between variables of interest and class membership (such interactions between gender, cortisol and class membership as mentioned in Chapter 4), which may have produced more insight into questions of prognosis.

The second question still outstanding for this field is the question of *how* treatment causes the symptom change patterns we have observed in this work. This is a question of mechanism of action. As Weersing and colleagues (2017) stated in their recent review of psychological treatments in adolescent depression, there is currently a lack of sufficient understanding of the mechanism behind treatment action for either psychological (Weersing et al., 2017) or, indeed, pharmacological treatments (Moncrieff, 2018). For instance, Moncrieff (2018) highlighted that while it is still commonplace to believe that antidepressants work by balancing chemical abnormalities in depressed individuals, the evidence for this assertion is not conclusive. In terms of psychological therapies, little research has been conducted on mediators of CBT (Kazdin & Nock, 2003; Weersing et al., 2017). Furthermore, although early work suggests that the behavioural component of CBT contributes most to its effectiveness (Jacobson et al., 1996), the active components of other psychological therapies remain to be adequately researched (Weersing et al., 2017; Weersing, Rozenman, & Gonzalez, 2009). This is true for the other psychological treatments used in IMPACT (BPI and STPP). Weersing and colleagues (2017) concluded by emphasising the effect this has on treatment development: *"If the core processes of an intervention model are unknown, there is little scientific basis to guide improvements to the treatment when poor effects are observed in practice settings*" (p.37) (Weersing et al., 2017). IMPACT was not designed to answer questions of mechanism, and contained no placebo group. However, the results presented in this thesis provide new information on the shape of trajectories of symptom change, and suggest that alterations might occur near treatment cessation. Consequently, this understanding of the trajectory of change during treatment contributes information that could be of value to understanding change mechanisms. Nevertheless, further focused mechanistic research is critical in advancing our treatment of adolescent depression (Weersing et al., 2017).

The practical issue of both the above questions is that they are currently hindered by a requirement for large samples and longitudinal data, which is costly, and typically results in underpowered research (Bhatt & Mehta, 2016). Consequently, the call for the adoption of adaptive designs in clinical trials is growing support (Bhatt & Mehta, 2016; Goodyer & Wilkinson, 2018), and would help address both questions outlined above. While these designs necessitate the adoption of more complex planning and statistical analyses, they allow for a flexible approach to treatment allocation (Bhatt & Mehta, 2016). Interim analyses can be used to answer the question of what treatment is working best for whom, and simultaneously optimise treatment allocation (Weisz et al., 2012). This helps increase power while shortening the necessary study duration (Bhatt & Mehta, 2016). Incorporating neurobiological tests into these designs could provide additional insight into how these treatments are exerting their effects.

*Conclusions and directions for future work*

Overall, this work has highlighted the importance of considering the temporal aspects of symptom change in the categorisation of patients to response classes. Without this information, we would not know that for a minority of patients, the active treatment period appears an important component for continued improvement, and thus may inform practice of extending treatment for such cases. It has also tentatively proposed a re-examination of biological predictors that have previously shown promise in relating to outcome, as this alternative categorical approach suggests these effects may be over-estimated or the hypotheses upon which they are based are underdeveloped.

It is imperative for future work to focus first on replicating the results of the GMM described here in another sample of patients with adolescent depression, as described in Chapter 1. With replication, we can begin to extend these findings to investigate what factors may alter a persons' trajectory path, which treatment is most optimal for which patient and indeed, the mechanisms by which treatment is acting to elucidate these trajectory paths.

More generally, while clinical research should continue to develop methods that aid our current diagnostic systems, exploratory work may benefit from less reliance on these criteria in study design and search for alternative approaches to diagnosis. This may elucidate clinical profiles in a more coherent manner to inform prescription of treatment modalities.

The development of treatments based on biologically validated classification methods may prove to better address cases of treatment resistance in the future, and finally advance the currently stagnated field of treatment efficacy (Insel et al., 2010).

**References**


Adam, E. K., Doane, L. D., Zinbarg, R. E., Mineka, S., Craske, M. G., & Griffith, J. W. (2010). Prospective prediction of major depressive disorder from cortisol awakening responses in adolescence. *Psychoneuroendocrinology*, *35*(6), 921–931. http://doi.org/10.1016/j.psyneuen.2009.12.007

Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., … Vollebergh, W. A. M. (2004). Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, *109*(s420), 21–27. http://doi.org/10.1111/j.1600-0047.2004.00327.x

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ (Clinical Research Ed.)*, *332*(7549), 1080. http://doi.org/10.1136/bmj.332.7549.1080

American Diabetes Association. (2003). Standards of medical care for patients with diabetes mellitus. *Diabetes Care*, *26 Suppl 1*(suppl 1), S33-50. http://doi.org/10.2337/DIACARE.26.2007.S33

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). American Psychiatric Association Press. Washington DC.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Publishing Arlington, VA. http://doi.org/10.1176/appi.books.9780890425596

Anacker, C., Zunszain, P. A., Cattaneo, A., Carvalho, L. A., Garabedian, M. J., Thuret, S., … Pariante, C. M. (2011). Antidepressants increase human hippocampal neurogenesis by activating the glucocorticoid receptor. *Molecular Psychiatry*, *16*(7), 738–750. http://doi.org/10.1038/mp.2011.26

Andrews, G., Issakidis, C., Sanderson, K., Corry, J., & Lapsley, H. (2004). Utilising survey data to inform public policy: Comparison of the cost-effectiveness of treatment of ten mental disorders. *British Journal of Psychiatry*, *184*(6), 526–533. http://doi.org/10.1192/bjp.184.6.526

Angold A, Costello E, Pickles A, W. F. (n.d.). The development of a questionnaire for use in epidemiological studies of depression in children and adolescents. *London: Medical Research Council Child Psychiatry Unit.*

Arnone, D., McIntosh, A. M., Ebmeier, K. P., Munafò, M. R., & Anderson, I. M. (2012). Magnetic resonance imaging studies in unipolar depression: systematic review and

meta-regression analyses. *European Neuropsychopharmacology*, *22*(1), 1–16. http://doi.org/10.1016/j.euroneuro.2011.05.003

Asarnow, J. R., Porta, G., Spirito, A., Emslie, G., Clarke, G., Wagner, K. D., … Brent, D. A. (2011). Suicide Attempts and Nonsuicidal Self-Injury in the Treatment of Resistant Depression in Adolescents: Findings from the TORDIA Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(8), 772–781. http://doi.org/10.1016/J.JAAC.2011.04.003

Augustine, J. R. (1996). Circuitry and fimctional aspects of the insular lobe in primates including humans. *Brain Research Reviews*, *22*, 229–244. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.587.386&rep=rep1&type=pdf

Austin, M., Whitehead, R., Edgar, C., & JE Janosky. (2002). Localized decrease in serotonin transporter-immunoreactive axons in the prefrontal cortex of depressed subjects committing suicide. *Neuroscience*, *114*(3), 807–815. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306452202002890

Avenevoli, S., Knight, E., Kessler, R. C., & Merikangas, K. R. (2008). *Epidemiology of depression in children and adolescents*. *Handbook of depression in children and adolescents*. New York: Guilford Press.

Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M., & Merikangas, K. R. (2015). Major Depression in the National Comorbidity Survey–Adolescent Supplement: Prevalence, Correlates, and Treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 37–44.e2. http://doi.org/10.1016/J.JAAC.2014.10.010

Bagby, R. M., Ryder, A. G., & Cristi, C. (2002). Psychosocial and clinical predictors of response to pharmacotherapy for depression. *Journal of Psychiatry & Neuroscience*, *27*(4), 250–257. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12174734

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? *American Journal of Psychiatry*, *161*(12), 2163–2177. http://doi.org/10.1176/appi.ajp.161.12.2163

Baghai, T. C., Schüle, C., Zwanzger, P., Minov, C., Holme, C., Padberg, F., … Rupprecht, R. (2002). Evaluation of a salivary based combined dexamethasone/CRH test in patients with major depression. *Psychoneuroendocrinology*, *27*, 385–399.

Bamber, D., Tamplin, A., Park, R. J., Kyte, Z. A., & Goodyer, I. M. (2002). Development of a short leyton obsessional inventory for children and adolescents. *Journal of the*

*American Academy of Child and Adolescent Psychiatry*, *41*(10), 1246–1252. http://doi.org/10.1097/00004583-200210000-00015

Barden, N., Reul, J. M. H. M., & Holsboer, F. (1995). Do antidepressants stabilize mood through actions on the hypothalamic-pituitary-adrenocortical system? *Trends in Neurosciences*, *18*(1), 6–11. http://doi.org/10.1016/0166-2236(95)93942-Q

Bartels, M., de Geus, E. J. C., Kirschbaum, C., Sluyter, F., & Boomsma, D. I. (2003). Heritability of Daytime Cortisol Levels in Children. *Behavior Genetics*, *33*(4), 421–433. http://doi.org/10.1023/A:1025321609994

Bauer, D. (2007). Observations on the Use of Growth Mixture Models in Psychological Research. *Multivariate Behavioural Research*, *42*(4), 757–786.

Bauer, D., & Curran, P. (2003). Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological Methods*, *8*(3), 338–363. Retrieved from http://psycnet.apa.org/buy/2003-09632-007

Beck, S., Choi, K., List, T., Okuhara, D., & Birnstiel, S. (1996). Corticosterone alters 5-HT receptor-mediated hyperpolarization in area CA1 hippocampal pyramidal neurons. *Neuropsychopharmacology*, *14*(1), 27–33. http://doi.org/10.1016/S0893-133X(96)80056-X

Beck AT, Ward CH, Mendelson M, Mock J, E. J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Belzung, C., Willner, P., & Philippot, P. (2015). Depression: from psychopathology to pathophysiology. *Current Opinion in Neurobiology*, *30*, 24–30. http://doi.org/10.1016/J.CONB.2014.08.013

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Statistical Methodology*, *57*(1), 289–300. http://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berlim, M. T., & Turecki, G. (2007). What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials. *European Neuropsychopharmacology*, *17*(11), 696–707. http://doi.org/10.1016/j.euroneuro.2007.03.009

Bhatt, D. L., & Mehta, C. (2016). Adaptive Designs for Clinical Trials. *New England Journal of Medicine*, *375*(1), 65–74. http://doi.org/10.1056/NEJMra1510061

Binder, E. B., Künzel, H. E., Nickel, T., Kern, N., Pfennig, A., Majer, M., … Holsboer, F. (2009).

HPA-axis regulation at in-patient admission is associated with antidepressant therapy outcome in male but not in female depressed patients. *Psychoneuroendocrinology*, *34*(1), 99–109. http://doi.org/10.1016/J.PSYNEUEN.2008.08.018

Birmaher, B., Brent, D. A., Kolko, D., Baugher, M., Bridge, J., Holder, D., … Ulloa, R. E. (2000). Clinical Outcome After Short-term Psychotherapy for Adolescents With Major Depressive Disorder. *Archives of General Psychiatry*, *57*(1), 29–36. http://doi.org/10.1001/archpsyc.57.1.29

Birmaher, B., Dahl, R. E., Perel, J., Williamson, D. E., Nelson, B., Stull, S., … Ryan, N. D. (1996). Corticotropin-releasing hormone challenge in prepubertal major depression. *Biological Psychiatry*, *39*(4), 267–277. http://doi.org/10.1016/0006-3223(95)00177-8

Björkenstam, E., Burström, B., Vinnerljung, B., & Kosidou, K. (2016). Childhood adversity and psychiatric disorder in young adulthood: An analysis of 107,704 Swedes. *Journal of Psychiatric Research*, *77*, 67–75. http://doi.org/10.1016/j.jpsychires.2016.02.018

Blake, M. J., Trinder, J. A., & Allen, N. B. (2018). Mechanisms underlying the association between insomnia, anxiety, and depression in adolescence: Implications for behavioral sleep interventions. *Clinical Psychology Review*, *63*, 25–40. http://doi.org/10.1016/J.CPR.2018.05.006

Bock, C., Bukh, J. D., Vinberg, M., Gether, U., & Kessing, L. V. (2010). The Influence of Comorbid Personality Disorder and Neuroticism on Treatment Outcome in First Episode Depression. *Psychopathology*, *43*, 197–204. http://doi.org/10.1159/000304176

Bockting, C. L. H., Spinhoven, P., Koeter, M. W. J., Wouters, L. F., Visser, I., & Schene, A. H. (2006). Differential Predictors of Response to Preventive Cognitive Therapy in Recurrent Depression : A 2 - Year Prospective Study. *Psychother Psychosom*, *75*, 229–236. http://doi.org/10.1159/000092893

Boes, A. D., McCormick, L. M., Coryell, W. H., & Nopoulos, P. (2008). Rostral anterior cingulate cortex volume correlates with depressed mood in normal healthy children. *Biological Psychiatry*, *63*(4), 391–7. http://doi.org/10.1016/j.biopsych.2007.07.018

Bora, E., Fornito, A., Pantelis, C., & Yücel, M. (2012). Gray matter abnormalities in Major Depressive Disorder: A meta-analysis of voxel based morphometry studies. *Journal of Affective Disorders*, *138*(1–2), 9–18. http://doi.org/10.1016/j.jad.2011.03.049

Bouma, E. M. C., Riese, H., Ormel, J., Verhulst, F. C., & Oldehinkel, A. J. (2009). Adolescents' cortisol responses to awakening and social stress; Effects of gender, menstrual phase and oral contraceptives. The TRAILS study. *Psychoneuroendocrinology*, *34*(6), 884–893.

http://doi.org/10.1016/J.PSYNEUEN.2009.01.003

Braun, D. L., Sunday, S. R., & Halmi, K. A. (1994). Psychiatric comorbidity in patients with eating disorders1. *Psychological Medicine*, *24*(4), 859–867. http://doi.org/10.1017/S0033291700028956

Braun, S. R., Gregor, B., & Tran, U. S. (2013). Comparing Bona Fide Psychotherapies of Depression in Adults with Two Meta-Analytical Approaches. *PLoS ONE*, *8*(6), e68135. http://doi.org/10.1371/journal.pone.0068135

Brendgen, M., Wanner, B., Morin, A. J. S., & Vitaro, F. (2005). Relations with Parents and with Peers, Temperament, and Trajectories of Depressed Mood During Early Adolescence. *Journal of Abnormal Child Psychology*, *33*(5), 579–594. http://doi.org/10.1007/s10802-005-6739-2

Brent, D. A., Holder, D., Kolko, D., Birmaher, B., Baugher, M., Roth, C., … Johnson, B. A. (1997). A Clinical Psychotherapy Trial for Adolescent Depression Comparing Cognitive, Family, and Supportive Therapy. *Archives of General Psychiatry*, *54*(9), 877–885. http://doi.org/10.1001/archpsyc.1997.01830210125017

Brent, D. A., Kolko, D. J., Birmaher, B., Baugher, M., Bridge, J., Roth, C., & Holder, D. (1998). Predictors ofTreatment Efficacy in a Clinical Trial of Three Psychosocial Treatments for Adolescent Depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*(9), 906–914. Retrieved from https://ac.els-cdn.com/S089085670962774X/1-s2.0-S089085670962774X-main.pdf?_tid=aac84017-cc97-4917-85ab-f433a17fa8d9&acdnat=1544803284_7ad51295f243035c0d377fbb48a9dc98

Brent, D., Emslie, G., Clarke, G., Wagner, K. D., Asarnow, J. R., Keller, M., … Zelazny, J. (2008). Switching to Another SSRI or to Venlafaxine With or Without Cognitive Behavioral Therapy for Adolescents With SSRI-Resistant Depression. *JAMA*, *299*(8), 901–913. http://doi.org/10.1001/jama.299.8.901

Brière, F. N., Janosz, M., Fallu, J.-S., & Morizot, J. (2015). Adolescent Trajectories of Depressive Symptoms: Codevelopment of Behavioral and Academic Problems. *Journal of Adolescent Health*, *57*(3), 313–319. http://doi.org/10.1016/J.JADOHEALTH.2015.05.012

Brière, F. N., Rohde, P., Stice, E., & Morizot, J. (2016). Group-based symptom trajectories in indicated prevention of adolescent depression. *Depression and Anxiety*, *33*(5), 444–451. http://doi.org/10.1002/da.22440

Buckley, P. F., Miller, B. J., Lehrer, D. S., & Castle, D. J. (2009). Psychiatric Comorbidities and Schizophrenia. *Schizophrenia Bulletin*, *35*(2), 383–402. http://doi.org/10.1093/schbul/sbn135

Buckley, T. M., & Schatzberg, A. F. (2005). REVIEW: On the Interactions of the Hypothalamic-Pituitary-Adrenal (HPA) Axis and Sleep: Normal HPA Axis Activity and Circadian Rhythm, Exemplary Sleep Disorders. *The Journal of Clinical Endocrinology and Metabolism*, *90*(5), 3106–3114.

Bullmore, E. (2012). The future of functional MRI in clinical medicine. *NeuroImage*, *62*(2), 1267–71. http://doi.org/10.1016/j.neuroimage.2012.01.026

Bulmer, M. G. (1979). *Principles of statistics*. Dover. Retrieved from https://books.google.co.uk/books?id=dh24EaSrmBkC&q=skewness#v=snippet&q=skewness&f=false

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*(4), 625–641. http://doi.org/10.1016/S0896-6273(02)00830-9

Burke, H., Davis, M., Otte, C., & Mohr, D. (2005). Depression and cortisol responses to psychological stress: A meta-analysis. *Psychoneuroendocrinology*, *30*, 846–856. Retrieved from file://me-filer1/home$/sed48/My Documents/MR-IMPACT/Cortisol Data/Cortisol papers/Burke-2005-PNEC.pdf

Burleson Daviss, W., Birmaher, B., Melhem, N. A., Axelson, D. A., Michaels, S. M., & Brent, D. A. (2006). Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. *Journal of Child Psychology and Psychiatry*, *47*(9), 927–934. http://doi.org/10.1111/j.1469-7610.2006.01646.x

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222. http://doi.org/10.1016/S1364-6613(00)01483-2

Bylund, D. B., & Reed, A. L. (2007). Childhood and adolescent depression: Why do children and adults respond differently to antidepressant drugs? *Neurochemistry International*, *51*(5), 246–253. http://doi.org/10.1016/J.NEUINT.2007.06.025

Cameron, O. G. (2006). Anxious-depressive comorbidity: effects on HPA axis and CNS noradrenergic functions. *Essential Psychopharmacology*, *7*(1), 24–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16989290

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain

development and its relation to cognitive development. *Biological Psychology*, *54*, 241–257. Retrieved from www.elsevier.com/locate/biopsycho

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the Role of the Insula in Human Cognition: Functional Parcellation and Large-Scale Reverse Inference. *Cerebral Cortex*, *23*, 739–749.

Chen, C.-H., Ridler, K., Suckling, J., Williams, S., Fu, C. H. Y., Merlo-Pich, E., & Bullmore, E. (2007). Brain Imaging Correlates of Depressive Symptom Severity and Predictors of Symptom Improvement After Antidepressant Treatment. *Biological Psychiatry*, *62*, 407–414. http://doi.org/10.1016/j.biopsych.2006.09.018

Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, *80*(3), 265–278. http://doi.org/10.1016/J.BIOPSYCHO.2008.10.004

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Joseph L. Izzo, J., … Committee, and the N. H. B. P. E. P. C. (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. *JAMA*, *289*(19), 2560–2571. http://doi.org/10.1001/jama.289.19.2560

Chrousos, G. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, *5*, 374–381.

Chuang, J.-Y., J Whitaker, K., Murray, G. K., Elliott, R., Hagan, C. C., Graham, J. M., … Suckling, J. (2016). Aberrant brain responses to emotionally valent words is normalised after cognitive behavioural therapy in female depressed adolescents. *Journal of Affective Disorders*, *189*, 54–61. http://doi.org/10.1016/j.jad.2015.09.008

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., … Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366. http://doi.org/10.1016/S0140-6736(17)32802-7

Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., … Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *The Lancet*, *388*(10047), 881–890. http://doi.org/10.1016/S0140-6736(16)30385-3

Clarke, G., Hops, H., Lewinsohn, P. M., Andrews, J., Seeley, J. R., & Williams, J. (1992).

Cognitive-Behavioral Group Treatment of Adolescent Depression: Prediction of Outcome. *Behavior Therapy*, *23*, 341–354.

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., … Tamminga, C. A. (2016). Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *American Journal of Psychiatry*, *173*(4), 373–384. http://doi.org/10.1176/appi.ajp.2015.14091200

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. http://doi.org/10.1177/001316446002000104

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. http://doi.org/10.1177/014662168300700301

Compton, S. N., March, J. S., Brent, D., Albano, A. M., Weersing, V. R., & Curry, J. (2004). Cognitive-Behavioral Psychotherapy for Anxiety and Depressive Disorders in Children and Adolescents: An Evidence-Based Medicine Review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(8), 930–959. http://doi.org/10.1097/01.CHI.0000127589.57468.BF

Cook, R. D., & Weisberg, S. (1982). *Detection of Influential Observation in Linear Regression*. New York: Chapman & Hall. http://doi.org/10.1080/00401706.1977.10489493

Corey-Lisle, P. K., Nash, R., Stang, P., & Swindle, R. (2004). Response, Partial Response, and Nonresponse in Primary Care Treatment of Depression. *Archives of Internal Medicine*, *164*(11), 1197–1204. http://doi.org/10.1001/archinte.164.11.1197

Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. Y. (2009). Prognostic and Diagnostic Potential of the Structural Neuroanatomy of Depression. *PLoS ONE*, *4*(7), e6353. http://doi.org/10.1371/journal.pone.0006353

Costello, D. M., Swendsen, J., Rose, J. S., & Dierker, L. C. (2008). Risk and protective factors associated with trajectories of depressed mood from adolescence to early adulthood. *Journal of Consulting and Clinical Psychology*, *76*(2), 173–83. http://doi.org/10.1037/0022-006X.76.2.173

Cotter, D. (2002). Reduced Neuronal Size and Glial Cell Density in Area 9 of the Dorsolateral Prefrontal Cortex in Subjects with Major Depressive Disorder. *Cerebral Cortex*, *12*(4), 386–394. http://doi.org/10.1093/cercor/12.4.386

Cousins, L., Whitaker, K. J., Widmer, B., Midgley, · Nick, Byford, S., Dubicka, · Bernadka, … Goodyer, I. M. (2016). Clinical characteristics associated with the prescribing of SSRI

medication in adolescents with major unipolar depression. *European Child & Adolescent Psychiatry*, *25*(12), 1287–1295. http://doi.org/10.1007/s00787-016-0849-y

Craig, A. (2009). How do you feel now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70.

Cristea, I. A., Huibers, M. J. H., David, D., Hollon, S. D., Andersson, G., & Cuijpers, P. (2015). The effects of cognitive behavior therapy for adult depression on dysfunctional thinking: A meta-analysis. *Clinical Psychology Review*, *42*, 62–71. http://doi.org/10.1016/J.CPR.2015.08.003

Cropley, M., Rydstedt, L. W., Devereux, J. J., & Middleton, B. (2015). The Relationship Between Work-Related Rumination and Evening and Morning Salivary Cortisol Secretion. *Stress and Health*, *31*(2), 150–157. http://doi.org/10.1002/smi.2538

Cuijpers, P. (2017). Four Decades of Outcome Research on Psychotherapies for Adult Depression: An Overview of a Series of Meta-Analyses. *Canadian Psychology*, *58*(1), 7–19. http://doi.org/10.1037/cap0000096

Cuijpers, P., Andersson, G., Donker, T., & van Straten, A. (2011). Psychological treatment of depression: Results of a series of meta-analyses. *Nordic Journal of Psychiatry*, *65*(6), 354–364. http://doi.org/10.3109/08039488.2011.596570

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison with other Treatments. *The Canadian Journal of Psychiatry*, *58*(7), 376–385. http://doi.org/10.1177/070674371305800702

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry*, *12*(2), 137–148. http://doi.org/10.1002/wps.20038

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. http://doi.org/10.1002/wps.20089

Cuijpers, P., Straten, A. van, Oppen, P. van, & Andersson, G. (2008). Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. Retrieved January 23, 2019, from https://www.ncbi.nlm.nih.gov/books/NBK98252/

Cuijpers, P., van Lier, P. A. C., van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders*, *89*(1–3), 137–146. http://doi.org/10.1016/j.jad.2005.09.001

Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2009a). Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Depression and Anxiety*, *26*(3), 279–288. http://doi.org/10.1002/da.20519

Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2009b). Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Depression and Anxiety*, *26*(3), 279–288. http://doi.org/10.1002/da.20519

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. http://doi.org/10.1002/da.22328

Cullen, K. R., Gee, D. G., Klimes-Dougan, B., Gabbay, V., Hulvershorn, L., Mueller, B. A., … Milham, M. P. (2009). A preliminary study of functional connectivity in comorbid adolescent depression. *Neuroscience Letters*, *460*, 227–231.

Cunningham, E. T., & Sawchenko, P. E. (1988). Anatomical specificity of noradrenergic inputs to the paraventricular and supraoptic nuclei of the rat hypothalamus. *The Journal of Comparative Neurology*, *274*(1), 60–76. http://doi.org/10.1002/cne.902740107

Curran, P. J., Obeidat, K., & Losardo, D. (2011). Twelve Frequently Asked Questions About Growth Curve Modeling. *J Cogn Dev*, *11*(2), 121–136. http://doi.org/10.1080/15248371003699969

Curran, P. J., & Willoughby, M. T. (2003). Implications of latent trajectory models for the study of developmental psychopathology. *Development and Psychopathology*, *15*, 581–612. http://doi.org/10.1017.S0954579403000300

Curry, J., Rohde, P., Simons, A., Silva, S., Vitiello, B., Kratochvil, C., … March, J. (2006). Predictors and Moderators of Acute Outcome in the Treatment for Adolescents With Depression Study (TADS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(12), 1427–1439. http://doi.org/10.1097/01.chi.0000240838.78984.e2

Dahl, R. E., Ryan, N. D., Puig-Antich, J., Nguyen, N. A., Al-Shabbout, M., Meyer, V. A., & Perel,

J. (1991). 24-Hour cortisol measures in adolescents with major depression: A controlled study. *Biological Psychiatry*, *30*(1), 25–36. http://doi.org/10.1016/0006-3223(91)90067-V

Dale, A., Fischl, B., & Sereno, M. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, *9*, 179–194.

Dedovic, K., & Ngiam, J. (2015). The cortisol awakening response and major depression: examining the evidence. *Neuropsychiatric Disease and Treatment*, *11*, 1181–1189. http://doi.org/10.2147/NDT.S62289

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980. http://doi.org/10.1016/j.neuroimage.2006.01.021

Dettenborn, L., Muhtz, C., Skoluda, N., Stalder, T., Steudte, S., Hinkelmann, K., … Otte, C. (2012). Introducing a novel method to assess cumulative steroid concentrations: Increased hair cortisol concentrations over 6 months in medicated patients with depression. *Stress*, *15*(3), 348–353. http://doi.org/10.3109/10253890.2011.619239

Devinsky, O., Morrell, M. J., & Vogt, B. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain*, *118*, 279–306. http://doi.org/10.1093/brain/118.1.279

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., … Koob, G. F. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science (New York, N.Y.)*, *325*(5940), 621–5. http://doi.org/10.1126/science.1171203

Diaz, S. L., Doly, S., Narboux-Nême, N., Fernández, S., Mazot, P., Banas, S. M., … Maroteaux, L. (2012). 5-HT2B receptors are required for serotonin-selective antidepressant actions. *Molecular Psychiatry*, *17*(2), 154–163. http://doi.org/10.1038/mp.2011.159

Dickstein, G., Shechner, C., Nicholson, W., Rosner, I., Shen-Orr, Z., Adawi, F., & Lahav, M. (1991). Adrenocorticotropin Stimulation Test: Effects of Basal Cortisol Level, Time of Day, and Suggested New Sensitive Low Dose Test*. *The Journal of Clinical Endocrinology & Metabolism*, *72*(4), 773–778. http://doi.org/10.1210/jcem-72-4-773

Dominguez, M. D. G., Wichers, M., Lieb, R., Wittchen, H.-U., & van Os, J. (2011). Evidence that onset of clinical psychosis is an outcome of progressively more persistent subclinical psychotic experiences: an 8-year cohort study. *Schizophrenia Bulletin*, *37*(1), 84–93. http://doi.org/10.1093/schbul/sbp022

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222. http://doi.org/10.1186/2193-1801-2-222

Dotson, V. M., Szymkowicz, S. M., Sozda, C. N., Kirton, J. W., Green, M. L., O'Shea, A., … Woods, A. J. (2016). Age Differences in Prefrontal Surface Area and Thickness in Middle Aged to Older Adults. *Frontiers in Aging Neuroscience*, *7*, 1–9. http://doi.org/10.3389/fnagi.2015.00250

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. http://doi.org/10.1038/nm.4246

Ducharme, S., Albaugh, M. D., Hudziak, J. J., Botteron, K. N., Nguyen, T.-V., Truong, C., … Karama, S. (2014). Anxious/depressed symptoms are linked to right ventromedial prefrontal cortical thickness maturation in healthy children and young adults. *Cerebral Cortex*, *24*(11), 2941–2950. http://doi.org/10.1093/cercor/bht151

Dudek, D., Rybakowski, J. ., Siwek, M., Pawlowski, T., Lojko, D., Roczén, R., & Kiejna, A. (2010). Risk factors of treatment resistance in major depression: Association with bipolarity. *Journal of Affective Disorders*, *126*, 268–271.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Dunlop, B. W., & Mayberg, H. S. (2014). Neuroimaging-based biomarkers for treatment selection in major depressive disorder. *Dialogues in Clinical Neuroscience*, *16*(4), 479–490.

Dunn, V., & Goodyer, I. M. (2006). Longitudinal investigation into childhood-and adolescence-onset depression: Psychiatric outcome in early adulthood. *British Journal of Psychiatry*, *188*(3), 216–222. http://doi.org/10.1192/bjp.188.3.216

Duval, E. R., Javanbakht, A., & Liberzon, I. (2015). Neural circuits in anxiety and stress disorders: a focused review. *Therapeutics and Clinical Risk Management*, *11*, 115–126. http://doi.org/10.2147/TCRM.S48528

Edwards, S., Evans, P., Hucklebridge, F., & Clow, A. (2001). Association between time of awakening and diurnal cortisol secretory activity. *Psychoneuroendocrinology*, *26*(6), 613–622. http://doi.org/10.1016/S0306-4530(01)00015-4

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., … Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences*,

*98*(12), 6917–6922. http://doi.org/10.1073/pnas.111134598

Ehlert, U., Gaab, J., & Heinrichs, M. (2001a). Psychoneuroendocrinological contributions to the etiology of depression, posttraumatic stress disorder, and stress-related bodily disorders: the role of the hypothalamus–pituitary–adrenal axis. *Biological Psychology*, *57*(1–3), 141–152. http://doi.org/10.1016/S0301-0511(01)00092-8

Ehlert, U., Gaab, J., & Heinrichs, M. (2001b). Psychoneuroendocrinological contributions to the etiology of depression, posttraumatic stress disorder, and stress-related bodily disorders: the role of the hypothalamus–pituitary–adrenal axis. *Biological Psychology*, *57*(1), 141–152. http://doi.org/10.1016/S0301-0511(01)00092-8

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable Functions in the Medial and Lateral Orbitofrontal Cortex: Evidence from Human Neuroimaging Studies. *Cerebral Cortex*, *10*(3), 308–317. http://doi.org/10.1093/cercor/10.3.308

Emslie, G. J., Heiligenstein, J. H., Wagner, K. D., Hoog, S. L., Ernest, D. E., Brown, E., … Jacobson, J. G. (2002). Fluoxetine for Acute Treatment of Depression in Children and Adolescents: A Placebo-Controlled, Randomized Clinical Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(10), 1205–1215. http://doi.org/10.1097/00004583-200210000-00010

Emslie, G. J., Rush, A. J., Weinberg, W. A., Kowatch, R. A., Hughes, C. W., Carmody, T., & Rintelmann, J. (1997). A Double-blind, Randomized, Placebo-Controlled Trial of Fluoxetine in Children and Adolescents With Depression. *Archives of General Psychiatry*, *54*(11), 1031–1037. http://doi.org/10.1001/archpsyc.1997.01830230069010

Etkin, A., & Wager, T. D. (2007). Functional Neuroimaging of Anxiety: A Meta-Analysis of Emotional Processing in PTSD, Social Anxiety Disorder, and Specific Phobia. *American Journal of Psychaitry*, *164*, 1476–1488.

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, *8*(1), 19–20. http://doi.org/10.1017/S1481803500013336

Farb, N. A. S., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, *2*(4), 313–322. http://doi.org/10.1093/scan/nsm030

Farmer, A., Korszun, A., Owen, M. J., Craddock, N., Jones, L., Jones, I., … McGuffin, P. (2008). Medical disorders in people with recurrent depression. *British Journal of Psychiatry*,

*192*(5), 351–355. http://doi.org/10.1192/bjp.bp.107.038380

Fineberg, N. A., Haddad, P. M., Carpenter, L., Gannon, B., Sharpe, R., Young, A. H., … Sahakian, B. J. (2013). The size, burden and cost of disorders of the brain in the UK. *Journal of Psychopharmacology*, *27*(9), 761–770. http://doi.org/10.1177/0269881113495118

Fischer, S., & Cleare, A. J. (2017). Cortisol as a predictor of psychological therapy response in anxiety disorders—Systematic review and meta-analysis. *Journal of Anxiety Disorders*, *47*, 60–68. http://doi.org/10.1016/J.JANXDIS.2017.02.007

Fischer, S., Macare, C., & Cleare, A. J. (2017). Hypothalamic-pituitary-adrenal (HPA) axis functioning as predictor of antidepressant response–Meta-analysis. *Neuroscience & Biobehavioral Reviews*, *83*, 200–211. http://doi.org/10.1016/J.NEUBIOREV.2017.10.012

Fischer, S., Strawbridge, R., Vives, A. H., & Cleare, A. J. (2017). Cortisol as a predictor of psychological therapy response in depressive disorders: Systematic review and meta-analysis. *British Journal of Psychiatry*, *210*(2), 105–109. http://doi.org/10.1192/bjp.bp.115.180653

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. http://doi.org/10.1016/j.neuroimage.2012.01.021

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*(20), 11050–11055. http://doi.org/10.1073/pnas.200033797

Fischl, B., Sereno, M., & Dale, A. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*, 195–207.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., … Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, *14*(1), 11–22. http://doi.org/10.1093/cercor/bhg087

Foland-Ross, L. C., Sacchet, M. D., Prasad, G., Gilbert, B., Thompson, P. M., & Gotlib, I. H. (2015). Cortical thickness predicts the first onset of major depression in adolescence. *International Journal of Developmental Neuroscience*, *46*, 125–31. http://doi.org/10.1016/j.ijdevneu.2015.07.007

Fombonne, E., Wostear, G., Cooper, V., Harrington, R., & Rutter, M. (2001). The Maudsley long-term follow-up of child and adolescent depression. 1. Psychiatric outcomes in adulthood. *The British Journal of Psychiatry*, *179*, 210–7.

Forbes, E. E., Olino, T. M., Ryan, N. D., Birmaher, B., Axelson, D., Moyles, D. L., & Dahl, R. E.

(2010). Reward-related brain function as a predictor of treatment response in adolescents with major depressive disorder. *Cognitive, Affective and Behavioural Neuroscience*, *10*(1), 107–118. http://doi.org/10.3758/CABN.10.1.107

Forbes, E. E., Williamson, D. E., Ryan, N. D., Birmaher, B., Axelson, D. A., & Dahl, R. E. (2006). Peri-Sleep-Onset Cortisol Levels in Children and Adolescents with Affective Disorders. *Biological Psychiatry*, *59*(1), 24–30. http://doi.org/10.1016/J.BIOPSYCH.2005.06.002

Fornito, A., Wood, S. J., Whittle, S., Fuller, J., Adamson, C., Saling, M. M., … Yücel, M. (2008). Variability of the paracingulate sulcus and morphometry of the medial frontal cortex: Associations with cortical thickness, surface area, volume, and sulcal depth. *Human Brain Mapping*, *29*(2), 222–236. http://doi.org/10.1002/hbm.20381

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: a review. *L'Encephale*, *28*(2), 97–107.

Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A., … Forgione, R. N. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine*, *41*(1), 151–162. http://doi.org/10.1017/S0033291710000553

Frank, E., Shear, M. K., Rucci, P., Cyranowski, J. M., Endicott, J., Fagiolini, A., … Cassano, G. B. (2000). Influence of Panic-Agoraphobic Spectrum Symptoms on Treatment Response in Patients With Recurrent Major Depression. *American Journal of Psychiatry*, *157*(7), 1101–1107. http://doi.org/10.1176/appi.ajp.157.7.1101

Frodl, T., Jäger, M., Born, C., Ritter, S., Kraft, E., Zetzsche, T., … Meisenzahl, E. (2008). Anterior cingulate cortex does not differ between patients with major depression and healthy controls, but relatively large anterior cingulate cortex predicts a good clinical course. *Psychiatry Research: Neuroimaging*, *163*, 76–83.

FsAverage - Free Surfer Wiki. (n.d.). Retrieved April 14, 2019, from https://surfer.nmr.mgh.harvard.edu/fswiki/FsAverage

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

Fu, C. H. Y., Williams, S. C. R., Cleare, A. J., Scott, J., Mitterschiffthaler, M. T., Walsh, N. D., … Murray, R. M. (2008). Neural Responses to Sad Facial Expressions in Major Depression Following Cognitive Behavioral Therapy. *Biological Psychiatry*, *64*, 505–512.

http://doi.org/10.1016/j.biopsych.2008.04.033

Fuligni, A. J., Arruda, E. H., Krull, J. L., & Gonzales, N. A. (2018). Adolescent Sleep Duration, Variability, and Peak Levels of Achievement and Mental Health. *Child Development*, *89*(2), e18–e28. http://doi.org/10.1111/cdev.12729

Fusar-Poli, P., Nelson, B., Valmaggia, L., Yung, A. R., & McGuire, P. K. (2014). Comorbid Depressive and Anxiety Disorders in 509 Individuals With an At-Risk Mental State: Impact on Psychopathology and Transition to Psychosis. *Schizophrenia Bulletin*, *40*(1), 120–131. http://doi.org/10.1093/schbul/sbs136

Galard, R., Catalán, R., Castellanos, J. M., & Gallart, J. M. (2002). Plasma corticotropin-releasing factor in depressed patients before and after the dexamethasone suppression test. *Biological Psychiatry*, *51*(6), 463–468.

Garber, J., & Weersing, V. R. (2010). Comorbidity of Anxiety and Depression in Youth: Implications for Treatment and Prevention. *Clinical Psychology: Science and Practice*, *17*(4), 293–306. http://doi.org/10.1111/j.1468-2850.2010.01221.x

Gartlehner, G., Thieda, P., Hansen, R. A., Gaynes, B. N., DeVeaugh-Geiss, A., Krebs, E. E., & Lohr, K. N. (2008). Comparative Risk for Harms of Second-Generation Antidepressants. *Drug Safety*, *31*(10), 851–865. http://doi.org/10.2165/00002018-200831100-00004

Gartlehner, G., Wagner, G., Matyas, N., Titscher, V., Greimel, J., Lux, L., … Lohr K, N. (2017). Pharmacological and nonpharmacological treatments for major depressive disorder: review of systematic reviews. *BMJ Open*, *7*. http://doi.org/10.1136/bmjopen-2016-014912

Gautam, P., Anstey, K. J., Wen, W., Sachdev, P. S., & Cherbuin, N. (2015). Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. *Behavioural Brain Research*, *287*, 331–339. http://doi.org/10.1016/J.BBR.2015.03.018

Ge, T., Yeo, B., & Winkler, A. (2018). A brief overview of permutation testing with examples.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, *15*(4), 870–878. http://doi.org/10.1006/NIMG.2001.1037

Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H. C., Greenhouse, J. B., … Watkins, J. T. (1993). Some Conceptual and Statistical Issues in Analysis of Longitudinal Psychiatric Data. *Archives of General Psychiatry*, *50*(9), 739. http://doi.org/10.1001/archpsyc.1993.01820210073009

Gibson, E., Checkley, S., Papadopoulos, A., Poon, L., Daley, S., & Wardle, J. (1999). Increased salivary cortisol reliably induced by a protein-rich midday meal. *Psychosomatic Medicine*, *61*(2), 214–224.

Giedd, J. N. (2008). The Teen Brain: Insights from Neuroimaging. *Journal of Adolescent Health*, *42*(4), 335–343. http://doi.org/10.1016/J.JADOHEALTH.2008.01.007

Girshkin, L., O'Reilly, N., Quidé, Y., Teroganova, N., Rowland, J. E., Schofield, P. R., & Green, M. J. (2016). Diurnal cortisol variation and cortisol response to an MRI stressor in schizophrenia and bipolar disorder. *Psychoneuroendocrinology*, *67*, 61–69. http://doi.org/10.1016/J.PSYNEUEN.2016.01.021

Goel, N., Workman, J. L., Lee, T. T., Innala, L., & Viau, V. (2014). Sex Differences in the HPA Axis. *Comprehensive Physiology*, *4*, 1121–1155. http://doi.org/10.1002/cphy.c130054

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179. http://doi.org/10.1073/pnas.201243998

Gold, P., Chrousos, G., Kellner, C., Post, R., Psychiatry, A. R.-A. J., & 1984, U. (1984). Psychiatric implications of basic and clinical studies with corticotropin-releasing factor. *American Journal of Psychaitry*, *141*(5), 619–627.

Gold, P. W. (2014). The organization of the stress system and its dysregulation in depressive illness. *Molecular Psychiatry*, *20*, 32. Retrieved from http://dx.doi.org/10.1038/mp.2014.163

Golden, S. H., Malhotra, S., Wand, G. S., Brancati, F. L., Ford, D., & Horton, K. (2007). Adrenal Gland Volume and Dexamethasone-Suppressed Cortisol Correlate with Total Daily Salivary Cortisol in African-American Women. *The Journal of Clinical Endocrinology & Metabolism*, *92*(4), 1358–1363. http://doi.org/10.1210/jc.2006-2674

Gong, Q., & He, Y. (2015). Depression, Neuroimaging and Connectomics: A Selective Overview. *Biological Psychiatry*, *77*(3), 223–235. http://doi.org/10.1016/J.BIOPSYCH.2014.08.009

Gong, Q., Wu, Q., Scarpazza, C., Lui, S., Jia, Z., Marquand, A., … Mechelli, A. (2011). Prognostic prediction of therapeutic response in depression using high-field MR imaging. *NeuroImage*, *55*(4), 1497–1503. http://doi.org/10.1016/j.neuroimage.2010.11.079

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., … Etkin,

A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, *72*(4), 305–315. http://doi.org/10.1001/jamapsychiatry.2014.2206

Goodyer, I., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S., … Harrington, R. (2007). Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. *BMJ*, *335*, 142. http://doi.org/10.1136/bmj.39224.494340.55

Goodyer, I. M. (2000). Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents. *The British Journal of Psychiatry*, *177*(6), 499–504. http://doi.org/10.1192/bjp.177.6.499

Goodyer, I. M., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S., … Harrington, R. (2008). A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial. *Health Technology Assessment*, *12*(14).

Goodyer, I. M., Herbert, J., Secher, S. M., & Pearson, J. (1997). Short-Term Outcome of Major Depression: I. Comorbidity and Severity at Presentation as Predictors of Persistent Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(2), 179–187. http://doi.org/10.1097/00004583-199702000-00008

Goodyer, I. M., Herbert, J., & Tamplin, A. (2003). Psychoendocrine antecedents of persistent first-episode major depression in adolescents: a community-based longitudinal enquiry. *Psychological Medicine*, *33*(4), S0033291702007286. http://doi.org/10.1017/S0033291702007286

Goodyer, I. M., Reynolds, S., Barrett, B., Byford, S., Dubicka, B., Hill, J., … Fonagy, P. (2017). Cognitive-behavioural therapy and short-term psychoanalytic psychotherapy versus brief psychosocial intervention in adolescents with unipolar major depression (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled trial. *Health Technology Assessment (Winchester, England)*, *21*(12), 1–94. http://doi.org/10.3310/hta21120

Goodyer, I. M., & Wilkinson, P. O. (2018). Practitioner Review: Therapeutics of unipolar major depressions in adolescents. *The Journal of Child Psychology and Psychiatry*. http://doi.org/10.1111/jcpp.12940

Goodyer, I., Reynolds, S., Barrett, B., Byford, S., Dubicka, B., Hill, J., … Fonagy, P. (2017). Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a

brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiori. *Lancet Psychiatry*, *4*(2), 109–119. http://doi.org/doi.org/10.1016/S2215-0366(16)30378-9

Goodyer, I., Tsancheva, S., Byford, S., Dubicka, B., Hill, J., Kelvin, R., … Fonagy, P. (2011). Improving mood with psychoanalytic and cognitive therapies (IMPACT): a pragmatic effectiveness superiority trial to investigate whether specialised psychological treatment reduces the risk for relapse in adolescents with moderate to severe unipolar depres. *Trials*, *12*(1), 175. http://doi.org/10.1186/1745-6215-12-175

Gorwood, P. (2008). Neurobiological mechanisms of anhedonia. *Dialogues in Clinical Neuroscience*, *10*(3), 291–299.

Gottfredson, N. C., Sterba, S. K., & Jackson, K. M. (2017). Explicating the Conditions Under Which Multilevel Multiple Imputation Mitigates Bias Resulting from Random Coefficient-Dependent Missing Longitudinal Data. *Prevention Science*, *18*(1), 12–19. http://doi.org/10.1007/s11121-016-0735-3

Gowers, S. G., Harrington, R. C., Whitton, A., Lelliot, P., Beevor, A., Wing, J., & Jezzard, R. (1999). Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *British Journal of Psychiatry*, *174*(5), 428–431. http://doi.org/10.1192/bjp.174.5.428

Graham, J. W. (2003). Adding Missing-Data-Relevant Variables to FIMLBased Structural Equation Models. *Structural Equation Modeling*, *10*(1), 80–100.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science*, *8*(3), 206–213. http://doi.org/10.1007/s11121-007-0070-9

Gray, T. S., Carney, M. E., & Magnuson, D. J. (1989). Direct projections from the central amygdaloid nucleus to the hypothalamic paraventricular nucleus: possible role in stress-induced adrenocorticotropin release. *Neuroendocrinology*, *50*(4), 433–46. http://doi.org/10.1159/000125260

Greeson, J. M., Smoski, M. J., Suarez, E. C., Brantley, J. G., Ekblad, A. G., Lynch, T. R., & Wolever, R. Q. (2015). Decreased Symptoms of Depression After Mindfulness-Based Stress Reduction: Potential Moderating Effects of Religiosity, Spirituality, Trait Mindfulness, Sex, and Age. *The Journal of Alternative and Complementary Medicine*, *21*(3), 166–174. http://doi.org/10.1089/acm.2014.0285

Grieve, S. M., Korgaonkar, M. S., Koslow, S. H., Gordon, E., & Williams, L. M. (2013).

Widespread reductions in gray matter volume in depression. *NeuroImage: Clinical*, *3*, 332–339. http://doi.org/10.1016/J.NICL.2013.08.016

Grimm, S., Boesiger, P., Beck, J., Schuepbach, D., Bermpohl, F., Walter, M., … Northoff, G. (2009). Altered Negative BOLD Responses in the Default-Mode Network during Emotion Processing in Depressed Subjects. *Neuropsychopharmacology*, *34*(4), 932–943. http://doi.org/10.1038/npp.2008.81

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), 1227–1237. http://doi.org/10.1001/archgenpsychiatry.2011.132

Guerry, J. D., & Hastings, P. D. (2011). In search of HPA axis dysregulation in child and adolescent depression. *Clinical Child and Family Psychology Review*, *14*(2), 135–60. http://doi.org/10.1007/s10567-011-0084-5

Gunlicks-Stoessel, M., Mufson, L., Cullen, K. R., & Klimes-Dougan, B. (2013). A pilot study of depressed adolescents' cortisol patterns during parent–adolescent conflict and response to interpersonal psychotherapy (IPT-A). *Journal of Affective Disorders*, *150*(3), 1125–1128. http://doi.org/10.1016/J.JAD.2013.05.037

Gunnar, M., & Quevedo, K. (2007). The Neurobiology of Stress and Development. *Annual Review of Psychology*, *58*(145–173). http://doi.org/10.1146/annurev.psych.58.110405.085605

Gunnar, M. R., Wewerka, S., Frenn, K., Long, J. D., & Griggs, C. (2009). Developmental changes in hypothalamus–pituitary–adrenal activity over the transition to adolescence: Normative changes and associations with puberty. *Development and Psychopathology*, *21*(1), 69–85. http://doi.org/10.1017/S0954579409000054

H. Kennedy, S., Downar, J., R. Evans, K., Feilotter, H., W. Lam, R., M. MacQueen, G., … Soares, C. (2012). The Canadian Biomarker Integration Network in Depression (CAN-BIND): Advances in Response Prediction. *Current Pharmaceutical Design*, *18*(36), 5976–5989. http://doi.org/10.2174/138161212803523635

Hagan, C. C., Graham, J. M. E., Tait, R., Widmer, B., van Nieuwenhuizen, A. O., Ooi, C., … Suckling, J. (2015). Adolescents with current major depressive disorder show dissimilar patterns of age-related differences in ACC and thalamus. *NeuroImage. Clinical*, *7*, 391–399. http://doi.org/10.1016/j.nicl.2014.12.019

Hagan, C. C., Graham, J. M., Widmer, B., Holt, R. J., Ooi, C., van Nieuwenhuizen, A. O., …

Suckling, J. (2013). Magnetic resonance imaging of a randomized controlled trial investigating predictors of recovery following psychological treatment in adolescents with moderate to severe unipolar depression: study protocol for Magnetic Resonance-Improving Mood with Psycho. *BMC Psychiatry*, *13*, 247. http://doi.org/10.1186/1471-244X-13-247

Halligan, S. L., Herbert, J., Goodyer, I., & Murray, L. (2007). Disturbances in Morning Cortisol Secretion in Association with Maternal Postnatal Depression Predict Subsequent Depressive Symptomatology in Adolescents. *Biological Psychiatry*, *62*(1), 40–46. http://doi.org/10.1016/j.biopsych.2006.09.011

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional Neuroimaging of Major Depressive Disorder: A Meta-Analysis and New Integration of Baseline Activation and Neural Response Data. *American Journal of Psychiatry*, *169*(7), 693–703. http://doi.org/10.1176/appi.ajp.2012.11071105

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, *23*, 56–62.

Hammack, S. E., Schmid, M. J., LoPresti, M. L., Der-Avakian, A., Pellymounter, M. A., Foster, A. C., … Maier, S. F. (2003). Corticotropin releasing hormone type 2 receptors in the dorsal raphe nucleus mediate the behavioral consequences of uncontrollable stress. *The Journal of Neuroscience*, *23*(3), 1019–1025.

Hammen, C. (2005). Stress and Depression. *Annual Review of Clinical Psychology*, *1*(1), 293–319. http://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Harmer, C. J., & Cowen, P. J. (2013). "It"s the way that you look at it'--a cognitive neuropsychological account of SSRI action in depression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*, 20120407. http://doi.org/10.1098/rstb.2012.0407

Harris, T. O., Borsanyi, S., Messari, S., Stanford, K., Cleary, S. E., Shiers, H. M., … Herbert, J. (2000). Morning cortisol as a risk factor for subsequent major depressive disorder in adult women. *The British Journal of Psychiatry*, *177*, 505–510.

Herbert, J. (2013). Cortisol and depression: three questions for psychiatry. *Psychological Medicine*, *43*, 449–469.

Herbert, J., Goodyer, I. M., Grossman, A. B., Hastings, M. H., De Kloet, E. R., Lightman, -S L, … Seckl, J. R. (2006). Do Corticosteroids Damage the Brain? *Journal of Neruoendocrinology*, *18*, 393–411. http://doi.org/10.1111/j.1365-2826.2006.01429.x

Herman, J. P., & Cullinan, W. E. (1997). Neurocircuitry of stress: central control of the hypothalamo–pituitary–adrenocortical axis. *Trends in Neurosciences*, *20*(2), 78–84. http://doi.org/10.1016/S0166-2236(96)10069-2

Heuser, I. J., Gotthardt, U., Schweiger, U., Schmider, J., Lammers, C.-H., Dettling, M., & Holsboer, F. (1994). Age-associated changes of pituitary-adrenocortical hormone regulation in humans: Importance of gender. *Neurobiology of Aging*, *15*(2), 227–231. http://doi.org/10.1016/0197-4580(94)90117-1

Hinkelmann, K., Muhtz, C., Dettenborn, L., Agorastos, A., Wingenfeld, K., Spitzer, C., … Otte, C. (2013). Association between childhood trauma and low hair cortisol in depressed patients and healthy control subjects. *Biological Psychiatry*, *74*(9), e15-7. http://doi.org/10.1016/j.biopsych.2013.04.021

Hipp, J. R., & Bauer, D. J. (2006). Local Solutions in the Estimation of Growth Mixture Models. *Psychological Methods*, *11*(1), 36–53. http://doi.org/10.1037/1082-989X.11.1.36

Hoaglin, D. C., & Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, *32*(1), 17–22. http://doi.org/10.1080/00031305.1978.10479237

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive Therapy and Research*, *36*(5), 427–440. http://doi.org/10.1007/s10608-012-9476-1

Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., & Fjell, A. M. (2013). The Structure of the Cerebral Cortex Across Adult Life: Age-Related Patterns of Surface Area, Thickness, and Gyrification. *Cerebral Cortex*, *23*(11), 2521–2530. http://doi.org/10.1093/cercor/bhs231

Holland, J. M., Schatzberg, A. F., O'Hara, R., Marquett, R. M., & Gallagher-Thompson, D. (2013). Pretreatment cortisol levels predict posttreatment outcomes among older adults with depression in cognitive behavioral therapy. *Psychiatry Research*, *210*(2), 444–450. http://doi.org/10.1016/j.psychres.2013.07.033

Holsboer, F., Bardeleben, U. v., Gerken, A., Stella, G. K., & Muller O.A. (1984). Blunted Corticotropin and Normal Cortisol Response to Human Corticotropin-Releasing Factor in Depression. *New England Journal of Medicine*, *311*(17), 1127. http://doi.org/10.1056/NEJM198410253111718

Honey, C. J., Thivierge, J. P., & Sporns, O. (2010). Can structure predict function in the human brain? *NeuroImage*, *52*(3), 766–776. http://doi.org/10.1016/j.neuroimage.2010.01.071

Horowitz, J. L., Garber, J., Ciesla, J. A., Young, J. F., & Mufson, L. (2007). Prevention of Depressive Symptoms in Adolescents: A Randomized Trial of Cognitive-Behavioral and

Interpersonal Prevention Programs. *Journal of Consulting and Clincal Psychology*, *75*(5), 693–706. http://doi.org/10.1037/0022-006X.75.5.693

Horstmann, S., & Binder, E. B. (2011). Glucocorticoids as Predictors of Treatment Response in Depression. *Harvard Review of Psychiatry*, *19*(3), 125–143.

Hsiao, F.-H., Yang, T.-T., Ho, R. T. H., Jow, G.-M., Ng, S.-M., Chan, C. L. W., … Wang, K.-C. (2010). The self-perceived symptom distress and health-related conditions associated with morning to evening diurnal cortisol patterns in outpatients with major depressive disorder. *Psychoneuroendocrinology*, *35*(4), 503–515. http://doi.org/10.1016/J.PSYNEUEN.2009.08.019

Huber, T. J., Issa, K., Schik, G., & Wolf, O. T. (2006). The cortisol awakening response is blunted in psychotherapy inpatients suffering from depression. *Psychoneuroendocrinology*, *31*(7), 900–904. http://doi.org/10.1016/J.PSYNEUEN.2006.03.005

Hui, C., Morcillo, C., Russo, D. A., Stochl, J., Shelley, G. F., Painter, M., … Perez, J. (2013). Psychiatric morbidity, functioning and quality of life in young people at clinical high risk for psychosis. *Schizophrenia Research*, *148*(1–3), 175–180. http://doi.org/10.1016/J.SCHRES.2013.05.026

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., 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.

Ising, M., Horstmann, S., Kloiber, S., Lucae, S., Binder, E. B., Kern, N., … Holsboer, F. (2007). Combined Dexamethasone/Corticotropin Releasing Hormone Test Predicts Treatment Response in Major Depression–A Potential Biomarker? *Biological Psychiatry*, *62*(1), 47–54. http://doi.org/10.1016/J.BIOPSYCH.2006.07.039

Ivleva, E. I., Clementz, B. A., Dutcher, A. M., Arnold, S. J. M., Jeon-Slaughter, H., Aslan, S., … Tamminga, C. A. (2017). Brain Structure Biomarkers in the Psychosis Biotypes: Findings From the Bipolar-Schizophrenia Network for Intermediate Phenotypes. *Biological Psychiatry*, *82*(1), 26–39. http://doi.org/10.1016/J.BIOPSYCH.2016.08.030

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., … Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*(2), 295–304. http://doi.org/10.1037//0022-006x.64.2.295

Jaeckle, R., Kathol, R., Lopez, J., Meller, W., & Krummel, S. (1987). Enhanced adrenal

sensitivity to exogenous cosyntropin (ACTH α1-24) stimulation in major depression: relationship to dexamethasone suppression test results. *Archives of General Psychiatry*, *44*(3), 233–240.

Jayson, D., Wood, A., Kroll, L., Fraser, J., & Harrington, R. (1998). Which Depressed Patients Respond to Cognitive-Behavioral Treatment? *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(1), 35–39. http://doi.org/10.1097/00004583-199801000-00014

Jensen, P. S., Hoagwood, K., & Petti, T. (1996). Outcomes of Mental Health Care for Children and Adolescents: II. Literature Review and Application of a Comprehensive Model. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(8), 1064–1077. http://doi.org/10.1097/00004583-199608000-00018

Jeppesen, P., Clemmensen, L., Munkholm, A., Rimvall, M. K., Rask, C. U., Jørgensen, T., … Skovgaard, A. M. (2015). Psychotic experiences co-occur with sleep problems, negative affect and mental disorders in preadolescence. *Journal of Child Psychology and Psychiatry*, *56*(5), 558–565. http://doi.org/10.1111/jcpp.12319

Joëls, M., Hesen, W., & de Kloet, E. R. (1991). Mineralocorticoid hormones suppress serotonin-induced hyperpolarization of rat hippocampal CA1 neurons. *The Journal of Neuroscience*, *11*(8), 2288–2294. http://doi.org/10.1523/JNEUROSCI.11-08-02288.1991

Jureidini, J. N., Doecke, C. J., Mansfield, P. R., Haby, M. M., Menkes, D. B., & Tonkin, A. L. (2004). Efficacy and safety of antidepressants for children and adolescents. *BMJ*, *328*, 879–883.

Jureidini, J. N., Mchenry, L. B., & Mansfield, P. R. (2008). Clinical trials and drug promotion: Selective reporting of study 329. *International Journal of Risk & Safety in Medicine*, *20*, 73–81. http://doi.org/10.3233/JRS-2008-0426

Juruena, M. F., Pariante, C. M., Papadopoulos, A. S., Poon, L., Lightman, S., & Cleare, A. J. (2009). Prednisolone suppression test in depression: prospective study of the role of HPA axis dysfunction in treatment resistance. *British Journal of Psychiatry*, *194*(4), 342–349. http://doi.org/10.1192/bjp.bp.108.050278

Jyotsna, V. P., Naseer, A., Sreenivas, V., Gupta, N., & Deepak, K. K. (2011). Effect of Cushing's syndrome — Endogenous hypercortisolemia on cardiovascular autonomic functions. *Autonomic Neuroscience*, *160*(1–2), 99–102. http://doi.org/10.1016/J.AUTNEU.2010.11.007

Kahl, K. G., Schweiger, U., Pars, K., Kunikowska, A., Deuschle, M., Gutberlet, M., … Hartung,

D. (2015). Adrenal gland volume, intra-abdominal and pericardial adipose tissue in major depressive disorder. *Psychoneuroendocrinology*, *58*, 1–8. http://doi.org/10.1016/J.PSYNEUEN.2015.04.008

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. *JAMA Psychiatry*, *72*(6), 603–11. http://doi.org/10.1001/jamapsychiatry.2015.0071

Karyotaki, Smit, Y., de Beurs, D. P., Henningsen, K. H., Robays, J., Huibers, M. J. H., … Cuijpers, P. (2016). The long-term efficacy of acute-phase psychotherapy for depression: A meta-analysis of randomised trials. *Depression and Anxiety*, *33*(5), 370–383. http://doi.org/10.1002/da.22491

Karyotaki, Smit, Y., Holdt Henningsen, K., Huibers, M. J. H., Robays, J., de Beurs, D., & Cuijpers, P. (2016). Combining pharmacotherapy and psychotherapy or monotherapy for major depression? A meta-analysis on the long-term effects. *Journal of Affective Disorders*, *194*, 144–152. http://doi.org/10.1016/J.JAD.2016.01.036

Katon, W., Unützer, J., & Russo, J. (2010). Major depression: the importance of clinical characteristics and treatment response to prognosis. *Depression and Anxiety*, *27*, 19–26.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., … Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(7), 980–988. http://doi.org/10.1097/00004583-199707000-00021

Kaufman, J., Martin, A., King, R. A., & Charney, D. (2001). Are child-, adolescent-, and adult-onset depression one and the same disorder? *Biological Psychiatry*, *49*(12), 980–1001. http://doi.org/10.1016/S0006-3223(01)01127-1

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, *44*(8), 1116–1129. http://doi.org/10.1111/1469-7610.00195

Keedwell, P. A., Drapier, D., Surguladze, S., Giampietro, V., Brammer, M., & Phillips, M. (2010). Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *Journal of Affective*

*Disorders*, *120*, 120–125.

Kelleher, I., Connor, D., Clarke, M. C., Devlin, N., Harley, M., & Cannon, M. (2012). Prevalence of psychotic symptoms in childhood and adolescence: a systematic review and meta-analysis of population-based studies. *Psychological Medicine*, *42*(9), 1857–1863. http://doi.org/10.1017/S0033291711002960

Kelleher, I., Keeley, H., Corcoran, P., Lynch, F., Fitzpatrick, C., Devlin, N., … Cannon, M. (2012). Clinicopathological significance of psychotic experiences in non-psychotic young people: evidence from four population-based studies. *The British Journal of Psychiatry : The Journal of Mental Science*, *201*(1), 26–32. http://doi.org/10.1192/bjp.bp.111.101543

Keller, M. B. (2003). Past, Present, and Future Directions for Defining Optimal Treatment Outcome in Depression. *JAMA*, *289*(23), 3152. http://doi.org/10.1001/jama.289.23.3152

Keller, M. B. (2005). Issues in Treatment-Resistant Depression. *Journal of Clinical Psychiatry*, *6666*(8), 5–12.

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., … Zajecka, J. (2000). A Comparison of Nefazodone, the Cognitive Behavioral-Analysis System of Psychotherapy, and Their Combination for the Treatment of Chronic Depression. *New England Journal of Medicine*, *342*(20), 1462–1470. http://doi.org/10.1056/NEJM200005183422001

Keller, M. B., Ryan, N. D., Strober, M., Klein, R. G., Kutcher, S. P., Birmaher, B., … Weller, R. A. (2001). Efficacy of Paroxetine in the Treatment of Adolescent Major Depression: A Randomized, Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(7), 762–772. http://doi.org/10.1097/00004583-200107000-00010

Kemp, A. H., Gordon, E., Rush, A. J., & Williams, L. M. (2008). Improving the Prediction of Treatment Response in Depression: Integration of Clinical, Cognitive, Psychophysiological, Neuroimaging, and Genetic Measures. *CNS Spectrums*, *13*(12), 1066–1086.

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal Relationship Between Stressful Life Events and the Onset of Major Depression. *http://dx.doi.org/10.1176/ajp.156.6.837*.

Kennard, B. D., Silva, S. G., Tonev, S., Rohde, P., Hughes, J. L., Vitiello, B., … March, J. (2009). Remission and Recovery in the Treatment for Adolescents With Depression Study

(TADS): Acute and Long-Term Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(2), 186–195. http://doi.org/10.1097/CHI.0B013E31819176F9

Kennedy, S. H., Evans, K. R., Krüger, S., Mayberg, H. S., Meyer, J. H., McCann, S., … Vaccarino, F. J. (2001). Changes in Regional Brain Glucose Metabolism Measured With Positron Emission Tomography After Paroxetine Treatment of Major Depression. *American Journal of Psychiatry*, *158*(6), 899–905. http://doi.org/10.1176/appi.ajp.158.6.899

Kennedy, S. H., Konarski, J. Z., Segal, Z. V, Lau, M. A., Bieling, P. J., McIntyre, R. S., & Mayberg, H. S. (2007). Differences in Brain Glucose Metabolism Between Responders to CBT and Venlafaxine in a 16-Week Randomized Controlled Trial. *American Journal of Psychiatry*, *164*, 778–788.

Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of Major and Minor Depression in Children and Adolescents: Evaluation of the Mood and Feelings Questionnaire. *Journal of Child Psychology and Psychiatry*, *38*(5), 565–573. http://doi.org/10.1111/j.1469-7610.1997.tb01543.x

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder. *JAMA*, *289*(23), 3095. http://doi.org/10.1001/jama.289.23.3095

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W. A. (2012). A Systematic Review of Comparative Efficacy of Treatments and Controls for Depression. *PLoS ONE*, *7*(7), e41778. http://doi.org/10.1371/journal.pone.0041778

Kiess, W., Meidert, A., Dressendörfer, R. A., Schriever, K., Kessler, U., Köunig, A., … Strasburger, C. J. (1995). Salivary Cortisol Levels throughout Childhood and Adolescence: Relation with Age, Pubertal Stage, and Weight. *Pediatric Research*, *37*(4), 502–506. http://doi.org/10.1203/00006450-199504000-00020

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*, *19*(4), 313–333.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1995). Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. *Psychoneuroendocrinology*, *20*(5), 509–14. http://doi.org/10.1016/0306-4530(94)00078-O

Kirschbaum, C., Schommer, N., Federenko, I., Gaab, J., Neumann, O., Oellers, M., … Hellhammer, D. H. (1996). Short-term estradiol treatment enhances pituitary-adrenal

axis and sympathetic responses to psychosocial stress in healthy young men. *The Journal of Clinical Endocrinology & Metabolism*, *81*(10), 3639–3643. http://doi.org/10.1210/jcem.81.10.8855815

Klimes-Dougan, B., Westlund Schreiner, M., Thai, M., Gunlicks-Stoessel, M., Reigstad, K., & Cullen, K. R. (2018). Neural and neuroendocrine predictors of pharmacological treatment response in adolescents with depression: A preliminary study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *81*, 194–202. http://doi.org/10.1016/J.PNPBP.2017.10.015

Knack, J. M., Jensen-Campbell, L. A., & Baum, A. (2011). Worse than sticks and stones? Bullying is associated with altered HPA axis functioning and poorer health. *Brain and Cognition*, *77*(2), 183–190. http://doi.org/10.1016/J.BANDC.2011.06.011

Knight, D. C., Smith, C. N., Cheng, D. T., Stein, E. A., & Helmstetter, F. J. (2004). Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 317–325. http://doi.org/10.3758/CABN.4.3.317

Koenig, J., Westlund Schreiner, M., Klimes-Dougan, B., Ubani, B., Mueller, B., Kaess, M., & Cullen, K. R. (2018). Brain structural thickness and resting state autonomic function in adolescents with major depression. *Social Cognitive and Affective Neuroscience*, *13*(7), 741–753. http://doi.org/10.1093/scan/nsy046

Konarski, J. Z., Kennedy, S. H., Segal, Z. V, Lau, M. A., Bieling, P. J., McIntyre, R. S., & Mayberg, H. S. (2009). Predictors of nonresponse to cognitive behavioural therapy or venlafaxine using glucose metabolism in major depressive disorder. *Journal of Psychiatry & Neuroscience*, *34*(3), 175–80.

Koolschijn, P. C. M. P., van Haren, N. E. M., Lensvelt-Mulders, G. J. L. M., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: A meta-analysis of magnetic resonance imaging studies. *Human Brain Mapping*, *30*(11), 3719–3735. http://doi.org/10.1002/hbm.20801

Kornstein, S., & Psychiatry, R. S. (2001). Clinical features of treatment-resistant depression. *Journal of Clinical Psychiatry*, *62*, 18–25.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. http://doi.org/10.1037/abn0000258

Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New Dimensions in the Quantitative Classification of Mental Illness. *Archives of General Psychiatry*, *68*(10), 1003–1011. http://doi.org/10.1001/archgenpsychiatry.2011.107

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and Moderators of Treatment Effects in Randomized Clinical Trials. *Archives of General Psychiatry*, *59*(10), 877–883. http://doi.org/10.1001/archpsyc.59.10.877

Kross, E., Davidson, M., Weber, J., & Ochsner, K. (2009). Coping with emotions past: the neural bases of regulating affect associated with negative autobiographical memories. *Biological Psychiatry*, *65*(5), 361–6. http://doi.org/10.1016/j.biopsych.2008.10.019

Kuehner, C. (2003). Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatrica Scandinavica*, *108*(3), 163–174. http://doi.org/10.1034/j.1600-0447.2003.00204.x

Kunz-Ebrecht, S. R., Kirschbaum, C., Marmot, M., & Steptoe, A. (2004). Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology*, *29*(4), 516–528. http://doi.org/10.1016/S0306-4530(03)00072-6

Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., … W, C. (2003). Regionally Localized Thinning of the Cerebral Cortex in Schizophrenia. *Archives of General Psychiatry*, *60*(9), 878. http://doi.org/10.1001/archpsyc.60.9.878

Lai, C.-H. (2013). Gray matter volume in major depressive disorder: A meta-analysis of voxel-based morphometry studies. *Psychiatry Research: Neuroimaging*, *211*, 37–46. http://doi.org/10.1016/j.pscychresns.2012.06.006

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function*, *214*(5–6), 579–591. http://doi.org/10.1007/s00429-010-0251-3

Lane, P. (2008). Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics*, *7*(2), 93–106. http://doi.org/10.1002/pst.267

Larson, R., & Ham, M. (1993). Stress and “storm and stress” in early adolescence: The relationship of negative events with dysphoric affect. *Developmental Psychology*, *29*(1), 130–140. http://doi.org/10.1037/0012-1649.29.1.130

Lee, K. H., & Siegle, G. J. (2012). Common and distinct brain networks underlying explicit emotional evaluation: a meta-analytic study. *Social Cognitive and Affective Neuroscience*, *7*, 521–534. http://doi.org/10.1093/scan/nsp001

Lemogne, C., le Bastard, G., Mayberg, H., Volle, E., Bergouignan, L., Lehéricy, S., … Fossati, P. (2009). In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Social Cognitive and Affective Neuroscience*, *4*(3), 305–12. http://doi.org/10.1093/scan/nsp008

Lener, M. S., & Iosifescu, D. V. (2015). In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Annals of the New York Academy of Sciences*, *1344*(1), 50–65. http://doi.org/10.1111/nyas.12759

Lener, M. S., Kundu, P., Wong, E., Dewilde, K. E., Tang, C. Y., Balchandani, P., & Murrough, J. W. (2016). Cortical abnormalities and association with symptom dimensions across the depressive spectrum. *Journal of Affective Disorders*, *190*, 529–36. http://doi.org/10.1016/j.jad.2015.10.027

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major Depression in Community Adolescents: Age at Onset, Episode Duration, and Time to Recurrence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *33*(6), 809–818. http://doi.org/10.1097/00004583-199407000-00006

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major Depression in Community Adolescents: Age at Onset, Episode Duration, and Time to Recurrence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *33*(6), 809–818. http://doi.org/10.1097/00004583-199407000-00006

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology*, *102*(1), 110–120. http://doi.org/10.1037/0021-843X.102.1.110

Li, C.-T., Lin, C.-P., Chou, K.-H., Chen, I.-Y., Hsieh, J.-C., Wu, C.-L., … Su, T.-P. (2010). Structural and cognitive deficits in remitting and non-remitting recurrent depression: A voxel-based morphometric study. *NeuroImage*, *50*, 347–356. http://doi.org/10.1016/j.neuroimage.2009.11.021

Li, M., Metzger, C. D., Li, W., Safron, A., van Tol, M.-J., Lord, A., … Walter, M. (2014). Dissociation of glutamate and cortical thickness is restricted to regions subserving trait but not state markers in major depressive disorder. *Journal of Affective Disorders*, *169*, 91–100. http://doi.org/10.1016/j.jad.2014.08.001

Liu, C.-H., Ma, X., Wu, X., Zhang, Y., Zhou, F.-C., Li, F., … Wang, C.-Y. (2012). Regional homogeneity of resting-state brain abnormalities in bipolar and unipolar depression. *Progress in Neuropsychopharmacology & Biological Psychiatry*, *41*, 52–59.

http://doi.org/10.1016/j.pnpbp.2012.11.010

Liu, C. H., Jing, B., Ma, X., Xu, P. F., Zhang, Y., Li, F., Wang, Y. P., … Wang C.Y. (2014). Voxel-based morphometry study of the insular cortex in female patients with current and remitted depression. *Neuroscience*, *262*, 190–199.

Liu, F., Guo, W., Yu, D., Gao, Q., Gao, K., Xue, Z., … Kuang, W. (2012). Classification of Different Therapeutic Responses of Major Depressive Disorder with Multivariate Pattern Analysis Method Based on Structural MR Scans. *PLoS ONE*, *7*(7), e40968. http://doi.org/10.1371/journal.pone.0040968

Liu, X., Kakeda, S., Watanabe, K., Yoshimura, R., Abe, O., Ide, S., … Korogi, Y. (2015). Relationship between the cortical thickness and serum cortisol levels in drug-naive, first-episode patients with major depressive disorder: A surface-based morphometric study. *Depression and Anxiety*, *32*, 702–708. http://doi.org/10.1002/da.22401

Liu, Z., Xu, C., Xu, Y., Wang, Y., Zhao, B., Lv, Y., … Du, C. (2010). Decreased regional homogeneity in insula and cerebellum: A resting-state fMRI study in patients with major depression and subjects at high risk for major depression. *Psychiatry Research: Neuroimaging*, *182*, 211–215.

Lopez-Duran, N. L., Kovacs, M., & George, C. J. (2009). Hypothalamic-pituitary-adrenal axis dysregulation in depressed children and adolescents: a meta-analysis. *Psychoneuroendocrinology*, *34*(9), 1272–1283. http://doi.org/10.1016/j.psyneuen.2009.03.016

Lovallo, W. R., Dickensheets, S. L., Myers, D. A., Thomas, T. L., & Nixon, S. J. (2000). Blunted Stress Cortisol Response in Abstinent Alcoholic and Polysubstance-Abusing Men. *Alcoholism: Clinical and Experimental Research*, *24*(5), 651–658. http://doi.org/10.1111/j.1530-0277.2000.tb02036.x

Lovallo, W. R., & Thomas, T. L. (2000). *Stress hormones in psychophysiological research: Emotional, behavioral, and cognitive implications.* Cambridge University Press.

Lui, S., Wu, Q., Qiu, L., Yang, X., Kuang, W., Chan, R. C. K., … Gong, Q. (2011). Resting-State Functional Connectivity in Treatment-Resistant Depression. *American Journal of Psychiatry*, *168*(6), 642–648. http://doi.org/10.1176/appi.ajp.2010.10101419

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. http://doi.org/10.1038/nrn2639

Luz Neto, L. M. da, Vasconcelos, F. M. N. de, Silva, J. E. da, Pinto, T. C. C., Sougey, É. B., &

Ximenes, R. C. C. (2018). Differences in cortisol concentrations in adolescents with eating disorders: a systematic review. *Jornal de Pediatria*, *In Press*. http://doi.org/10.1016/J.JPED.2018.02.007

Ly, J., McGrath, J. J., & Gouin, J.-P. (2015). Poor sleep as a pathophysiological pathway underlying the association between stressful experiences and the diurnal cortisol profile among children and adolescents. *Psychoneuroendocrinology*, *57*, 51–60. http://doi.org/10.1016/j.psyneuen.2015.03.006

MacQueen, G., M., Yucel, K., Taylor, V. H., Macdonald, K., & Joffe, R. (2008). Posterior hippocampal volumes are associated with remission rates in patients with major depressive disorder. *Biological Psychiatry*, *64*, 880–883.

Malhi, G. S., & Byrow, Y. (2016). Is treatment-resistant depression a useful concept? *Evidence Based Mental Health*, *19*(1), 1–3. http://doi.org/10.1136/eb-2015-102299

March, J., Silva, S., Curry, J., Wells, K., Fairbank, J., Burns, B., … Team, T. for A. W. D. S. (TADS). (2009). The Treatment for Adolescents With Depression Study (TADS): Outcomes Over 1 Year of Naturalistic Follow-Up. *American Journal of Psychiatry*, *166*(10), 1141–1149. http://doi.org/10.1176/appi.ajp.2009.08111620

March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., … Treatment for Adolescents With Depression Study (TADS) Team. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*, *292*(7), 807–820. http://doi.org/10.1001/jama.292.7.807

Massat, I., Souery, D., Del-Favero, J., Nothen, M., Blackwood, D., Muir, W., … Mendlewicz, J. (2005). Association between COMT (Val158Met) functional polymorphism and early onset in patients with major depressive disorder in a European multicenter genetic association study. *Molecular Psychiatry*, *10*(6), 598–605. http://doi.org/10.1038/sj.mp.4001615

Mathers, C. D., & Loncar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, *3*(11), e442. http://doi.org/10.1371/journal.pmed.0030442

Mathew, S. J., Coplan, J. D., Goetz, R. R., Feder, A., Greenwald, S., Dahl, R. E., … Weissman, M. M. (2003). Differentiating Depressed Adolescent 24 h Cortisol Secretion in Light of Their Adult Clinical Outcome. *Neuropsychopharmacology*, *28*, 1336–1343. http://doi.org/10.1038/sj.npp.1300184

Mayberg, H., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R., Jerabek, P., … Fox, P. (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *The American Journal of Psychiatry*, *156*(5), 675–682.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., … Kennedy, S. H. (2005). Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron*, *45*(5), 651–660. http://doi.org/10.1016/J.NEURON.2005.02.014

Mazure, C. M. (1998). Life Stressors as Risk Factors in Depression. *Clinical Psychology: Science and Practice*, *5*(3), 291–313. http://doi.org/10.1111/j.1468-2850.1998.tb00151.x

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), 821–829.

McKnight, D. L., Nelson-Gray, R. O., & Barnhill, J. (1992). Dexamethasone suppression test and response to cognitive therapy and antidepressant medication. *Behavior Therapy*, *23*(1), 99–111. http://doi.org/10.1016/S0005-7894(05)80311-9

Meewisse, M.-L., Reitsma, J. B., De Vries, G.-J., Gersons, B. P. R., & Olff, M. (2007). Cortisol and post-traumatic stress disorder in adults. *British Journal of Psychiatry*, *191*(5), 387–392. http://doi.org/10.1192/bjp.bp.106.024877

Melamed, S., Ugarten, U., Shirom, A., Kahana, L., Lerman, Y., & Froom, P. (1999). Chronic burnout, somatic arousal and elevated salivary cortisol levels. *Journal of Psychosomatic Research*, *46*(6), 591–598. http://doi.org/10.1016/S0022-3999(99)00007-0

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506. http://doi.org/10.1016/j.tics.2011.08.003

Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*, *11*(1), 7–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19432384

Mezuk, B., & Kendler, K. S. (2012). Examining variation in depressive symptoms over the life course: a latent class analysis. *Psychological Medicine*, *42*(10), 2037–2046. http://doi.org/10.1017/S003329171200027X

Midgley, N., & Kennedy, E. (2011). Psychodynamic psychotherapy for children and adolescents: a critical review of the evidence base. *Journal of Child Psychotherapy*,

*37*(3), 232–260. http://doi.org/10.1080/0075417X.2011.614738

Miller, C. H., Hamilton, J. P., Sacchet, M. D., & Gotlib, I. H. (2015). Meta-analysis of Functional Neuroimaging of Major Depressive Disorder in Youth. *JAMA Psychiatry*, *72*(10), 1045–1053. http://doi.org/10.1001/jamapsychiatry.2015.1376

Mohanty, A., Engels, A. S., Herrington, J. D., Heller, W., Ringo Ho, M.-H., Banich, M. T., … Miller, G. A. (2007). Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. *Psychophysiology*, *44*, 343–351. http://doi.org/10.1111/j.1469-8986.2007.00515.x

Mokhtari, M., Arfken, C., & Boutros, N. (2013). The DEX/CRH test for major depression: A potentially useful diagnostic test. *Psychiatry Research*, *208*(2), 131–139. http://doi.org/10.1016/J.PSYCHRES.2012.09.032

Moncrieff, J. (2018). What does the latest meta-analysis really tell us about antidepressants? *Epidemiology and Psychiatric Sciences*, *27*, 430–432. http://doi.org/10.1017/S2045796018000240

Moncrieff, J., & Kirsch, I. (2005). Efficacy of antidepressants in adults. *BMJ*, *331*, 155–157.

Moncrieff, J., & Kirsch, I. (2015). Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemporary Clinical Trials*, *43*, 60–62.

Montgomery, S. A. (1994). Clinically relevant effect sizes in depression. *European Neruopharmacology*, *4*, 283–284.

Montgomery, S. A., & Asberg, M. (1979). A New Depression Scale Designed to be Sensitive to Change. *The British Journal of Psychiatry*, *134*(9), 382–389.

Mueller, T., Leon, A., Keller, M., Solomon, D., Endicott, J., Coryell, W., … Maser, J. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *American Journal of Psychiatry*, *156*(7), 1000–1006.

Mulder, R. T., Joyce, P. R., & Frampton, C. (2003). Relationships among measures of treatment outcome in depressed patients. *Journal of Affective Disorders*, *76*(1–3), 127–135. http://doi.org/10.1016/S0165-0327(02)00080-0

Murray, C. J., Richards, M. A., Newton, J. N., Fenton, K. A., Anderson, H. R., Atkinson, C., … Davis, A. (2013). UK health performance: findings of the Global Burden of Disease Study 2010. *The Lancet*, *381*(9871), 997–1020. http://doi.org/10.1016/S0140-6736(13)60355-4

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide*. (Muthén & Muthén, Ed.) (8th

ed.). Los Angeles, CA. Retrieved from https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

Nemeroff, C. B., Krishnan, K. R. R., Reed, D., Leder, R., Beam, C. A., & Dunnick, N. R. (1992). Adrenal gland enlargement in major depression; A computer tomographic study. *Archives of General Psychiatry*, *49*, 384–387. http://doi.org/10.1001/archpsyc.1992.01820050048008

Nemeroff, C. B., Owens, M. J., Bissette, G., Andorn, A. C., & Stanley, M. (1988). Reduced Corticotropin Releasing Factor Binding Sites in the Frontal Cortex of Suicide Victims. *Archives of General Psychiatry*, *45*(6), 577–579. http://doi.org/10.1001/archpsyc.1988.01800300075009

Niarchou, M., Zammit, S., & Lewis, G. (2015). The Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort as a resource for studying psychopathology in childhood and adolescence: a summary of findings for depression and psychosis. *Social Psychiatry and Psychiatric Epidemiology*, *50*(7), 1017–1027. http://doi.org/10.1007/s00127-015-1072-8

NICE. (2015). *Depression in children and young people: identification and management | 1-Recommendations | Guidance and guidelines | NICE*.

NICE. (2016). *Depression in adults: recognition and management | Guidance and guidelines | NICE*.

Nicolson, N. (2008). *Measurement of cortisol*. *Handbook of physiological research methods*.

Nierenberg, A. A., & DeCecco, L. M. (2001). Definitions of Antidepressant Treatment Response, Remission, Nonresponse, Partial Response, and Other Relevant Outcomes: A Focus on Treatment-Resistant Depression. *The Journal of Clinical Psychiatry*, *62*(suppl 16), 5–9.

Nierenberg, A. A., Husain, M. M., Trivedi, M. H., Fava, M., Warden, D., Wisniewski, S. R., … Rush, A. J. (2010). Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: a STAR*D report. *Psychological Medicine*, *40*(1), 41–50. http://doi.org/10.1017/S0033291709006011

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, *100*(4), 569–582. http://doi.org/10.1037/0021-843X.100.4.569

Nouretdinov, I., Costafreda, S. G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., & Fu, C. H. Y. (2011). Machine learning classification with confidence: Application of

transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. *NeuroImage*, *56*(2), 809–813. http://doi.org/10.1016/J.NEUROIMAGE.2010.05.023

Ntali, G., Grossman, A., & Karavitaki, N. (2015). Clinical and biochemical manifestations of Cushing's. *Pituitary*, *18*(2), 181–187. http://doi.org/10.1007/s11102-014-0631-4

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. http://doi.org/10.1080/10705510701575396

O'Keeffe, S., Martin, P., Goodyer, I. M., Wilkinson, P., Consortium, I., & Midgley, N. (2018). Predicting dropout in adolescents receiving therapy for depression. *Psychotherapy Research*, *28*(5), 708–721. http://doi.org/10.1080/10503307.2017.1393576

O'Toole, S. M., Sekula, L. K., & Rubin, R. T. (1997). Pituitary-adrenal cortical axis measures as predictors of sustained remission in major depression. *Biological Psychiatry*, *42*(2), 85–89. http://doi.org/10.1016/S0006-3223(96)00293-4

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D. E., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, *23*(2), 483–99. http://doi.org/10.1016/j.neuroimage.2004.06.030

Ongur, D., Drevets, W. C., & Price, J. L. (1998). Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proceedings of the National Academy of Sciences*, *95*(22), 13290–13295. http://doi.org/10.1073/pnas.95.22.13290

Oquendo, M. A., Echavarria, G., Galfalvy, H. C., Grunebaum, M. F., Burke, A., Barrera, A., … John Mann, J. (2003). Lower Cortisol Levels in Depressed Patients with Comorbid Post-Traumatic Stress Disorder. *Neuropsychopharmacology*, *28*(3), 591–598. http://doi.org/10.1038/sj.npp.1300050

Owens, M., Herbert, J., Jones, P. B., Sahakian, B. J., Wilkinson, P. O., Dunn, V. J., … Goodyer, I. M. (2014). Elevated morning cortisol is a stratified population-level biomarker for major depression in boys only with high depressive symptoms. *Proceedings of the National Academy of Sciences*, *111*(9), 3638–3643. http://doi.org/10.1073/pnas.1318786111

Paillè Re Martinot, M.-L., Martinot, J.-L., Ringuenet, D., Galinowski, A., Gallarda, T., Bellivier, F., … Artiges, E. (2011). Baseline Brain Metabolism in Resistant Depression and Response to Transcranial Magnetic Stimulation. *Neuropsychopharmacology*, *36*, 2710–

2719. http://doi.org/10.1038/npp.2011.161

Pancheri, P., Picardi, A., Pasquini, M., Gaetano, P., & Biondi, M. (2002). Psychopathological dimensions of depression: a factor study of the 17-item Hamilton depression rating scale in unipolar depressed outpatients. *Journal of Affective Disorders*, *68*(1), 41–47. http://doi.org/10.1016/S0165-0327(00)00328-1

Pariante, C. M., & Lightman, S. L. (2008). The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences*, *31*(9), 464–468. http://doi.org/10.1016/j.tins.2008.06.006

Pariante, C. M., Makoff, A., Lovestone, S., Feroli, S., Heyden, A., Miller, A. H., & Kerwin, R. W. (2001). Antidepressants enhance glucocorticoid receptor function *in vitro* by modulating the membrane steroid transporters. *British Journal of Pharmacology*, *134*(6), 1335–1343. http://doi.org/10.1038/sj.bjp.0704368

Pariante, C. M., & Miller, A. H. (2001). Glucocorticoid receptors in major depression: relevance to pathophysiology and treatment. *Biological Psychiatry*, *49*(5), 391–404. http://doi.org/10.1016/S0006-3223(00)01088-X

Park, M., Cuijpers, P., van Straten, A., & Reynolds, C. F. (2014). The Effects of Psychotherapy for Adult Depression on Social Support: A Meta-Analysis. *Cognitive Therapy and Research*, *38*(6), 600–611. http://doi.org/10.1007/s10608-014-9630-z

Paykel, E. S., Ramana, R., Cooper, Z., Hayhurst, H., Kerr, J., & Barocka, A. (1995). Residual symptoms after partial remission: an important outcome in depression. *Psychological Medicine*, *25*(6), 1171–80.

Perez, J., Russo, D. A., Stochl, J., Clarke, J., Martin, Z., Jassi, C., … Jones, P. B. (2018). Common mental disorder including psychotic experiences: Trailblazing a new recovery pathway within the Improving Access to Psychological Therapies programme in England. *Early Intervention in Psychiatry*, *12*(3), 497–504. http://doi.org/10.1111/eip.12434

Perlis, R. H., Uher, R., Ostacher, M., Goldberg, J. F., Trivedi, M. H., Rush, A. J., & Fava, M. (2010). Association Between Bipolar Spectrum Features and Treatment Outcomes in Outpatients With Major Depressive Disorder. *Archives of General Psychiatry*, *68*(4), 351–360. http://doi.org/10.1001/archgenpsychiatry.2010.179

Perugi, G., Medda, P., Zanello, S., Toni, C., & Cassano, G. B. (2012). Episode length and mixed features as predictors of ECT nonresponse in patients with medication-resistant major depression. *Brain Stimulation*, *5*(1), 18–24. http://doi.org/10.1016/j.brs.2011.02.003

Peterson, B. S., Warner, V., Bansal, R., Zhu, H., Hao, X., Liu, J., … Weissman, M. M. (2009).

Cortical thinning in persons at increased familial risk for major depression. *Proceedings of the National Academy of Sciences*, *106*(15), 6273–6278.

Phillips, M. L., Chase, H. W., Sheline, Y. I., Etkin, A., Almeida, J. R. C., Deckersbach, T., & Trivedi, M. H. (2015). Identifying Predictors, Moderators, and Mediators of Antidepressant Response in Major Depressive Disorder: Neuroimaging Approaches. *American Journal of Psychiatry*, *172*(2), 124–138. http://doi.org/10.1176/appi.ajp.2014.14010076

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression: Critical review. *British Journal of Psychiatry*, *177*(6), 486–492. http://doi.org/10.1192/bjp.177.6.486

Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and Effectiveness of Antidepressants: Current Status of Research. *Psychotherapy and Psychosomatics*, *79*, 267–279. http://doi.org/10.1159/000318293

Pizzagalli, D. A. (2011). Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology*, *36*(1), 183–206. http://doi.org/10.1038/npp.2010.166

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., … Mann, J. J. (2011). The Columbia–Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults. *American Journal of Psychiatry*, *168*(12), 1266–1277. http://doi.org/10.1176/appi.ajp.2011.10111704

Protopopescu, X., Pan, H., Tuescher, O., Cloitre, M., Goldstein, M., Engelien, A., … Silbersweig, D. (2006). Increased brainstem volume in panic disorder: a voxel-based morphometric study. *NeuroReport*, *17*(4), 361–363. http://doi.org/10.1097/01.wnr.0000203354.80438.1

Pruessner, J. C., Gaab, J., Hellhammer, D. H., Lintz, D., Schommer, N., & Kirschbaum, C. (1997). Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. *Psychoneuroendocrinology*, *22*(8), 615–625. http://doi.org/10.1016/S0306-4530(97)00072-3

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., von Auer, K., Jobst, S., … Kirschbaum, C. (1997). Free Cortisol Levels after Awakening: A Reliable Biological Marker for the Assessment of Adrenocortical Activity. *Life Sciences*, *61*(26), 2539–2549. http://doi.org/10.1016/S0024-3205(97)01008-4

Pruessner, M., Hellhammer, D. H., Pruessner, J. C., & Lupien, S. J. (2003). Self-Reported

Depressive Symptoms and Stress Levels in Healthy Young Men: Associations With the Cortisol Response to Awakening. *Psychosomatic Medicine*, *65*(1), 92–99. http://doi.org/10.1097/01.PSY.0000040950.22044.10

Putnam, K. M., Pizzagalli, D. A., Gooding, D. C., Kalin, N. H., & Davidson, R. J. (2008). Neural activity and diurnal variation of cortisol: evidence from brain electrical tomography analysis and relevance to anhedonia. *Psychophysiology*, *45*(6), 886–95. http://doi.org/10.1111/j.1469-8986.2008.00697.x

Pyne, J. M., Patterson, T. L., Kaplan, R. M., Gillin, J. C., Koch, W. L., & Grant, I. (1997). Assessment of the quality of life of patients with major depression. *Psychiatric Services*, *48*(2), 224–230. http://doi.org/10.1176/ps.48.2.224

R: The R Project for Statistical Computing. (n.d.). Retrieved April 14, 2019, from https://www.r-project.org/

Raadsheer, F. C., Hoogendijk, W. J., Stam, F. C., Tilders, F. J., & Swaab, D. F. (1994). Increased numbers of corticotropin-releasing hormone expressing neurons in the hypothalamic paraventricular nucleus of depressed patients. *Neuroendocrinology*, *60*(4), 436–444. http://doi.org/10.1159/000126778

Raison, C. L., & Miller, A. H. (2003). When Not Enough Is Too Much: The Role of Insufficient Glucocorticoid Signaling in the Pathophysiology of Stress-Related Disorders. *American Journal of Psychiatry*, *160*(9), 1554–1565. http://doi.org/10.1176/appi.ajp.160.9.1554

Rajkowska, G. (2000). Postmortem studies in mood disorders indicate altered numbers of neurons and glial cells. *Biological Psychiatry*, *48*(8), 766–777. http://doi.org/10.1016/S0006-3223(00)00950-1

Rajkowska, G., Miguel-Hidalgo, J. J., Wei, J., Dilley, G., Pittman, S. D., Meltzer, H. Y., … Stockmeier, C. A. (1999). Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biological Psychiatry*, *45*(9), 1085–1098. http://doi.org/10.1016/S0006-3223(99)00041-4

Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in Neurosciences*, *18*(9), 383–388.

Ram, N., & Grimm, K. J. (2009). Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *International Journal of Behavioral Development*, *33*(6), 565–576. http://doi.org/10.1177/0165025409343765

Rao, U., Hammen, C., Ortiz, L. R., Chen, L.-A., & Poland, R. E. (2008). Effects of Early and Recent Adverse Experiences on Adrenal Response to Psychosocial Stress in Depressed

Adolescents. *Biological Psychiatry*, *64*(6), 521–526. http://doi.org/10.1016/j.biopsych.2008.05.012

Resmini, E., Santos, A., & Webb, S. M. (2016). Cortisol Excess and the Brain. *Frontiers of Hormone Research*, *46*, 74–86. http://doi.org/10.1159/000443868

Reul, J. M. H. M., & De KLoet, E. R. (1985). Two Receptor Systems for Corticosterone in Rat Brain: Microdistribution and Differential Occupation. *Endocrinology*, *117*(6), 2505–2511. http://doi.org/10.1210/endo-117-6-2505

Reynolds, C. R., & Richmond, B. O. (1978). What i think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, *6*(2), 271–280. http://doi.org/10.1007/BF00919131

Reynolds, S., Carrey, N., Jaworska, N., Langevin, L. M., Yang, X.-R., & MacMaster, F. P. (2014). Cortical thickness in youth with major depressive disorder. *BMC Psychiatry*, *14*(1), 83. http://doi.org/10.1186/1471-244X-14-83

Rhebergen, D., Lamers, F., Spijker, J., De Graaf, R., Beekman, A. T. F., & Penninx, B. W. J. H. (2012). Course trajectories of unipolar depressive disorders identified by latent class growth analysis. *Psychological Medicine*, *42*, 1383–1396. http://doi.org/10.1017/S0033291711002509

Ritchey, M., Dolcos, F., Eddington, K. M., Strauman, T. J., & Cabeza, R. (2011). Neural correlates of emotional processing in depression: Changes with cognitive behavioral therapy and predictors of treatment response. *Journal of Psychiatric Research*, *45*(5), 577–587. http://doi.org/10.1016/j.jpsychires.2010.09.007

Rizvi, S. J., Salomons, T. V., Konarski, J. Z., Downar, J., Giacobbe, P., McIntyre, R. S., & Kennedy, S. H. (2013). Neural response to emotional stimuli associated with successful antidepressant treatment and behavioral activation. *Journal of Affective Disorders*, *151*, 573–581.

Robbins, D. R., Alessi, N. E., & Colfer, M. V. (1989). Treatment of adolescents with major depression: implications of the DST and the melancholic clinical subtype. *Journal of Affective Disorders*, *17*(2), 99–104. http://doi.org/10.1016/0165-0327(89)90031-1

Rodenbeck, A., Huether, G., Rüther, E., & Hajak, G. (2002). Interactions between evening and nocturnal cortisol secretion and sleep parameters in patients with severe chronic primary insomnia. *Neuroscience Letters*, *324*(2), 159–163. http://doi.org/10.1016/S0304-3940(02)00192-1

Roelfsema, F., van den Berg, G., Frölich, M., Veldhuis, J. D., van Eijk, A., Buurman, M. M., &

Etman, B. H. (1993). Sex-dependent alteration in cortisol response to endogenous adrenocorticotropin. *The Journal of Clinical Endocrinology & Metabolism*, *77*(1), 234–240. http://doi.org/10.1210/jcem.77.1.8392084

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1991). Comorbidity of unipolar depression: II. Comorbidity with other mental disorders in adolescents and adults. *Journal of Abnormal Psychology*, *100*(2), 214–222. http://doi.org/10.1037/0021-843X.100.2.214

Rohleder, N., Wolf, J. M., & Wolf, O. T. (2010). Glucocorticoid sensitivity of cognitive and inflammatory processes in depression and posttraumatic stress disorder. *Neuroscience and Biobehavioural Reviews*, *35*, 104–114. http://doi.org/10.1016/j.neubiorev.2009.12.003

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

Rubin, R. T., Phillips, J. J., Sadow, T. F., & McCracken, J. T. (1995). Adrenal Gland Volume in Major Depression. *Archives of General Psychiatry*, *52*(3), 213–218. http://doi.org/10.1001/archpsyc.1995.03950150045009

Rudolph, K. D., & Hammen, C. (1999). Age and Gender as Determinants of Stress Exposure, Generation, and Reactions in Youngsters: A Transactional Perspective. *Child Development*, *70*(3), 660–677. http://doi.org/10.1111/1467-8624.00048

Ruhé, H. G., Booij, J., Veltman, D. J., Michel, M. C., & Schene, A. H. (2008). *Successful Pharmacological Treatment of Major Depression Disorder Attenuates Amygdala Activation to Negative Facial Expressions. An fMRI Study*.

Ruhé, H. G., Khoenkhoen, S. J., Ottenhof, K. W., Koeter, M. W., Mocking, R. J. T., & Schene, A. H. (2015). Longitudinal effects of the SSRI paroxetine on salivary cortisol in Major Depressive Disorder. *Psychoneuroendocrinology*, *52*, 261–271. http://doi.org/10.1016/J.PSYNEUEN.2014.10.024

Rupprecht, R., Kornhuber ’, J., Wodarz, N., Lugauer, J., Gijbel ’, C., Riederer ’, P., & Beckmann ’ ’, H. (1991). Lymphocyte glucocorticoid receptor binding during depression and after clinical recovery. *Journal of Affective Disorders*, *22*, 31–35.

Rush, A. J., Kraemer, H. C., Sackeim, H. A., Fava, M., Trivedi, M. H., Frank, E., … ACNP Task Force. (2006). Report by the ACNP Task Force on Response and Remission in Major Depressive Disorder. *Neuropsychopharmacology*, *31*(9), 1841–1853. http://doi.org/10.1038/sj.npp.1301131

Rush, J., Giles, D. E., Schlesser, M. A., Orsulak, P. J., Richard Parker, C., Weissenburger, J. E., … Murphy, A. L. (1996). The Dexamethasone Suppression Test in Patients With Mood Disorders. *J Clin Psychiatry*, *571057*, 470–484.

Rush, J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Maurizio Fava, M. (2006). Acute and Longer-Term Outcomes in Depressed Outpatients Requiring One or Several Treatment Steps: A STAR*D Report. *Am J Psychiatry*, *163*(11), 1905–1917.

Russell, E., Koren, G., Rieder, M., & van Uum, S. (2012). Hair cortisol as a biological marker of chronic stress: Current status, future directions and unanswered questions. *Psychoneuroendocrinology*, *37*, 589–601.

Saarni, S. I., Suvisaari, J., Sintonen, H., Pirkola, S., Koskinen, S., Aromaa, A., & Lönnqvist, J. (2007). Impact of psychiatric disorders on health-related quality of life: general population survey. *British Journal of Psychiatry*, *190*(4), 326–332. http://doi.org/10.1192/bjp.bp.106.025106

Sallee, F., Nesbitt, L., Dougherty, D., & Hilal, R. (1995). Lymphocyte glucocorticoid receptor: Predictor of sertraline response in adolescent major depressive disorder (MDD). *Psychopharmacology Bulletin*, *31*(2), 339–345.

Salvadore, G., Nugent, A. C., Lemaitre, H., Luckenbaugh, D. A., Tinsley, R., Cannon, D. M., … Drevets, W. C. (2011). Prefrontal cortical abnormalities in currently depressed versus currently remitted patients with major depressive disorder. *NeuroImage*, *54*(2643–2651).

Sanford, M., Szatmari, P., Spinner, M., Munroe-Blum, H., Jamieson, E., Walsh, C., & Jones, D. (1995). Predicting the One-Year Course of Adolescent Major Depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34*(12), 1618–1628.

Saravanan, C., & Wilks, R. (2014). Medical students' experience of and reaction to stress: the role of depression and anxiety. *The Scientific World Journal*, *2014*, 1–8. http://doi.org/10.1155/2014/737382

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, *33*(4), 545–571. http://doi.org/10.1207/s15327906mbr3304_5

Schlosser, N., Wolf, O. T., & Wingenfeld, K. (2011). Cognitive correlates of hypothalamic–pituitary–adrenal axis in major depression. *Expert Review of Endocrinology & Metabolism*, *6*(1), 109–126. http://doi.org/10.1586/eem.10.79

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., … Veltman, D. J. (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular Psychiatry*, *22*, 900–909. http://doi.org/10.1038/mp.2016.60

Schmaal, L., Veltman, D. J., van Erp, T. G. M., Sämann, P. G., Frodl, T., Jahanshad, N., … Hibar, D. P. (2016). Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry*, *21*(6), 806–812. http://doi.org/10.1038/mp.2015.69

Schmaal, L., Yücel, M., Ellis, R., Vijayakumar, N., Simmons, J. G., Allen, N. B., & Whittle, S. (2017). Brain Structural Signatures of Adolescent Depressive Symptom Trajectories: A Longitudinal Magnetic Resonance Imaging Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(7), 593–601.e9. http://doi.org/10.1016/j.jaac.2017.05.008

Schmidt, U., Laessle, R., & Hellhammer, D. (2013). Major depression in young girls is related to altered cortisol awakening response. *European Child & Adolescent Psychiatry*, *22*(6), 379–384. http://doi.org/10.1007/s00787-012-0371-9

Schüle, C. (2007). Neuroendocrinological Mechanisms of Actions of Antidepressant Drugs. *Journal of Neuroendocrinology*, *19*(3), 213–226. http://doi.org/10.1111/j.1365-2826.2006.01516.x

Schüle, C., Baghai, T. C., Eser, D., & Rupprecht, R. (2009). Hypothalamic–pituitary–adrenocortical system dysregulation and new treatment strategies in depression. *Expert Review of Neurotherapeutics*, *9*(7), 1005–1019. http://doi.org/10.1586/ern.09.52

Scott, K., Lewis, C. C., & Marti, C. N. (2019). Trajectories of Symptom Change in the Treatment for Adolescents With Depression Study. *Journal of the American Academy of Child & Adolescent Psychiatry*. http://doi.org/10.1016/J.JAAC.2018.07.908

Seeley, W., Menon, V., Schatzberg, A., Keller, J., Glover, G., Kenna, H., … Greicius, M. (2007). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, *27*(9), 2349–2356. http://doi.org/20024039

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *The Journal of Neruoscience*, *27*(9), 2349–2356. http://doi.org/10.1523/JNEUROSCI.5587-06.2007

Seifter, J., Sloane, D., & Ratner, A. (2005). *Concepts in medical physiology*. Lippincott Williams & Wilkins.

Sharma, T., Guski, L. S., Freund, N., & Gøtzsche, P. C. (2016). Suicidality and aggression during antidepressant treatment: systematic review and meta-analyses based on clinical study reports. *BMJ*, *352*, i65. http://doi.org/10.1136/bmj.i65

Sharp, C., Goodyer, I. M., & Croudace, T. J. (2006). The Short Mood and Feelings Questionnaire (SMFQ): A Unidimensional Item Response Theory and Categorical Data Factor Analysis of Self-Report Ratings from a Community Sample of 7-through 11-Year-Old Children. *Journal of Abnormal Child Psychology*, *34*(3), 365–377. http://doi.org/10.1007/s10802-006-9027-x

Shelton, R. C., Prakash, A., Mallinckrodt, C. H., Wohlreich, M. M., Raskin, J., Robinson, M. J., & Detke, M. J. (2007). Patterns of depressive symptom response in duloxetine-treated outpatients with mild, moderate or more severe depression. *International Journal of Clinical Practice*, *61*(8), 1337–1348. http://doi.org/10.1111/j.1742-1241.2007.01444.x

Shelton, R. C., Williamson, D. J., Corya, S. A., Sanger, T. M., Van Campen, L. E., Case, M., … Tollefson, G. D. (2005). Olanzapine/Fluoxetine for Treatment Resistant MDD Olanzapine/Fluoxetine Combination for Treatment-Resistant Depression: A Controlled Study of SSRI and Nortriptyline Resistance. *J Clin Psychiatry*, *66*(10), 1289–1297.

Sher, L. (2004). Daily hassles, cortisol, and the pathogenesis of depression. *Medical Hypotheses*, *62*(2), 198–202. http://doi.org/10.1016/S0306-9877(03)00320-7

Shin, L. M., Davis, F., VanElzakker, M. B., Dahlgren, M. K., & Dubois, S. J. (2013). Neuroimaging predictors of treatment response in anxiety disorders. *Biology of Mood & Anxiety Disorders*, *3*(1), 15. http://doi.org/10.1186/2045-5380-3-15

Shirk, S. R., Deprince, A. P., Crisostomo, P. S., & Labus, J. (2013). Cognitive Behavioral Therapy for Depressed Adolescents Exposed to Interpersonal Trauma: An Initial Effectiveness Trial. *Psychotherapy*, *51*(1), 167–179. http://doi.org/10.1037/a0034845

Siegle, G. J., Carter, C. S., & Thase, M. E. (2006). Use of fMRI to Predict Recovery From Unipolar Depression With Cognitive Behavior Therapy. *Am J Psychiatry*, *163*(4), 735–737.

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., … Pell, J. P. (2013). Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. *PLoS ONE*, *8*(11), e75362. http://doi.org/10.1371/journal.pone.0075362

Smyth, J., Ockenfels, M. C., Porter, L., Kirschbaum, C., Hellhammer, D. H., & Stone, A. A. (1998). Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. *Psychoneuroendocrinology*, *23*(4), 353–370. http://doi.org/10.1016/S0306-4530(98)00008-0

Soriano-Mas, C., Hernández-Ribas, R., Pujol, J., Urretavizcaya, M., Deus, J., Harrison, B. J., … Cardoner, N. (2011). Cross-Sectional and Longitudinal Assessment of Structural Brain Alterations in Melancholic Depression. *Biological Psychiatry*, *69*(4), 318–325. http://doi.org/10.1016/J.BIOPSYCH.2010.07.029

Sotsky, S. M., Glass, D. R., Shea, M. T., Pilkonis, P. A., Collins, F., Elkin, I., … Oliveri, M. E. (1991). Patient Predictors of Response to Psychotherapy and Pharmacotherapy: Findings in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, *148*(8), 997–1008. http://doi.org/10.1176/foc.4.2.278

Souery, D., Oswald, P., Massat, I., Bailer, U., Bollen, J., Demyttenaere, K., … Mendlewicz, J. (2007). Clinical Factors Associated With Treatment Resistance in Major Depressive Disorder: Results from a European Multicenter Study. *Journal of Clinical Psychiatry*, *68*, 1062–1070.

Späti, J., Hänggi, J., Doerig, N., Ernst, J., Sambataro, F., Brakowski, J., … Spinelli, S. (2015). Prefrontal thinning affects functional connectivity and regional homogeneity of the anterior cingulate cortex in depression. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *40*(7), 1640–8. http://doi.org/10.1038/npp.2015.8

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioural Reviews*, *24*, 417–463.

Sramek, J. J., Murphy, M. F., & Cutler, N. R. (2016). Sex differences in the psychopharmacological treatment of depression. *Dialogues in Clinical Neruoscience*, *18*(4), 447–457.

Staufenbiel, S. M., Penninx, B. W. J. H., Spijker, A. T., Elzinga, B. M., & Van Rossum, E. F. C. (2012). Hair cortisol, stress exposure, and mental health in humans: A systematic review. *Psychoneuroendocrinology*, *38*(8), 1220–1235. http://doi.org/10.1016/j.psyneuen.2012.11.015

Steiner, J., Walter, M., Gos, T., Guillemin, G. J., Bernstein, H.-G., Sarnyai, Z., … Myint, A.-M. (2011). Severe depression is associated with increased microglial quinolinic acid in subregions of the anterior cingulate gyrus: Evidence for an immune-modulated

glutamatergic neurotransmission? *Journal of Neuroinflammation*, *8*(1), 94. http://doi.org/10.1186/1742-2094-8-94

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

Stetler, C., & Miller, G. (2005). Blunted cortisol response to awakening in mild to moderate depression: regulatory influences of sleep patterns and social contacts. *Journal of Abnormal Psychology*, *114*(4), 697–705.

Stetler, C., & Miller, G. E. (2011). Depression and Hypothalamic-Pituitary-Adrenal Activation: A Quantitative Summary of Four Decades of Research. *Psychosomatic Medicine*, *73*(2), 114–126. http://doi.org/10.1097/PSY.0b013e31820ad12b

Steudte-Schmiedgen, S., Wichmann, S., Stalder, T., Hilbert, K., Muehlhan, M., Lueken, U., & Beesdo-Baum, K. (2017). Hair cortisol concentrations and cortisol stress reactivity in generalized anxiety disorder, major depression and their comorbidity. *Journal of Psychiatric Research*, *84*, 184–190. http://doi.org/10.1016/J.JPSYCHIRES.2016.09.024

Stevens, J. (James P. (2009). *Applied multivariate statistics for the social sciences*. Routledge.

Stochl, J., Khandaker, G. M., Lewis, G., Perez, J., Goodyer, I. M., Zammit, S., … Jones, P. B. (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychological Medicine*, *45*(7), 1483–1493. http://doi.org/10.1017/S003329171400261X

Streit, F., Memic, A., Hasandedić, L., Rietschel, L., Frank, J., Lang, M., … Rietschel, M. (2016). Perceived stress and hair cortisol: Differences in bipolar disorder and schizophrenia. *Psychoneuroendocrinology*, *69*, 26–34. http://doi.org/10.1016/J.PSYNEUEN.2016.03.010

Stulz, N., Thase, M. E., Klein, D. N., Manber, R., & Crits-Christoph, P. (2010). Differential effects of treatments for chronic depression: A latent growth model reanalysis. *Journal of Consulting and Clinical Psychology*, *78*(3), 409–419.

Sundquist, K., Frank, G., & Sundquist, J. (2004). Urbanisation and incidence of psychosis and depression: follow-up study of 4.4 million women and men in Sweden. *The British Journal of Psychiatry*, *184*(4), 293–8. http://doi.org/10.1192/bjp.184.4.293

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. *The Journal of Clinical Psychiatry*,

*70*(3), 344–353. http://doi.org/10.4088/JCP.07m03780

Takahashi, T., Yücel, M., Lorenzetti, V., Tanino, R., Whittle, S., Suzuki, M., … Allen, N. B. (2010). Volumetric MRI study of the insular cortex in individuals with current and past major depression. *Journal of Affective Disorders*, *121*, 231–238. http://doi.org/10.1016/j.jad.2009.06.003

Talairach, J. (Jean), & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging*. G. Thieme.

Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., … Mills, K. L. (2017). Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *The Journal of Neuroscience*, *37*(12), 3402–3412. http://doi.org/10.1523/JNEUROSCI.3302-16.2017

Tarpey, T., & Petkova, E. (2010). Latent Regression Analysis. *Statistical Modelling*, *10*(2), 133–158. http://doi.org/10.1177/1471082X0801000202

Tedlow, J., Fava, M., Uebelacker, L., Nierenberg, A. A., Alpert, J. E., & Rosenbaum, J. (1998). Outcome Definitions and Predictors in Depression. *Psychotherapy and Psychosomatics*, *67*(4–5), 266–270. http://doi.org/10.1159/000012290

Tedlow, J., Fava, M., Uebelacker, L., Nierenberg, A. A., Alpert, J. E., & Rosenbaum, J. (1998). Outcome definitions and predictors in depression. *Psychotherapy and Psychosomatics*, *67*(4–5), 266–270. http://doi.org/10.1159/000012290

Thase, M. E., Dubé, S., Bowler, K., Howland, R. H., Myers, J. E., Friedman, E., & Jarrett, D. B. (1996). Hypothalamic-pituitary-adrenocortical activity and response to cognitive behavior therapy in unmedicated, hospitalized depressed patients. *The American Journal of Psychiatry*, *153*(7), 886–891. http://doi.org/10.1176/ajp.153.7.886

Thase, M. E., Simons, A. D., & Reynolds, C. F. (1993). Psychobiological correlates of poor response to cognitive behavior therapy: Potential indications for antidepressant pharmacotherapy. *Psychopharmacology Bulletin*, *29*(2), 293–301.

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

Thomas, C. M., & Morris, S. (2003). Cost of depression among adults in England in 2000. *The British Journal of Psychiatry*, *183*(6), 514–519. http://doi.org/10.1192/00-000

Thomas, C. M., & Morris, S. (2003). Cost of depression among adults in England in 2000. *The British Journal of Psychiatry*, *183*(6), 514–519.

Thomas, L., Kessler, D., Campbell, J., Morrison, J., Peters, T. J., Williams, C., … Wiles, N. (2013). Prevalence of treatment-resistant depression in primary care: cross-sectional data. *British Journal of General Practice*, 852–858. http://doi.org/10.3399/bjgp13X675430

Tiffin, P. A., & Rolling, K. (2012). Structure of the Health of the Nation Outcome Scales for Children and Adolescents: An ordinal factor analysis of clinician ratings of a sample of young people referred to community mental health services. *Psychiatry Research*, *197*(1–2), 154–162. http://doi.org/10.1016/J.PSYCHRES.2012.02.007

Topic 6: Mplus Short Course Videos and Handouts. (n.d.). Retrieved April 14, 2019, from https://www.statmodel.com/course_materials.shtml

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V, Kurian, B. T., Phillips, M. L., … Weissman, M. M. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research*, *78*, 11–23. http://doi.org/10.1016/j.jpsychires.2016.03.001

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., … Team, S. S. (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. http://doi.org/10.1176/appi.ajp.163.1.28

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. *New England Journal of Medicine*, *358*(3), 252–260. http://doi.org/10.1056/NEJMsa065779

Tyrer, P. (2014). Comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment*, *20*(4), 280–285. http://doi.org/10.1192/apt.bp.113.011296

Uchida, R. R., Del-Ben, C. M., Busatto, G. F., Duran, F. L. S., Guimarães, F. S., Crippa, J. A. S., … Graeff, F. G. (2008). Regional gray matter abnormalities in panic disorder: A voxel-based morphometry study. *Psychiatry Research: Neuroimaging*, *163*, 21–29.

Uhart, M., Chong, R. Y., Oswald, L., Lin, P.-I., & Wand, G. S. (2006). Gender differences in hypothalamic–pituitary–adrenal (HPA) axis reactivity. *Psychoneuroendocrinology*, *31*(5), 642–652. http://doi.org/10.1016/J.PSYNEUEN.2006.02.003

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., … McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants.

*Psychological Medicine*, *40*(8), 1367–1377. http://doi.org/10.1017/S0033291709991528

Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., & Murray, C. J. L. (2004). Global burden of depressive disorders in the year 2000. *British Journal of Psychiatry*, *184*(5), 386–392. http://doi.org/10.1192/bjp.184.5.386

Valentino, R. J., & Commons, K. G. (2005). Peptides that fine-tune the serotonin system. *Neuropeptides*, *39*(1), 1–8. http://doi.org/10.1016/J.NPEP.2004.09.005

Van den Bergh, B. R. H., & Van Calster, B. (2009). Diurnal cortisol profiles and evening cortisol in post-pubertal adolescents scoring high on the Children's Depression Inventory. *Psychoneuroendocrinology*, *34*, 791–794. http://doi.org/10.1016/j.psyneuen.2008.12.008

Van den Bergh, B. R. H., Van Calster, B., Pinna Puissant, S., & Van Huffel, S. (2008). Self-reported symptoms of depressed mood, trait anxiety and aggressive behavior in post-pubertal adolescents: Associations with diurnal cortisol profiles. *Hormones and Behavior*, *54*(2), 253–257. http://doi.org/10.1016/j.yhbeh.2008.03.015

van Eijndhoven, P., van Wingen, G., Katzenbauer, M., Groen, W., Tepest, R., Fernández, G., … Tendolkar, I. (2013). Paralimbic Cortical Thickness in First-Episode Depression: Evidence for Trait-Related Differences in Mood Regulation. *American Journal of Psychiatry*, *170*(12), 1477–1486. http://doi.org/10.1176/appi.ajp.2013.12121504

van Tol, M.-J., Li, M., Metzger, C. D., Hailla, N., Horn, D. I., Li, W., … Walter, M. (2014). Local cortical thinning links to resting-state disconnectivity in major depressive disorder. *Psychological Medicine*, *44*(10), 2053–65. http://doi.org/10.1017/S0033291713002742

van Zoonen, K., Buntrock, C., Ebert, D. D., Smit, F., Reynolds, C. F., Beekman, A. T., & Cuijpers, P. (2014). Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *International Journal of Epidemiology*, *43*(2), 318–329. http://doi.org/10.1093/ije/dyt175

Varghese, F. P., & Brown, E. S. (2001). The Hypothalamic-Pituitary-Adrenal Axis in Major Depressive Disorder: A Brief Primer for Primary Care Physicians. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*(4), 151–155. http://doi.org/10.2174/1566523013348869

Verma, R., Balhara, Y. P. S., & Gupta, C. S. (2011). Gender differences in stress response: Role of developmental and biological determinants. *Industrial Psychiatry Journal*, *20*(1), 4–10. http://doi.org/10.4103/0972-6748.98407

Videbech, P., & Ravnkilde, B. (2004). Hippocampal Volume and Depression: A Meta-Analysis

of MRI Studies. *American Journal of Psychiatry*, *161*(11), 1957–1966. http://doi.org/10.1176/appi.ajp.161.11.1957

Vijayakumar, N., Allen, N. B., Youssef, G., Dennison, M., Yücel, M., Simmons, J. G., & Whittle, S. (2016). Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Human Brain Mapping*, *37*(6), 2027–2038. http://doi.org/10.1002/hbm.23154

Vitiello, B., Emslie, G., Clarke, G., Wagner, K. D., Asarnow, J. R., Keller, M. B., … Brent, D. A. (2011). Long-Term Outcome of Adolescent Depression Initially Resistant to Selective Serotonin Reuptake Inhibitor Treatment. *The Journal of Clinical Psychiatry*, *72*(3), 388–396. http://doi.org/10.4088/JCP.09m05885blu

Vitiello, M. V., Larsen, L. H., Moe, K. E., Borson, S., Schwartz, R. S., & Prinz, P. N. (1996). Objective Sleep Quality of Healthy Older Men and Women Is Differentially Disrupted by Nighttime Periodic Blood Sampling via Indwelling Catheter. *Sleep*, *19*(4), 304–311. http://doi.org/10.1093/sleep/19.4.304

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. http://doi.org/10.1176/appi.ajp.2015.15040492

Vreeburg, S. A., Hoogendijk, W. J. G., van Pelt, J., DeRijk, R. H., Verhagen, J. C. M., van Dyck, R., … Penninx, B. W. J. H. (2009). Major Depressive Disorder and Hypothalamic-Pituitary-Adrenal Axis Activity. *Archives of General Psychiatry*, *66*(6), 617–626. http://doi.org/10.1001/archgenpsychiatry.2009.50

Vreeburg, S. A., Zitman, F. G., van Pelt, J., DeRijk, R. H., Verhagen, J. C. M., van Dyck, R., … Penninx, B. W. J. H. (2010). Salivary Cortisol Levels in Persons With and Without Different Anxiety Disorders. *Psychosomatic Medicine*, *72*(4), 340–347. http://doi.org/10.1097/PSY.0b013e3181d2f0c8

Vrouva, I., Fonagy, P., Fearon, P. R. M., & Roussow, T. (2010). The risk-taking and self-harm inventory for adolescents: Development and psychometric evaluation. *Psychological Assessment*, *22*(4), 852–865. http://doi.org/10.1037/a0020583

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. http://doi.org/10.3758/BF03206482

Wagner, G., Schultz, C. C., Koch, K., Schachtzabel, C., Sauer, H., & Schlösser, R. G. (2012). Prefrontal cortical thickness in depressed patients with high-risk for suicidal behavior.

*Journal of Psychiatric Research*, *46*(11), 1449–55. http://doi.org/10.1016/j.jpsychires.2012.07.013

Walker, E., Mittal, V., & Tessner, K. (2008). Stress and the Hypothalamic Pituitary Adrenal Axis in the Developmental Course of Schizophrenia. *Annual Review of Clinical Psychology*, *4*(1), 189–216. http://doi.org/10.1146/annurev.clinpsy.4.022007.141248

Walker, E. R., & Druss, B. G. (2015). Rate and Predictors of Persistent Major Depressive Disorder in a Nationally Representative Sample. *Community Mental Health Journal*, *51*(6), 701–707. http://doi.org/10.1007/s10597-014-9793-9

Wang, L., Xia, M., Li, K., Zeng, Y., Su, Y., Dai, W., … Si, T. (2015). The effects of antidepressant treatment on resting-state functional brain networks in patients with major depressive disorder. *Human Brain Mapping*, *36*(2), 768–778. http://doi.org/10.1002/hbm.22663

Weersing, V. R., Jeffreys, M., Do, M.-C. T., Schwartz, K. T. G., & Bolano, C. (2017). Evidence Base Update of Psychosocial Treatments for Child and Adolescent Depression. *Journal of Clinical Child & Adolescent Psychology*, *46*(1), 11–43. http://doi.org/10.1080/15374416.2016.1220310

Weersing, V. R., Rozenman, M., & Gonzalez, A. (2009). Core Components of Therapy in Youth. *Behavior Modification*, *33*(1), 24–47. http://doi.org/10.1177/0145445508322629

Weissman, M. M., Bland, R., Joyce, P. R., Newman, S., Wells, J. E., & Wittchen, H.-U. (1993). Sex differences in rates of depression: cross-national perspectives. *Journal of Affective Disorders*, *29*(2–3), 77–84. http://doi.org/10.1016/0165-0327(93)90025-F

Weissman, M. M., Wolk, S., Goldstein, R. B., Moreau, D., Adams, P., Greenwald, S., … Wickramaratne, P. (1999). Depressed Adolescents Grown Up. *JAMA*, *281*(18), 1707. http://doi.org/10.1001/jama.281.18.1707

Weisz, J., Kuppens, S., Ng, M., Eckshtain, D., Ugueto, A., Vaughn-Coaxum, R., … Fordwood, S. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *The American Psychologist*, *72*(2), 79–117.

Weisz, J., McCarty, C., & Valeri, S. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychological Bulletin*, *132*(1), 132–149.

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., … Health, and the R. N. on Y. M. (2012). Testing Standard and Modular Designs for Psychotherapy Treating Depression, Anxiety, and Conduct Problems in Youth. *Archives of General Psychiatry*, *69*(3), 274. http://doi.org/10.1001/archgenpsychiatry.2011.147

Whitaker, K. (2015). FreeSurfer Vertex Analysis. Retrieved from https://github.com/KirstieJane/NSPN_CODE/blob/master/freesurfer_vertex_analysis_multipleSmoothing.sh

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. http://doi.org/10.1002/sim.4067

Whitfield-Gabrieli, S., Ghosh, S. S., Nieto-Castanon, A., Saygin, Z., Doehrmann, O., Chai, X. J., … Gabrieli, J. D. E. (2015). Brain connectomics predict response to treatment in social anxiety disorder. *Molecular Psychiatry*, *21*, 680–685.

Whittle, S., Lichter, R., Dennison, M., Vijayakumar, N., Schwartz, O., Byrne, M. L., … Allen, N. B. (2014). Structural Brain Development and Depression Onset During Adolescence: A Prospective Longitudinal Study. *American Journal of Psychiatry*, *171*, 564–571.

Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-order growth curves and mixture modeling with Mplus : a practical guide*.

Wickrama, T., & Wickrama, K. A. S. (2010). Heterogeneity in Adolescent Depressive Symptom Trajectories: Implications for Young Adults' Risky Lifestyle. *Journal of Adolescent Health*, *47*(4), 407–413. http://doi.org/10.1016/J.JADOHEALTH.2010.02.013

Wierenga, L. M., Bos, M. G. N., Schreuders, E., vd Kamp, F., Peper, J. S., Tamnes, C. K., & Crone, E. A. (2018). Unraveling age, puberty and testosterone effects on subcortical brain development across adolescence. *Psychoneuroendocrinology*, *91*, 105–114. http://doi.org/10.1016/J.PSYNEUEN.2018.02.034

Wigman, J. T. W., van Nierop, M., Vollebergh, W. A. M., Lieb, R., Beesdo-Baum, K., Wittchen, H.-U., & van Os, J. (2012). Evidence that psychotic symptoms are prevalent in disorders of anxiety and depression, impacting on illness onset, risk, and severity--implications for diagnosis and ultra-high risk research. *Schizophrenia Bulletin*, *38*(2), 247–257. http://doi.org/10.1093/schbul/sbr196

Wigman, J. T. W., van Os, J., Abidi, L., Huibers, M. J. H., Roelofs, J., Arntz, A., … Peeters, F. P. M. L. (2014). Subclinical psychotic experiences and bipolar spectrum features in depression: association with outcome of psychotherapy. *Psychological Medicine*, *44*(2), 325–336. http://doi.org/10.1017/S0033291713000871

Wilkinson, P., Dubicka, B., Kelvin, R., Roberts, C., & Goodyer, I. (2009). Treated depression in adolescents: predictors of outcome at 28 weeks. *The British Journal of Psychiatry*, *194*(4), 334–341. http://doi.org/10.1192/bjp.bp.108.052381

Wilkinson, P., Kelvin, R., Roberts, C., Dubicka, B., & Goodyer, I. (2011). Clinical and Psychosocial Predictors of Suicide Attempts and Nonsuicidal Self-Injury in the Adolescent Depression Antidepressants and Psychotherapy Trial (ADAPT). *American Journal of Psychiatry*, *168*(5), 495–501. http://doi.org/10.1176/appi.ajp.2010.10050718

William K.D.S., & Yurgelun-Todd, D. A. (2006). Ventromedial prefrontal activity correlates with depressed mood in adolescent children. *NeuroReport*, *17*(2), 167–171. http://doi.org/10.1097/WNR.0000000000000761

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., … Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–46. http://doi.org/10.1016/j.neuroimage.2009.12.028

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., … Steinhausen, H.-C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655–679. http://doi.org/10.1016/J.EURONEURO.2011.07.018

Wong, M.-L., & Licinio, J. (2004). From monoamines to genomic targets: a paradigm shift for drug discovery in depression. *Nature Reviews Drug Discovery*, *3*(2), 136–151. http://doi.org/10.1038/nrd1303

World Health Organisation. (1992). *The ICD-10 classification of mental and behavioural disorder: Clinical descriptions and diagnostic guidelines.* World Health Organisation: Geneva.

Wüst, S., Federenko, I., Hellhammer, D. H., & Kirschbaum, C. (2000). Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology*, *25*(7), 707–720. http://doi.org/10.1016/S0306-4530(00)00021-4

Yamagishi, T., Takagishi, H., Fermin, A. de S. R., Kanai, R., Li, Y., & Matsumoto, Y. (2016). Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. *Proceedings of the National Academy of Sciences*, *113*(20), 5582–5587. http://doi.org/10.1073/PNAS.1523940113

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Ueda, K., Suzuki, S., & Shigetoyamawaki. (2010). Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *Journal of Affective Disorders*, *122*(1–2), 76–85.

http://doi.org/10.1016/j.jad.2009.06.017

Young, E. A., Abelson, J., & Lightman, S. L. (2004). Cortisol pulsatility and its role in stress regulation and health. *Frontiers in Neuroendocrinology*, *25*, 69–76. http://doi.org/10.1016/j.yfrne.2004.07.001

Young, E. A., Altemus, M., Lopez, J. F., Kocsis, J. H., Schatzberg, A. F., DeBattista, C., & Zubieta, J.-K. (2004). HPA axis activation in major depression and response to fluoxetine: a pilot study. *Psychoneuroendocrinology*, *29*(9), 1198–204. http://doi.org/10.1016/j.psyneuen.2004.02.002

Zeng, L.-L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., … Hu, D. (2012). Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain*, *135*(5), 1498–1507. http://doi.org/10.1093/brain/aws059

Zhang, W.-N., Chang, S.-H., Guo, L.-Y., Zhang, K.-L., & Wang, J. (2013). The neural correlates of reward-related processing in major depressive disorder: a meta-analysis of functional magnetic resonance imaging studies. *Journal of Affective Disorders*, *151*(2), 531–9. http://doi.org/10.1016/j.jad.2013.06.039

Zhao, K., Liu, H., Yan, R., Hua, L., Chen, Y., Shi, J., … Yao, Z. (2017). Cortical thickness and subcortical structure volume abnormalities in patients with major depression with and without anxious symptoms. *Brain and Behavior*, *7*(8), e00754. http://doi.org/10.1002/brb3.754

Zhao, Y.-J., Du, M.-Y., Huang, X.-Q., Lui, S., Chen, Z.-Q., Liu, J., … Gong, Q.-Y. (2014). Brain grey matter abnormalities in medication-free patients with major depressive disorder: a meta-analysis. *Psychological Medicine*, *44*, 2927–2937. http://doi.org/10.1017/S0033291714000518

Zhao, Y.-J., Du, M.-Y., Huang, X.-Q., Lui, S., Chen, Z.-Q., Liu, J., … Gong, Q.-Y. (2014). Brain grey matter abnormalities in medication-free patients with major depressive disorder: a meta-analysis. *Psychological Medicine*, *44*(14), 2927–37. http://doi.org/10.1017/S0033291714000518

Zheng, H., Xu, L., Xie, F., Guo, X., Zhang, J., Yao, L., & Wu, X. (2015). The Altered Triple Networks Interaction in Depression under Resting State Based on Graph Theory. *BioMed Research International*, *2015*, 386326. http://doi.org/10.1155/2015/386326

Zhou, X., Hetrick, S. E., Cuijpers, P., Qin, B., Barth, J., Whittington, C. J., … Xie, P. (2015). Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis. *World Psychiatry*,

*14*(2), 207–222. http://doi.org/10.1002/wps.20217

Zimmerman, M., Martinez, J. A., Attiullah, N., Friedman, M., Toba, C., Boerescu, D. A., & Rahgeb, M. (2012). Why Do Some Depressed Outpatients Who Are in Remission According to the Hamilton Depression Rating Scale Not Consider Themselves to Be in Remission? *The Journal of Clinical Psychiatry*, *73*(6), 790–795. http://doi.org/10.4088/JCP.11m07203

Zisook, S., Lesser, I., Stewart, J. W., Wisniewski, S. R., Balasubramani, G. K., Fava, M., … Rush, A. J. (2007). Effect of Age at Onset on the Course of Major Depressive Disorder. *American Journal of Psychiatry*, *164*(10), 1539–1546. http://doi.org/10.1176/appi.ajp.2007.06101757

Zobel, A. W., Nickel, T., Sonntag, A., Uhr, M., Holsboer, F., & Ising, M. (2001). Cortisol response in the combined dexamethasone/CRH test as predictor of relapse in patients with remitted depression. *Journal of Psychiatric Research*, *35*(2), 83–94. http://doi.org/10.1016/S0022-3956(01)00013-9

Zobel, A. W., Yassouridis, A., Frieboes, R.-M., & Holsboer, F. (1999). Prediction of Medium-Term Outcome by Cortisol Response to the Combined Dexamethasone-CRH Test in Patients With Remitted Depression. *Am J Psychiatry*, *156*(6), 949–951.

Zorn, J. V., Schür, R. R., Boks, M. P., Kahn, R. S., Joëls, M., & Vinkers, C. H. (2017). Cortisol stress reactivity across psychiatric disorders: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *77*, 25–36. http://doi.org/10.1016/J.PSYNEUEN.2016.11.036

**Supplement: Paper submitted for peer review**

**Title:** Trajectories Of Depression Symptom Change During And Following Treatment In Adolescents With Unipolar Major Depression

**Running Title:** Depression, Treatment, Trajectories, Outcome.

Sian Emma Davies, MSc, Sharon AS Neufeld BSc, Eleonore van Sprang, MSc, Lizanne Schweren, PhD, Rogier Keivit, PhD, Peter Fonagy, PhD FMedSci, Bernadka Dubicka, MD, R Kelvin, M.B.B.S., N Midgley, PhD, S Reynolds PhD, M Target, PhD, PO Wilkinson, MD, A van Harmelen, PhD & Ian M Goodyer, MD FMedSci.

Sian Emma Davies(sed48@medschl.cam.ac.uk) , Sharon Neufeld (sasn2@medschl.cam.ac.uk) , Raphael Kelvin (raphael.kelvin@btinternet.com) , Paul Oliver Wilkinson(pow12@cam.ac.uk) , and Ian Michael Goodyer (ig104@cam.ac.uk) , are from the Department of Psychiatry, University of Cambridge, Douglas House, 18b Trumpington Raod, Cambridge CB2 8AH. Eleonore van Sprang (e.vansprang@ggzingeest.nl.) is from the Department of Psychiatry, Amsterdam UMC, location VUmc, Oldenaller 1, 1081HJ Amsterdam, the Netherlands. Lizanne Schweren (l.j.s.schweren@umcg.nl) , is from the Department of Psychiatry of the University Medical Center Groningen, The Netherlands. Peter Fonagy(p.fonagy@ucl.ac.uk) , Nick Midgley (nicholas.midgley@ucl.ac.uk) , Mary Target (M.Target@ucl.ac.uk) are from the Research Department of Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London, London, UK. Shirley Reynolds (s.reynolds@reading.ac.uk) is at the School Of Psychology And Clinical Language Sciences, University of Reading Whiteknights PO Box 217 Reading Berkshire RG6 6AH United Kingdom.

Coresponding Authors: Sian Emma Davies(sed48@medschl.cam.ac.uk) and Ian Michael Goodyer (ig104@cam.ac.uk)

**Contributors**

SED and IMG wrote the initial draft of the manuscript. All authors contributed to and approved the final manuscript.

**Declaration of interests**

IMG, POW and RK received consulting fees from Lundbeck.

All other authors declare no competing interests.

**Acknowledgments**

The study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project number 06/05/01). ED was funded by a doctoral studentship awarded by the Neuroscience in Psychiatry (NSPN.Org) Consortium itself funded by a strategic award from the Wellcome Trust (095844/Z/11/Z) awarded to IMG and PF. AVH is in receipt of a Dorothy Hodgkin Research Fellowship from the Royal Society.

The authors would like to thank the project manager and all research assistants who collected and curated the impact data over the three research sites.

The views expressed in this publication are those of the authors and do not necessarily reflect those of the HTA programme, NIHR UK, National Health Service, or the Department of Health, UK.

**Abstract**

**Objective:** To classify a cohort of depressed adolescents recruited to the UK IMPACT trial. We examined for predictors and determined the dadat driven outomes of pateitnes with a priori operational definitions of treatment response.

**Method:** Secondary data analysis using growth mixture modelling (GMM). Missing data was imputed.Trajectories of self-reported depressive symptoms were plotted using measures taken at 6 nominal time points over 86 weeks from randomisation in all 465 patients.

**Results:** A piecewise GMM categorised patients into two classes with initially similar and subsequently distinct trajectories. Both groups had a significant decline in depressive symptoms over the first 18 weeks. Eighty-four percent (84.1%, n=391) of patients were classed as "continued improvers" with symptoms reducing over the full duration of the study. A further class of 15.9% (n=74) of patients were termed "halted improvers" with higher baseline depression scores, faster early recovery but no further improvement after 18 weeks. Presence of baseline co-morbidity moderately increased membership (OR = 1.40, CI 1.00-1.96) of the halted improvers class. Compared to the data driven classes, by end of stud a clinical cut-off (<=27) and symptom reduction (>=50%) scores for defining response. misclasiffied 15% for 31% cases respectively.

**Conclusion:** A fast reduction in depressive symptoms in the first few weeks of treatment may not indicate a good prognosis. Halted improvement may only be apparent after 18 weeks of treatment. Utilising longitudinal modeling may improve the precision of revealing differential responses to treatment. Clinical progress maybe somewhat better in the year after treatment than previously considered.

**Background/rationale**

Adolescence denotes the highest incidence risk rate for the emergence of major depression over the lifecourse [1]. A quarter of depressed adults report that their illness began during adolescence [2], with early onset being a risk factor for subsequent relapse and recurrence[3,4] and impairment [2]. The effectiveness of current treatment strategies, which may involve psychological and SSRI medication alone or in combination [5], have reported moderate effect sizes of between 0.3-0.6 [6,7]. At least 20% of adolescents with major depression show no response to available treatments[8] but the reasons for this are unclear.

To improve our recognition of patients who may or may not benefit from available treatments, requires more precision in defining treatment response [9–11]. Currently there are large variations between trials with respect to the definition of response[12]. Such discrepancies in measurement can diminish comparability between studies and impact the proportions of patients considered responders or non-responders respectively [13,14]. Definitions are commonly based on percentage reduction in symptoms or on final scores below an a priori clinical cut-off . These methods are arbitrary and may lack clinical meaning[11,13]. Furthermore there can be a substantial overlap of patients who simultaneously meet criteria for non-remission (e.g. a final Hamilton Rating Scale for Depression (HRSD) score of >=7) but achieve a positive clinical response (reduction of >=50% in HRSD)[15].

Empirical person centered modelling techniques, such as growth mixture modelling (GMM), may address some of the validity issues with a priori definitions, by categorising patients post-hoc[16]. This computational technique searches for naturally occurring heterogeneity to categorize patients into particular latent classes that follow similar trajectories. Such approaches make no assumptions on what percentage reduction constitutes a meaningful response, providing a far less arbitrary definition and more homogenous subgroups of patients[11,13,17]. Homogeneity of groups is a particularly important characteristic for research

investigating predictors of response types that are likely to have small effect sizes. In modelling patterns of change over time, GMM is additionally able to provide descriptions of behavioural trajectory, which reveals how groups differ in their shape over time[16,18].

GMM analyses of treatment trials in depression have all supported the presence of multiple, qualitatively distinct classes of symptom trajectories in patients, and specific predictors of such classes[11,13,17–20]. Some studies also reported differential treatment preferences between classes[13,17,19,20]. However, most studies in Major Depressive Disorders (MDD) to date have been with adult patients with primarily pharmacological/combination interventions[11,13,18–20] . One recent report was from the Treatment of Adolescent Depression Study(TADS) noting that, at 12 weeks, there were 2 groups that improved over the trial, and a further group showing limited change [21]. A report of healthy adolescents recruited to a depression prevention study carried out a longitudinal latent class growth analysis over a 2 year period. The authors noted 2 groups where symptoms gradually reduced over time; 1 group which showed no change, and a further whose symptom count resurged within 6 months of the intervention[17].

As shown current GMM studies of treatment in depression have also focused on short-term response of typically 8-12 weeks[13,18,19]. However, two studies have found that the duration of follow-up plays a contributory role in determining response classes[11,17]. Thibodeau and colleagues[11] found that short-term follow-up mistakenly classified some responders as non-responders, while Brière and colleagues[17] found a subgroup of adolescents that showed a significant decline in symptoms during the first 12 weeks, but relapsed after this point. These fidnings highlight that longer-term follow-ups may improve the precision of denoting true responders, sustained non-responders and relapsing patients[11,17].

**Objectives and Hypotheses**

Our primary objective was to reveal trajectories of depression symptoms from randomisation to the final assessment one year following end of treatment. The specific aims were to: i) define the number and shape of longitudinal classes pf patients revealed from depression symptoms only; and ii) compare the defined groups with standard a priori definitions of response/non-response. Prior literature would suggest that 4 classes could emerge with favourable trajectories showing rapid and gradual improvement in symptoms[13], and unfavourable trajectories showing either no improvement, or a relapsing trajectory shape[17].

Our second objective was then to test whether selected baseline clinical characteristics would predict class membership. Prior GMM studies have associated a number of baseline clinical variables with class membership and unfavourable trajectory including: older age[13], higher baseline anxiety[18], baseline level of function[11], emotional stability and comorbidities[11]. Further, psychotic experiences (PEs), even at the subclinical level, have been noted as a potential risk factor for both relapse and treatment resistance of mood disorders [22 23,24]. The prevalence of PEs in community samples of adolescents may be higher than in adult samples [25,26]. Therefore, we specifically hypothesised that subclinical PEs and non-depressive comorbidity at baseline would be associated with unfavourable trajectory classes.

## Methods

*Study Design*

This study was a re-analysis of the Improving Mood with Psychoanalytic and Cognitive Therapies (IMPACT) trial. The IMPACT study was a multicentre, pragmatic, observer-blind, randomised controlled trial investigating whether there was a superior effect for two specialist psychological treatments (cognitive behavioural therapy, short-term psychoanalytic psychotherapy) compared with a reference treatment of brief psychosocial intervention (BPI) on reducing self-reported depressive symptoms by end of follow up 12 months after end of treatment[27]. Participants were randomly assigned to one of the three treatment arms, with stochastic minimisation by age, sex, self-reported depression sum score, and region, as per study protocol[28]. The primary findings from the trial demonstrated no differences in depression symptoms sum scores between treatment groups over the course and by the end of the study [27]. Consequently, treatment group was collapsed for the present study, to investigate the symptom trajectories in this whole population. Self-reported depressive symptomatology was measured at 6 nominal time points: baseline, 6, 12, 36, 52 and 86 weeks post-randomisation. The last 2 time points were post treatment which was completed by 36 weeks in >95% of the cohort [26]. Based on each individual's sum score symptom change over the trial, a series of growth mixture models were conducted to determine the best fitting model, and the number of classes of individuals present within the dataset. Baseline clinical characteristics were used to describe the sub-classes, and estimate the predictive value of these characteristics in determining class membership.

*Setting*

The IMPACT trial[28] recruited patients from 15 National Health Service child and adolescent mental health service (CAMHS) clinics across 3 geographical regions in the UK: East Anglia, North London and North-West England covering an estimated 1,000,000 adolescents aged

11-17 years. The study recruited, assessed and followed up all participants between June 29, 2010 and Jan 17, 2013[27].

*Participants*

Adolescents aged between 11 and 17 years, with a current diagnosis of major depression (DSM-IV[29]) were enrolled. Patients were randomized to either cognitive behavioural therapy (CBT), short-term psychoanalytic psychotherapy (STPP) or a reference treatment of brief psychological intervention (BPI), as per study protocol. Full details on patient inclusion and exclusion criteria can be found in the study protocol[28].

*Variables*

Symptom trajectory class membership was defined through growth mixture modelling (see below) using the self-reported Mood and Feelings Questionnaire (MFQ) score across all time-points. This is a 33-item Questionnaire[30] of depressive symptomatology covering the past 2 weeks. MFQ items were measured on a 3-point scale (almost never, sometimes, often/almost always). Total scores (range of 0-66) were used in GMMs. Higher scores indicated more severe depressive symptoms and were positively correlated with greater psychosocial impairment[31].

As noted above, a number of baseline clinical variables were investigated for their potential predictive value over class membership. These were derived sum scores from self-report measures for anxiety (the Revised Children's Manifest Anxiety Scale, RCMAS)[32], obsessionality (the short Leyton Obsessional Inventory for adolescents, LOI)[33] and overall psychiatric symptomatology/impairment (the Health of the Nation Outcome Scales for Children and Adolescents, HONOSCA)[34]. Lifetime suicidal thoughts and suicide attempts were defined as binary variables (yes, no) from data derived from the Columbia Suicide

Inventory[35]. Lifetime non-suicidal self-injury was measured using the self-report Risk and Self Harm Inventory[36]. The Kiddie-Schedule for Affective Disorder and Schizophrenia (K-SADS)[37] interview was used to assess the presence of psychiatric diagnoses and psychotic symptoms at baseline. Comorbidity was defined on an ordinal scale, as the number of additional mental illnesses other than depression that met threshold criteria during interview. Psychotic symptoms were also defined on an ordinal scale (absent, present: subthreshold, or present: threshold), as answering positively to either of the two screening questions for psychosis present in the K-SADS interview.

**Bias**

The recruitment sites were dependent on referrals from primary care sources including family physicians, community mental health teams and self-referral. Clinics who participated were invited and not selected randomly. Therefore, we cannot be certain that the sample is necessarily representative of major depression in the adolescent population at large nor of cases usually referred to child and adolescent mental health clinics in the UK. However, there are no other referral options for the primary care services other than their local NHS services and therefore the clinics are likely to be receiving the majority of referrals for major depression. Finally these clinics are part of routine NHS mental health services and not set up solely for the purposes of the IMPACT study.

**Study size**

Of 557 participants screened for eligibility into the IMPACT trial, 87 were excluded (73 did not meet criteria for major depression, 4 had mania, 4 had a primary substance use disorder, 2 had received previous treatment used in the trial, 1 had autism, 1 was pregnant, 1 would not engage, 1 was unable to read or understand information). 470 adolescents were therefore

randomised to 3 treatment arms. Subsequently, 5 withdrew from the study, leaving 465 patients included and in the current analysis[27].

## Statistical methods

*Imputation*

Despite follow-up rates of 80% at 86-weeks[27], multiple imputation was required in order to maximize sample size and achieve convergence for the GMM. Due to a wealth of auxiliary variables predicting missingness, data was presumed to be missing at random. Also due to these auxiliary variables, multiple imputation was favoured over Full Information Maximum Likelihood as auxiliary variables can easily be incorporated into a multiple imputation model and help decrease bias and increase efficiency[38]. Variables at all time points were assessed for inclusion in the imputation model in addition to MFQ items. Those related to outcome ($p<0.05$ or $r>=0.3$) and/or missingness in outcome, and variables used in final analyses, were included in the model[39]. Additional non-missing variables were also included to improve model prediction. This resulted in imputation of 24 variables plus the 33 MFQ items, repeated over six assessments, yielding a dataset too large to impute in wide format. Thus, time-varying data was imputed in long format, a method which is less biased under conditions of less missing data, more repeated measures, and a reliable outcome measure[40], as is the case in the present data. For each model, fifty datasets were multiply imputed using chained equations[39]. As it is not possible to obtain the VLMR and LMR fit statistics for model comparison in a multiply imputed dataset, multiple imputations were averaged prior to estimation of GMM. While we acknowledge it is more optimal to obtain model estimates from each of the multiply imputed datasets and then combine estimates[39], our approach allowed us to obtain these fit statistics which are crucial for determining the most optimal model.

*Growth Mixture Models*

Total MFQ scores at baseline, 6, 12, 36, 52 and 86 weeks were the intended time-points of the trial. However, substantial variation existed across individuals in the timing of their assessments. To model the symptom change experienced by patients accurately, the mean time of each actual assessment (weeks) was taken as the focal time points for GMMs. This therefore corresponded to baseline, 12, 18, 43, 60 and 95 weeks post-randomisation. Variation in time of assessment was further included as a covariate in all growth models.

Growth mixture models were tested in the Mplus program version 8.0[41]. Four growth trends were considered: a linear and quadratic growth trend respectively , and two linear-piecewise growth trends. We investigated piecewise growth trends because we hypothesised that different rates of improvement might occur during treatment vs follow-up stages of the trial. Due to the average length of treatment falling between the means of two assessment time-points, we considered two possible transition points in the models: the first was placed at the third assessment (18 weeks on average from baseline) and the second at the fourth assessment (43 weeks from baseline).

Classes were incrementally added to the single class model to determine the best fit. All models allowed for within-class variation, letting patient's symptom scores vary around the mean of the group. However, upon testing, there was no evidence of significant variation between-classes, (See Supplementary Matieral, Table 1), meaning that the variation around each group mean was not significantly different for the respective classes. Consequently, between-class variation was held equal for all growth factors. Only solutions that were replicated with different starting values were accepted.

We considered models with 1 to 5 trajectory classes due to our a priori hypothesis, and retained the most parsimonious model based on the following criteria:

Firstly, the model with the lowest AIC and BIC was retained, with BIC values favoured in cases of discrepancy. Models were only considered as favourable over the model with one fewer latent class if the BIC difference was 10 or more[13]. Secondly, entropy values were considered. This is a measure of the uncertainty of the model in the classification of subjects into the correct class. Values closer to 1 are preferred. Finally, clinical interpretability and relevance of the class trajectories, as well as class size, were taken into account. Given the sample size of our cohort, models where classes contained less than 10% of the sample were rejected as these were not considered numerically stable[13]. Patients were assigned to their most likely class based on model probabilities.

*Baseline clinical characteristics and predictors of class membership*

We undertook a two-step approach to investigating trajectories of symptom change, and predictors of these trajectories. After selecting the best fitting trajectory model, we saved the information on most likely class membership for all patients and conducted analyses of associations in a separate step. Univariate analyses were conducted (chi-square, or t-tests) to determine whether there were significant differences between classes on baseline demographic and clinical characteristics. Mann-Whitney-Wilcoxon tests were performed where data were non-normal. Multinomial logistic regression was then used to determine which variables predicted class membership. R-squared statistic indicated how much variance the regression model explained. Odds Ratios are reported for any predictors.

*Agreement between categorical definitions and GMM model result*

Cohen's Kappa coefficient[42] was used to test the agreement between trajectory classes and the two most commonly used definitions of response/non-response in clinical trials. The first

definition reclassified patients as clinical responders/non-responders if they had a 50% reduction/or not in MFQ score by end of study (12 months following treatment). The second definition reclassified patients as clinical/non-responders if their MFQ score at 12 months following treatment was below or above 27 respectively [43].

## Results

*Participants*

All 465 participants who entered the trial were available for the longitudinal analysis. Across all assessments, 65% or more of the sample (304/465) had full data on all MFQ items.

*Outcome data*

Fit information for all piecewise class models tested are provided in Table 1. A two-class, piecewise model that separately modelled the change in depressive symptoms, linearly, over the first 18 weeks of treatment (on average; assessments 0-2), and then the remaining period of the trial (assessments 3-5), was identified as the optimal model. This is illustrated in Figure 1. BIC showed a favourable decrease of approximately 42 with the addition of a second class from the single class solution. Although the 3 class solution yielded a slightly lower BIC ($\Delta$BIC=4.039), the decrease in entropy below commonly accepted thresholds suggests poor classification precision in the 3 class solution(entropy=0.844 vs 0.734). The Mplus code for the two-class model is provided in the Supplementary Material.

Table One Here

Figure One Here

The two-class piecewise GMM divided patients into a comparatively large class of 391 (84.1%; class 1), and a small class of 74 (15.9%; class 2). The mean depression scores and the % change between time points is shown in table 2.

Table Two Here

Class 2 on average showed significantly higher baseline levels of MFQ scores than class 1 (Wald $X^2(1)= 25.577$, $p<.001$). Both classes showed a significant decrease in MFQ score over the first 18 weeks of the trial (-6.466, $p<.001$, and -8.794, $p<.001$ respectively). However, class 2 showed a significantly faster rate of MFQ reduction compared with class 1 (Wald $X^2(1)=5.446$, $p=.0196$).

These two classes departed markedly from each other after 18 weeks. While class 2 showed a significantly faster rate of improvement over the first 18 weeks, they showed no further improvement statistically over the rest of the trial (0.899, $p=.183$). Conversely, class 1 on average, continued to show a significant decline in MFQ score, albeit slower than in the initial 18 weeks (-1.639, $p=.014$). This difference between the second linear slopes of the two classes was statistically significant (Wald $X^2(1) =167.075$, $p<.001$). By the end of the trial, Class 1 showed on average a 60.5% improvement in depressive symptoms, compared with 11.0% in class 2. We labelled class 2 as “halted-improvers” due to the statistically revealed transition point and class 1 as “continued-improvers”.

*Baseline characteristics of each class*

The characteristics of patients following each trajectory class are described in Table 3.

Table Three Here

The two classes differed significantly on their proportions of males and females. Eighty-five percent of halted improvers were female, compared with only 73% of continued improvers. No other demographic characteristics were significantly different between groups.

Halted improvers on average showed significantly higher obsessionality scores, higher likelihood of NSSI, higher HoNOSCA (greater overall psychiatric symptoms/impairment) and more comorbid disorders than continued improvers at baseline. Interestingly neither baseline levels of suicidality, psychotic symptoms nor treatment group discriminated between the classes. There were no significant differences between the two classes on any characteristic related to treatment.

*Predictors of Trajectory Class Membership*

Class 1 acted as the reference class for the logistic regression. Results from the logistic regression are shown in Table 4. When controlling for variables included in the model, only the number of co-morbid psychiatric diagnoses significantly predicted a higher probability of membership to halted-improvers compared with continual-improvers (Table 3). With each increasing number of co-morbid diagnoses the odds of a patient belonging to the halted-improvers class compared to the continual-improvers class increased by a factor of 1.4. This analysis produced a significant improvement in the fit of the model over the constant ($X^2(4)=46.03$, $p<.001$) but only explained 5.4% of the total variance in class membership allocation (Cox and Snell $R^2= 0.054$).

Table 4 Here

Agreement between GMM classes and a priori categorical definitions of response

Continued-improvers were considered the comparative for “clinical responders”, and halted-improvers were considered the comparative for “clinical non-responders”. Comparisons are illustrated in Figure 2.

Figure Two Here

*Percentage reduction by end of study*

There was moderate agreement between trajectory class membership and clinical categorical outcome when 'treatment response' is defined by percentage reduction of 50% by end of study (k=0.412, p<.001). All halted-improvers were also clinical non-responders by this definition. However, only 269 of 391 (69%) of continued-improvers were also clinical responders. The remaining continued-improvers (122 of 391; 31%) were classified as clinical non-responders on the percent reduction category.

*Clinical cut-off*

There was stronger, albeit still moderate agreement between trajectory membership and clinical categorical outcomes when defined by a cut-off score of 27 on the MFQ by end of study (k=0.642, p<.001). All halted-improvers were also clinical non-responders. However, only 332 of 391 (85%) of continued-improvers were also clinical responders. The remaining continued-improvers (59 of 391; 15%) were classified as clinical non-responders.

Overall if either a priori category definitions have been used in this study a false negative classification rate of between 15%-31% would have been reported.

## Discussion

*Key results and Interpretation*

Using a cohort of depressed adolescents recruited into a clinical trial we computed a depression symptom trajectory that revealed a piecewise function, with two separate linear trajectories. The best fit model was for two classes of individuals differing on specific characteristics. We defined a large (84.1%) group of continued improvers and a small (15.9%) group of halted improvers. We noted that both groups improved significantly over the first 18 weeks of the trial. The halted group showed a cessation in improvement from thereon and this may index a putative relapsing group by traditional end-of-study measures. As the slope value for the second phase of the model did not reach significance in this cohort this requires replicaiton. These findings are consistent with secondary findings from two other clinical trials investigating the effectiveness of antidepressants in adult samples[13,18]. Although there are a number of studies that support the existence of more than two classes, these applied constraints on investigated trajectory shape[11,19], and eliminated within-class variation from their models[17]. These methodological choices are often necessary when models struggle to fit the data, and both lead models to favour more classes[44]. To assume that no individual variation exists within classes in depressive patients we felt was not representative of real data. We therefore favoured a GMM analysis at outset, which would allow for within-class variation. Our results were stable across different sets of random starting values without these constraints, offering a much more representative model of patient experience of depressive symptom change.

A striking finding in the present study was the great contrast between trajectories of the two groups across both parts of the model. Compared to the continuous improvers, the halted improvers group actually showed significantly greater symptom reduction between the start of treatment and 18 week assessment. This indicates that clinicians may need to consider that a

fast reduction in depressive symptoms does not necessarily indicate a good prognosis. Further, treatment response cannot be prognostically assessed adequately before 18 weeks. The precise psychological treatment implications however cannot be determined as there was no placebo control group: we do not know whether this period of rapid improvement is due to the psychological therapy, a non-specific response to receiving assessment and treatment, or regression to the mean.

Previous studies reporting two classes describe their trajectory groups as either rapid and gradual responders, or responders and non-responders over each piece of the piecewise model[13,18]. One recent study of adolescents reported 3 trajectory classes with the 2 improving classes merging by end of treatment[21]. However, those studies were limited to short trial durations of 12 weeks or less and not able to assess longer-term outcome. The shape of the halted-improvers trajectory in this clinical study resonates with the symptom resurgence group reported by Briere and colleagues in their community-sample depression prevention study[17]. Their resurgence group showed a similar rapid initial improvement in symptoms, followed by a rapid decline in symptoms over time. While the slope value for the second section of our trajectory in the halted improvers did not reach significance with this cohort, visual inspection suggests this trend. Briere's samples was one third larger than ours[17] and therefore it is possible that a larger sample may have provided the power to detect a significant decline in condition in these patients. This trajectory shape suggests that different underlying therapeutic mechanisms may be activated in early and subsequent treatment responses. What factors account for the optimal break point of responding to be at the 18 week assessment is not clear. Longer-term follow-up is essential in future studies to disentangle group trajectory patterns and their related underlying mechanisms more accurately and to reveal the most valid prognostic markers for treatment response both early and later in follow up.

We attempted to identify baseline predictors of class membership as a parsimonious beginning to establishing candidate moderators for future mechanistic research. Univariate analyses suggested potential predictors of halted responding may include being female, higher current obsessionality, self-harm, more impairment and greater comorbidity at baseline. These findings are consistent with prior reports[45–48] except with regard to gender: one previous study associated female sex with a better outcome by end of treatment [17] . Interestingly, and in contrast to previous reports[23,47,49], suicidality and psychotic symptoms were not associated as potential predictors of our unfavourable trajectory class; halted improvers. The reason for this dissonance may be the shorter follow-up of those studies, especially terminating before 18 weeks. It is possible as well that these associations (except comorbidity) were due to confounding, given their non-significance in our multivariate model, or that our two-step approach decreased power to detect these differences.

Despite baseline univariate differences, a predictive model of demographic and clinical characteristics explained very little of the variance between these two classes (5.4%). Only comorbidities at baseline was retained as the only significant independent predictor with a small odds ratio (<2). This finding suggests that baseline demographic and clinical observations are insufficient in predicting depression symptom change over time. Including non-depressive symptoms in a more multidimensional longitudinal analysis of all symptoms to further disaggregate the behavioural phenotypes over time and determine if non depressive symptoms improve the classes, traejctories and break points in longitudinal course may be of value. Such an analysis may improve the signal for putative moderators for treatment response over that shown here.

In line with the Research Domain Criteria (RDoC)[50] framework, future work could also investigate the significance of biological predictors of such trajectories. Furthermore a much fuller inclusion of social environment factors should be considered in moderator and indeed

mediator models of treatment response. Incorporating qualitative data to incorporate patient experience of treatment may also help to explain some of the differences found in patterns of symptom change over time.

Finally, we considered whether the 2-class solution is in agreement with currently used measures based on self-report. The findings showed only moderate agreement between empirical and a priori definitions of symptom change, similar to those of previous studies[13,18]. The difference resides in 15% of improvers identified through GMM who are classified as non-responders by a priori methods. These differences demonstrate that the choice of methodology to determine outcome is particularly important if homogeneity is a goal for revealing the best group of non-responding individuals. The findings here have provided a clear cut homogeneous group (defined by narrow range of depressive symptom scores over time) of halted improvers by end of study, potentially predicted by higher depression scores at entry combined with halted improvement by 18 weeks.

This current work would also caution the use of cut-off or percentage change measures within 36 weeks of beginning therapy as it is possible that a significant percentage of potentially good responding patients might be misclassified to quickly as false negatives to treatment. Indeed the impact of misclassification in clinical trials has been highlighted in previous GMM work and research has suggested that current clinical response definitions may be too strict [13,18]. The current results support these prior findings from studies of adults, showing that the level of correspondence between these two contrasting approaches is currently inadequate. Until more is known about sub-classes of depression, researchers must take care in their choice of outcome measure and in particular to try to minimise a false negative result.

**Limitations**

We cannot use these findings to generalise to population based trials where recruitment takes place from schools, community settings or patients with distinct cultural differences to those in this UK NHS study. The current work could be improved in future studies through the collection of more time-points to allow for a more detailed investigation of trajectory shape. Additionally, GMM are large-sample statistical techniques, and while a sample of 465 is sufficient, a sample of 600 or more may have seen the emergence of a more stable 3-class model. The lack of a non-symptom driven 86-week outcome validator, such as interpersonal function, is another limitation of this paper. However, investigation of HONOSCA, a measure of function as well as psychiatric symptomatology, showed similar trends to MFQ (Supplementary material), which provides preliminary external validity for our findings. GMM trajectories in clinical populations with multiple symptom profile could utilise non-depressive symptoms in their longitudinal analyses. The predominance of females in the trial prevents an investigation of sex differentiated trajectories.

**References**


1. Avenevoli S, Knight E, Kessler RC, Merikangas KR. *Epidemiology of Depression in Children and Adolescents*. New York: Guilford Press; 2008.
2. Zisook S, Lesser I, Stewart JW, et al. Effect of Age at Onset on the Course of Major Depressive Disorder. *Am J Psychiatry*. 2007;164(10):1539-1546. doi:10.1176/appi.ajp.2007.06101757.
3. Fombonne E, Wostear G, Cooper V, Harrington R, Rutter M. The Maudsley long-term follow-up of child and adolescent depression. 1. Psychiatric outcomes in adulthood. *Br J psychiatry J Ment Sci*. 2001;179:210-217.
4. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major Depression in Community Adolescents: Age at Onset, Episode Duration, and Time to Recurrence. *J Am Acad Child Adolesc Psychiatry*. 1994;33(6):809-818. doi:10.1097/00004583-199407000-00006.
5. Depression in children and young people: identification and management | 1-Recommendations | Guidance and guidelines | NICE.
6. March J, Silva S, Petrycki S, et al. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*. 2004;292(7):807-820. doi:10.1001/jama.292.7.807.
7. Weisz J, Kuppens S, Ng M, et al. What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *Am Psychol*. 2017;72(2):79-117. http://psycnet.apa.org/fulltext/2017-07146-001.html. Accessed February 9, 2018.
8. Goodyer IM, Dubicka B, Wilkinson P, et al. A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial. *Health Technol Assess (Rockv)*. 2008;12(14).

9. Malhi GS, Byrow Y. Is treatment-resistant depression a useful concept? *Evid Based Ment Heal*. 2016;19(1):1-3. doi:10.1136/eb-2015-102299.

10. Zimmerman M, Martinez JA, Attiullah N, et al. Why Do Some Depressed Outpatients Who Are in Remission According to the Hamilton Depression Rating Scale Not Consider Themselves to Be in Remission? *J Clin Psychiatry*. 2012;73(6):790-795. doi:10.4088/JCP.11m07203.

11. Thibodeau MA, Quilty LC, De Fruyt F, De Bolle M, Rouillon F, Bagby RM. Latent classes of nonresponders, rapid responders, and gradual responders in depressed outpatients receiving antidepressant medication and psychotherapy. *Depress Anxiety*. 2015;32(3):213-220. doi:10.1002/da.22293.

12. Berlim MT, Turecki G. What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials. *Eur Neuropsychopharmacol*. 2007;17(11):696-707. doi:10.1016/j.euroneuro.2007.03.009.

13. Uher R, Muthén B, Souery D, et al. Trajectories of change in depression severity during treatment with antidepressants. *Psychol Med*. 2010;40(8):1367-1377. doi:10.1017/S0033291709991528.

14. Vitiello B, Emslie G, Clarke G, et al. Long-Term Outcome of Adolescent Depression Initially Resistant to Selective Serotonin Reuptake Inhibitor Treatment. *J Clin Psychiatry*. 2011;72(3):388-396. doi:10.4088/JCP.09m05885blu.

15. Fu CHY, Mourao-Miranda J, Costafreda SG, et al. Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression. *Biol Psychiatry*. 2008;63(7):656-662. doi:10.1016/j.biopsych.2007.08.020.

16. Ram N, Grimm KJ. Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *Int J Behav Dev*. 2009;33(6):565-576. doi:10.1177/0165025409343765.

17. Brière FN, Rohde P, Stice E, Morizot J. Group-based symptom trajectories in indicated

prevention of adolescent depression. *Depress Anxiety*. 2016;33(5):444-451. doi:10.1002/da.22440.

18. Gueorguieva R, Mallinckrodt C, Krystal JH. Trajectories of depression severity in clinical trials of duloxetine: insights into antidepressant and placebo responses. *Arch Gen Psychiatry*. 2011;68(12):1227-1237. doi:10.1001/archgenpsychiatry.2011.132.
19. Stulz N, Thase ME, Klein DN, Manber R, Crits-Christoph P. Differential effects of treatments for chronic depression: A latent growth model reanalysis. *J Consult Clin Psychol*. 2010;78(3):409-419.
20. Cuijpers P, van Lier PAC, van Straten A, Donker M. Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *J Affect Disord*. 2005;89(1-3):137-146. doi:10.1016/j.jad.2005.09.001.
21. Scott K, Lewis CC, Marti CN. Trajectories of Symptom Change in the Treatment for Adolescents With Depression Study. *J Am Acad Child Adolesc Psychiatry*. 2019. doi:10.1016/J.JAAC.2018.07.908.
22. Jeppesen P, Clemmensen L, Munkholm A, et al. Psychotic experiences co-occur with sleep problems, negative affect and mental disorders in preadolescence. *J Child Psychol Psychiatry*. 2015;56(5):558-565. doi:10.1111/jcpp.12319.
23. Wigman JTW, van Os J, Abidi L, et al. Subclinical psychotic experiences and bipolar spectrum features in depression: association with outcome of psychotherapy. *Psychol Med*. 2014;44(2):325-336. doi:10.1017/S0033291713000871.
24. Perlis RH, Uher R, Ostacher M, et al. Association Between Bipolar Spectrum Features and Treatment Outcomes in Outpatients With Major Depressive Disorder. *Arch Gen Psychiatry*. 2010;68(4):351-360. doi:10.1001/archgenpsychiatry.2010.179.
25. Wigman JTW, van Nierop M, Vollebergh WAM, et al. Evidence that psychotic symptoms are prevalent in disorders of anxiety and depression, impacting on illness onset, risk, and severity--implications for diagnosis and ultra-high risk research.

*Schizophr Bull*. 2012;38(2):247-257. doi:10.1093/schbul/sbr196.

26. Kelleher I, Connor D, Clarke MC, Devlin N, Harley M, Cannon M. Prevalence of psychotic symptoms in childhood and adolescence: a systematic review and meta-analysis of population-based studies. *Psychol Med*. 2012;42(9):1857-1863. doi:10.1017/S0033291711002960.

27. Goodyer I, Reynolds S, Barrett B, et al. Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiori. *Lancet Psychiatry*. 2017;4(2):109-119. doi:doi.org/10.1016/S2215-0366(16)30378-9.

28. Goodyer IM, Tsancheva S, Byford S, et al. Improving mood with psychoanalytic and cognitive therapies (IMPACT): a pragmatic effectiveness superiority trial to investigate whether specialised psychological treatment reduces the risk for relapse in adolescents with moderate to severe unipolar depres. *Trials*. 2011;12(1):175. doi:10.1186/1745-6215-12-175.

29. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association Press. Washington DC.; 2000.

30. Burleson Daviss W, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA. Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. *J Child Psychol Psychiatry*. 2006;47(9):927-934. doi:10.1111/j.1469-7610.2006.01646.x.

31. Goodyer IM, Reynolds S, Barrett B, et al. Cognitive-behavioural therapy and short-term psychoanalytic psychotherapy versus brief psychosocial intervention in adolescents with unipolar major depression (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled trial. *Health Technol Assess*. 2017;21(12):1-94. doi:10.3310/hta21120.

32. Reynolds CR, Richmond BO. What i think and feel: A revised measure of children's manifest anxiety. *J Abnorm Child Psychol*. 1978;6(2):271-280. doi:10.1007/BF00919131.

33. Bamber D, Tamplin A, Park RJ, Kyte ZA, Goodyer IM. Development of a short leyton obsessional inventory for children and adolescents. *J Am Acad Child Adolesc Psychiatry*. 2002;41(10):1246-1252. doi:10.1097/00004583-200210000-00015.

34. Gowers SG, Harrington RC, Whitton A, et al. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *Br J Psychiatry*. 1999;174(5):428-431. doi:10.1192/bjp.174.5.428.

35. Posner K, Brown GK, Stanley B, et al. The Columbia–Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults. *Am J Psychiatry*. 2011;168(12):1266-1277. doi:10.1176/appi.ajp.2011.10111704.

36. Vrouva I, Fonagy P, Fearon PRM, Roussow T. The risk-taking and self-harm inventory for adolescents: Development and psychometric evaluation. *Psychol Assess*. 2010;22(4):852-865. doi:10.1037/a0020583.

37. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36(7):980-988. doi:10.1097/00004583-199707000-00021.

38. Graham JW. Adding Missing-Data-Relevant Variables to FIMLBased Structural Equation Models. *Struct Equ Model*. 2003;10(1):80-100.

39. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067.

40. Gottfredson NC, Sterba SK, Jackson KM. Explicating the Conditions Under Which Multilevel Multiple Imputation Mitigates Bias Resulting from Random Coefficient-

Dependent Missing Longitudinal Data. *Prev Sci*. 2017;18(1):12-19. doi:10.1007/s11121-016-0735-3.

41. Muthén LK, Muthén BO. *Mplus User's Guide*. 8th ed. (Muthén & Muthén, ed.). Los Angeles, CA; 2017. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.
42. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas*. 1960;20(1):37-46. doi:10.1177/001316446002000104.
43. Kent L, Vostanis P, Feehan C. Detection of Major and Minor Depression in Children and Adolescents: Evaluation of the Mood and Feelings Questionnaire. *J Child Psychol Psychiatry*. 1997;38(5):565-573. doi:10.1111/j.1469-7610.1997.tb01543.x.
44. Wickrama KAS, Lee TK, O'Neal CW, Lorenz FO. *Higher-Order Growth Curves and Mixture Modeling with Mplus : A Practical Guide*.; 2016.
45. Goodyer IM, Herbert J, Secher SM, Pearson J. Short-Term Outcome of Major Depression: I. Comorbidity and Severity at Presentation as Predictors of Persistent Disorder. *J Am Acad Child Adolesc Psychiatry*. 1997;36(2):179-187. doi:10.1097/00004583-199702000-00008.
46. Wilkinson P, Dubicka B, Kelvin R, Roberts C, Goodyer I. Treated depression in adolescents: predictors of outcome at 28 weeks. *Br J psychiatry J Ment Sci*. 2009;194(4):334-341. doi:10.1192/bjp.bp.108.052381.
47. Curry J, Rohde P, Simons A, et al. Predictors and Moderators of Acute Outcome in the Treatment for Adolescents With Depression Study (TADS). *J Am Acad Child Adolesc Psychiatry*. 2006;45(12):1427-1439. doi:10.1097/01.chi.0000240838.78984.e2.
48. Nock MK, Joiner TE, Gordon KH, Lloyd-Richardson E, Prinstein MJ. Non-suicidal self-injury among adolescents: diagnostic correlates and relation to suicide attempts. *Psychiatry Res*. 2006;144(1):65-72. doi:10.1016/j.psychres.2006.05.010.
49. Souery D, Oswald P, Massat I, et al. Clinical Factors Associated With Treatment

Resistance in Major Depressive Disorder: Results from a European Multicenter Study. *J Clin Psychiatry*. 2007;68:1062-1070. https://s3.amazonaws.com/academia.edu.documents/42429351/Clinical_Factors_Associated_With_Treatme20160208-1018-1acok94.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1544967880&Signature=h2i%2FfxnT0knsuPhPDy3dU4SoIPQ%3D&response-content-disposition=inlin. Accessed December 16, 2018.

50. Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., Wang P. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *Am J Psychiatry*. 2010;167(7):748-751.

**Table One.** Model fit information for piecewise GMM-CIs

| Fit Statistics | | 1 Class | 2 Classes | 3 Classes | 4 Classes | 5 Classes |
|---|---|---|---|---|---|---|
| GMM-CIs | | | | | | |
| LL (No. of parameters) | | -10487.996 | -10454.836 | -10440.533 | -10431.253 | -10423.269 |
| AIC | | 21015.992 | 20957.672 | 20937.065 | 20926.506 | 20918.538 |
| BIC | | 21098.832 | 21057.081 | 21053.042 | 21059.051 | 21067.651 |
| Entropy | | 1 | .844 | .734 | .718 | .729 |
| Group size (%) | C1 | 465(100%) | 391(84.1%) | 329(70.7%) | 191(41.1%) | 219(47.1%) |
| | C2 | - | 74(15.9%) | 77(16.6%) | 161(34.6%) | 109(23.5%) |
| | C3 | - | - | 59(12.7%) | 57(12.3%) | 56(12.0%) |
| | C4 | - | - | - | 56(12.0%) | 54(11.6%) |
| | C5 | - | - | - | - | 27(5.8%) |

**Table Two.** Estimated and observed mean values for MFQ scores and observed mean percentage improvement in MFQ scores are given for both latent classes.

| | Class 1: Continual improvement (n=391) | | | | Class 2: Halted improvement (n=74) | | | |
|---|---|---|---|---|---|---|---|---|
| | MFQ scores | | | | MFQ scores | | | |
| Assessment Point in average weeks from baseline | Estimated | Weighted Estimates | Observed | % observed improvement from baseline | Estimated | Weighted Estimates | Observed | % observed improvement from baseline |
| 0 | 44.828 | 44.810 | 44.774 | | 51.638 | 51.721 | 52.096 | |
| 12 | 37.073 | 34.649 | 34.623 | 22.672 | 41.090 | 38.191 | 38.420 | 26.252 |
| 18 | 32.881 | 32.752 | 32.689 | 26.991 | 35.390 | 36.134 | 36.558 | 29.826 |
| 43 | 25.877 | 23.938 | 25.920 | 42.109 | 39.232 | 38.243 | 38.669 | 25.774 |
| 60 | 23.039 | 22.291 | 22.224 | 50.364 | 40.789 | 39.034 | 39.835 | 23.535 |
| 95 | 17.247 | 17.797 | 17.673 | 60.528 | 43.966 | 44.978 | 46.357 | 11.016 |

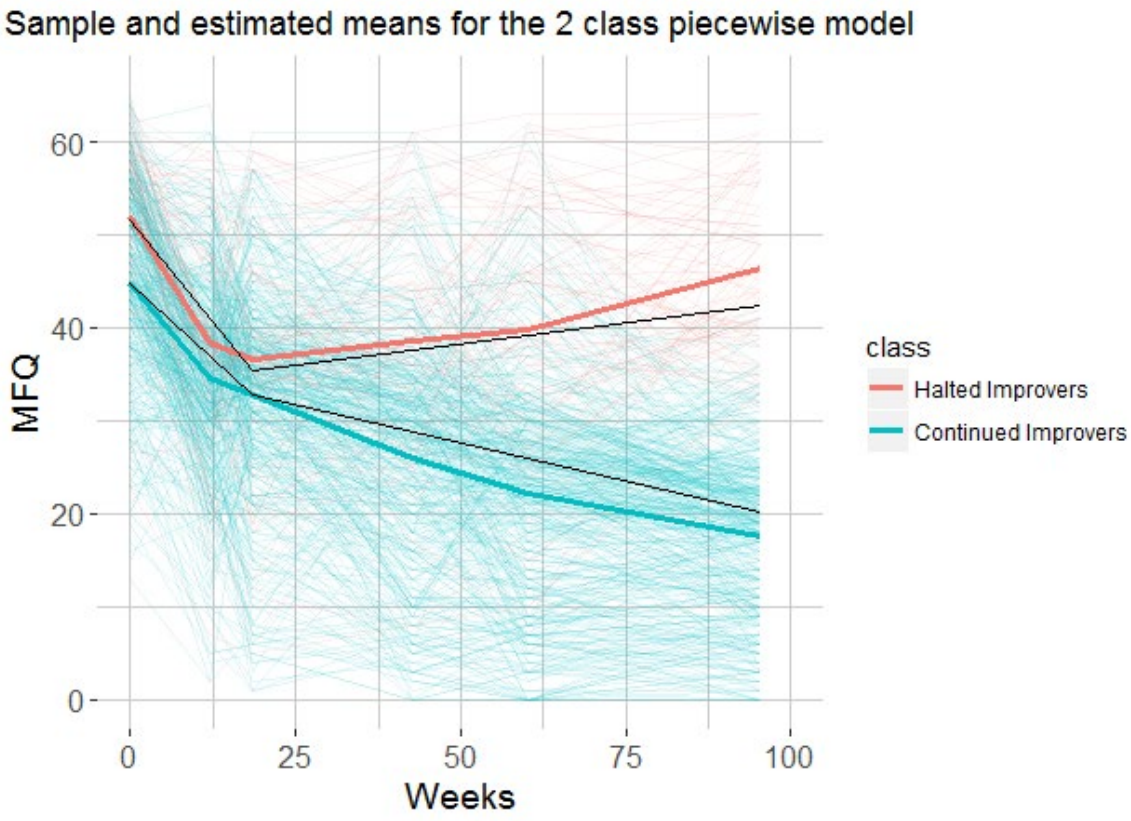
**Table Three.** Characteristics of subjects following the two latent trajectories

| | Class 1: Continued improvers (n=391) | | Class 2: Halted improvers (n=74) | | Comparison | |
|---|---|---|---|---|---|---|
| | Mean(n) | S.D(%) | Mean(n) | S.D(%) | $X^2$/t/W | p |
| Demographics | | | | | | |
| Female | 285 | 72.8% | 63 | 85.1% | 4.955 | .026 |
| Age | 15.6 | 1.4 | 15.7 | 1.3 | 0.459 | .647 |
| Region | - | - | - | - | 2.035 | .361 |
| East Anglia | 161 | 41.2% | 24 | 32.4% | - | - |
| North London | 105 | 26.9% | 22 | 29.7% | - | - |
| North-West England | 125 | 32.0% | 28 | 37.8% | - | - |
| Ethnicity(white) | 325 | 83.1% | 65 | 87.8% | 1.024 | .312 |
| Index of multiple deprivation(IMD) | 23.4 | - | 27.7 | - | 13446 | .336 |
| Baseline clinical characteristics | | | | | | |
| RCMAS | 40.7 | 7.3 | 42.3 | 6.7 | 1.863 | .065 |
| LOI | 9.6 | 5.1 | 11.8 | 5.6 | 3.124 | .002 |
| Suicidal thoughts | 345 | 88.2% | 69 | 93.2% | 1.600 | .206 |
| Suicidal attempts | 131 | 33.5% | 28 | 37.8% | 0.519 | .471 |
| NSSI | 218 | 55.8% | 53 | 71.6% | 6.443 | .011 |
| HONOSCA (available for 435 patients) | 18.3 | 6.0 | 19.9 | 6.3 | 2.018 | .046 |
| Comorbidity^ | - | - | - | - | 10.20 | .006 |
| 1 | 121 | 30.9% | 26 | 35.1% | - | - |
| 2 | 59 | 15.1% | 18 | 24.3% | - | - |
| 3 | 5 | 1.3% | 3 | 4.1% | - | - |
| 4 | 0 | 0% | 1 | 1.4% | - | - |
| Psychotic symptoms | - | - | - | - | 2.024 | .363 |
| Subthreshold | 87 | 23.6% | 16 | 22.5% | - | - |
| Threshold | 32 | 8.8% | 10 | 14.1% | - | - |
| Treatment characteristics | | | | | | |
| Treatment arm: | - | - | - | - | 2.463 | .292 |
| BPI | 127 | 32.5% | 28 | 37.8% | - | - |
| CBT | 127 | 32.5% | 27 | 36.5% | - | - |
| STPP | 137 | 35.0% | 19 | 25.7% | - | - |
| Baseline SSRI prescription | 87 | 22.3% | 10 | 13.5% | 2.877 | .090 |

IMD: *Median reported for non-parametric tests:

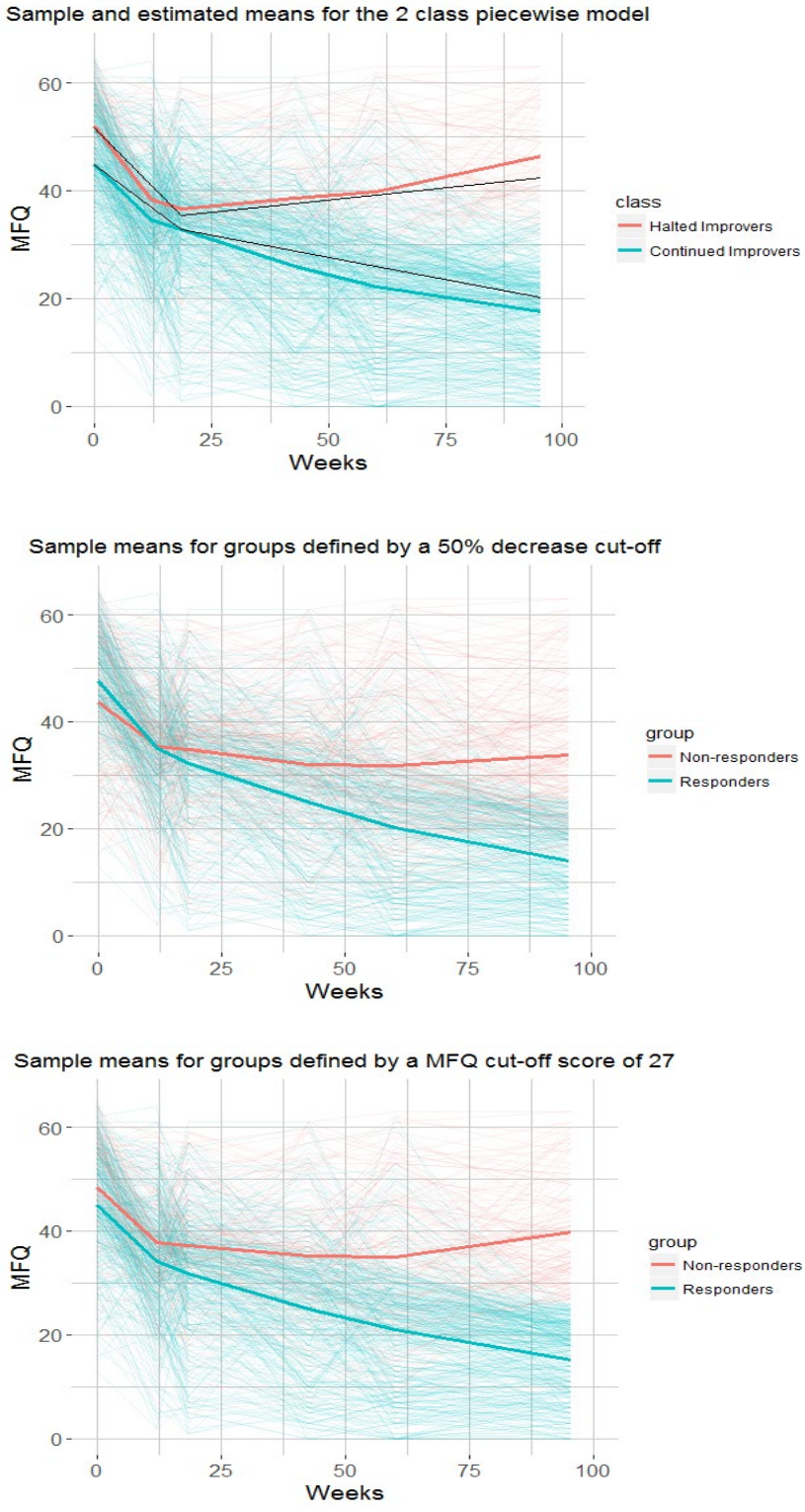
^due to insufficient cell size, variable was recorded as 0,1 and 2+ to meet assumptions of chi-square test.

Figure One




Sample (shown in red/blue) and estimated (shown in black) means for the 2 class piecewise growth mixture model. Class 1 reveal a continued improver class, n=391(84.1%) of the population. Class 2 reveal halted improver class , n= 74(15.9%) of the population. Behind plots every individual patient's trajectory, colour coded to their respective classes.

Figure 2




Sample means for the three categorical approaches. (a) Trajectory classes. (b) Percentage reduction. (c) Clinical cut-off. Behind plots every individual patient's trajectory, colour coded to their respective classes.