Understanding mental health treatment effectiveness in young people

Sharon Ann Scovil Neufeld

Department of Psychiatry

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Jesus College

September, 2019

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 dissertation 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 Degree Committee of Clinical Medicine and Clinical Veterinary Medicine.

.

Sharon Neufeld, September 2019

Understanding mental health treatment effectiveness in young people

by Sharon Ann Scovil Neufeld

Mental disorders are one of the leading causes of non-communicable disease burden worldwide, with distress emerging as a common factor among such disorders. Most lifetime mental disorders emerge during adolescence and young adulthood, yet no studies have robustly assessed changes in common distress over time during this broad developmental epoch. Common distress is particularly pertinent when assessing mental health treatment effectiveness, as comorbidity rates are higher amongst those who seek treatment. In this dissertation, I seek to redress the paucity of evidence regarding mental health treatment effectiveness in young people. Herein I longitudinally validate a common distress factor, enabling me to use specific and broad mental health outcomes to assess the effectiveness of treatment-as-usual mental health services in two separate samples.

In chapter 2, I use a community sample of adolescents to test the relationship of age 14 service contact on depressive symptoms by age 17. I found that 14-year-old adolescents who had contact with mental health services in the past year had a greater decrease in depressive symptoms than those without contact. By age 17 years, the odds of reporting clinical depression were higher in individuals without contact than in service users who had been similarly depressed at baseline. In addition to propensity score weighting to adjust for participants' initial likelihood to access services, I used a clinically relevant cutoff and adjusted for a wide range of time-varying confounding variables. These adjustments give greater confidence than previous studies to the notion that mental health service contact is related to meaningful improvements in subsequent mental health. Policy implications of these findings are addressed.

In chapter 3, I develop a transdiagnostic measure of mental health to be used to assess treatment effectiveness with relevancy across multiple disorders. This transdiagnostic measure is the general factor (common distress or *p*) from a bifactor model of 118 self-report items from previously validated measures of depression, anxiety, behavioural problems, obsessions, psychotic symptoms, wellbeing, and self-esteem. Longitudinal and gender measurement invariance and reliability of this model was demonstrated in a sample of 14- to 24-year olds assessed annually three times. Predictive validity of the general and specific factors was demonstrated using an extensive set of external variables covering factors such as social environment, personality, and risk behaviour. Accelerated growth modelling revealed developmental changes in the factors from ages 14-27 largely consistent with epidemiological patterns of the associated disorders.

I use this validated distress factor in chapter 4 to further test the effectiveness of mental health treatment-as-usual, in a broader age range with a broader outcome than the adolescent sample. Family functioning and friendship support were also explored as potential mechanisms of action. Treatment was related to decreased distress over two years, in unadjusted and adjusted models. In order to propensity weight, I focused analyses on those above the population mean in baseline distress, which contained 85% of young people reporting treatment for a mental disorder.

Such individuals were more likely to have required treatment, so treated and untreated groups are more meaningful and comparable. In propensity score weighted models, treatment was related to decreased subsequent distress. Treatment was also related to improved family functioning over two years, but only in adolescents aged 14-18, more likely to be living with their family. In this younger group, the best fitting model revealed distress as a significant mediator: the majority of improvements in family functioning occurred through improvements in distress. However, treatment was not related to a change in friendship support over time. Thus, not only is mental health treatment-as-usual related to a significant decline in young people's distress, but also improvements in adolescent family functioning.

Finally, in chapter 5 I discuss the broader implications of the findings. Analyses in two separate community samples of young people strongly support the notion that treatment-as-usual is related to improvements in both specific and broad measures of mental health. Such treatment-related improvements in mental health appear to also yield benefits in family functioning in adolescents. Findings as a whole argue for increased access to mental health services.

Publications arising from this dissertation

Chapter 2

Neufeld SAS, Dunn V, Jones PB, Croudace T, & Goodyer IM. (2017). Reduction in adolescent depression following contact with mental health services: a longitudinal cohort study in the United Kingdom. ***Lancet Psychiatry, 4(2):120-127.***

Contribution: SASN coded health services data, and SASN and VJD resolved any queries. SASN analysed the data, and SASN, IMG, PBJ, and VJD interpreted the data. SASN drafted the manuscript, and IMG, PBJ, and VJD criticised the manuscript for intellectual content.

Neufeld SAS, Jones PB, Goodyer IM. (2017). Reduction in adolescent depression following contact with mental health services: a longitudinal cohort study in the United Kingdom-authors' reply. ***Lancet Psychiatry, 4(5):e8-9.***

Contribution: SASN drafted the reply, and IMG and PBJ criticised the manuscript for intellectual content.

Neufeld SAS, Jones PB, Goodyer IM. (2017). Child and Adolescent Mental Health Services: Longitudinal data sheds light on current policy for psychological interventions in the community. ***Journal of Public Mental Health, 16(3):96-99.***

Contribution: SASN drafted the manuscript, and IMG and PBJ criticised the manuscript for intellectual content.

Acknowledgements

I've been thrilled to have a wonderful set of experts at my fingertips throughout my time spent doing this work – people who are more than able to scientifically motivate, but are also happy to let me explore my ideas.

Paul's direct clinical experience and inherent understanding of young people's mental health services has greatly helped me to broaden my own understanding. The ability to discuss ideas and challenges with Paul when needed has been invaluable. I am so grateful for Peter's continued staunch support - your broad but deep knowledge of this ever-expanding field is always thought-provoking.

Of course, no acknowledgement would be complete without giving a big thank you to Ian Goodyer, who has been instrumental in setting the wheels turning for many of these chapters.

All three of you have had your turn listening patiently to me while I explain yet another technique I'm learning. I've really appreciated the wise advice you've given in helping hone my ideas. This work would not have been possible without your kind and able mentorship.

I am also very grateful to my other colleagues. Val Dunn's great knowledge of all the details of ROOTS helped immensely with coding the health services data. Michelle St Clair has always been a great sounding board and generous with her time. Thanks also go to Jeannette Brodbeck, for introducing me to new methods, being keen to hear the details of any analytical challenges I was facing, and always curious about ideas. I cannot stop here without thanking Adisa Broadhurt, who has been so patient when I've come to ask her yet another question about an unusual circumstance for the rather unusual PhD student I've been (PhD student *and* Research Associate *and* mother of three…?).

Finally, I would like to thank my family who have been so supportive and patient as I have been finishing this thesis. To Wendell, who frequently showered me with kisses while I was writing; to Miles, who goaded me out to play football or badminton for breaks; and to Cayley, for all our talks as you embark on your own journey of adolescence. Most of all, thanks go to Jerome, who has taken on extra chores without complaint, been a Pied Piper with the children as needed, and has been so encouraging of me taking this on. Having such a vibrant home to return to is certainly an incentive to work hard!

I look forward to what lies ahead and want to thank everyone for enabling me to do this research. I very much hope that something of my work below will help connect young people to the services that so many dearly need.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

1.1 Epidemiology of mental disorders

The prevention of mental disorders is a pressing issue of our day. Mental and substance use disorders are one of the leading causes of non-communicable disease burden worldwide, following only behind cardiovascular diseases and cancers, and representing nearly 200,000 Disability-Adjusted Life Years (DALYs (Kyu *et al.*, 2018)). DALYs are a composite measure of disease burden which capture premature mortality, prevalence, and severity of ill health). Mental disorders show no signs of abating: globally, DALYs for mental disorders increased by 14% between 2007 to 2017 (Kyu *et al.*, 2018). Across sociodemographic levels, DALYs for mental disorders have been shown to first emerge at discernable levels from ages 5-9, and burgeon from ages 10-24, where mental disorders were the highest contributer to DALYs across all diseases (Kyu *et al.*, 2018). This underscores the importance of prevention of mental disorders, particularly in this age range.

Age of onset data from the World Health Organization mental health surveys concurs with the above increased DALYS related to mental disorders in young people. Across the 28 countries studied, half of all lifetime mental disorders emerged by the mid-teens and three-quarters by the mid-twenties (Kessler *et al.*, 2007). These figures were based predominantly on cross-sectional retrospective data, which are biased by poor recall and cohort differences (Moffitt *et al.*, 2010). However, findings from the Dunedin birth cohort study, a longitudinal epidemiological study, also demonstrate an increase in prevalence of mental disorder across adolescence (Newman *et al.*, 1996). With disorder rates obtained every two to three years from ages 11 to 21, significant increases were seen from ages 13-15, and from ages 15-18, with rates remaining high into the twenties. (Both studies used the diagnostic classification system of the Diagnostic and Statistical Manual of Mental Disorders (DSM), commonly used in many studies of mental disorder (Clark *et al.*, 2017). Internationally, the age of onset of particular disorders is consistent (Kessler *et al.*, 2007). Phobias and impulse-control disorders manifest earlier (median age of onset for both disorder groups is approximately 7-14, with 25% emerging at age 4 (Kessler *et al.*, 2007)), explaining the first discernable DALYs being exhibited at ages 5-9 (Kyu *et al.*, 2018). Other anxiety disorders and depressive disorders manifest later: 25% have emerged by the mid-teens, with the lower bound to the median being age 25 (Kessler *et al.*, 2007). Most psychoses become apparent in the third decade of life, with a median onset in the early 20s. As mental disorders emerge in both adolescence and emerging adulthood (ages 18-25 (Arnett, 2000) or ages 18-29 (Arnett, Žukauskiene and Sugimura, 2014)) the subsequent discussion will encompass these age ranges. (The term “young people” will be used when I collectively refer to both adolescents and emerging adults.)

The onset of new mood and anxiety disorders in young people is particularly troubling due to the high prevalence of these disorders. Internationally, anxiety disorders have been shown to be consistently the most prevalent of mental disorders followed by mood disorders, with substance use disorders and impulse disorders being less prevalent (The WHO World Mental Health Survey Consortium, 2004). This finding was confirmed by a systematic review and meta-analysis of all population surveys published over 33 years focusing on the first three disorders (Steel *et al.*, 2014). This study revealed pooled prevalence estimates for anxiety and mood disorders at 6.7% and 5.4% of the general population, respectively, followed by 3.8% for substance use. This relatively higher increased prevalence has resulted in mood and anxiety disorders being termed 'common mental disorders' (CMD). When DALYs are considered, depressive disorders rank above anxiety disorders, due to the former being more functionally disabling and having a higher association with premature death (Kyu *et al.*, 2018). Together these two disorder categories represent 57% of DALYs across all 10 main categories of mental disorders.

Coupled with a high prevalence, CMD also exhibit strong continuity from adolescence to adulthood. Homotypic continuity exists when symptoms of the same mental disorder persist over time. This continuity could be due to chronic or recurrent symptoms (Ormel *et al.*, 2015). In a representative cross-sectional sample of 1584 adolescents, relatively high 12-month to lifetime prevalence ratios (59% and 66% for mood and anxiety disorders respectively) indicated substantial disorder persistence (Ormel *et al.*, 2015). Longitudinal cohort studies have revealed that of those with a CMD diagnosis in emerging adulthood, 60%-74% had a prior diagnosis in adolescence (Newman *et al.*, 1996; Patton *et al.*, 2014). This persistence over time of the same mental disorder has been termed homotypic continuity. Homotypic continuity is a shown even further into adulthood in findings from a prospective British birth cohort. Those suffering from CMD in their mid-teens were more likely to also exhibit CMD at ages 36 and 43; those with persistent CMD across ages 13 and 15 had higher rates of CMD at ages 36 and 43, and were even at fourfold risk of CMD at age 53 (Colman *et al.*, 2007).

Rates of homotypic continuity vary depending on the sample and disorder studied. A review of longitudinal data from clinical and community samples revealed that 40-70% of depressed adolescents are diagnosed with major depressive disorder in adulthood (Rutter, Kim-Cohen and Maughan, 2006). Not included in the above review, a prospective sample of young people who were diagnosed with a childhood or adolescent depressive disorder concurs with the review findings - 40% experienced recurrent depression and 18% experienced persistent depression at age 23 (Dunn and Goodyer, 2006). However, a longitudinal cohort study of nearly 1000 young people revealed more modest rates of continuity, with 23% reporting two or more episodes of major depression (MDD) when repeatedly assessed from ages 16-21 (Fergusson, Boden and Horwood, 2007). For adults who had experienced remission from MDD, similar recurrence rates were found (27% and 34% for those who had been treated in specialist mental healthcare and primary care

respectively, (Hardeveld *et al.*, 2013). For adults with an episode of anxiety disorder, recovery rates 12 years later varied depending on type of anxiety disorder (social phobia had the lowest recovery rate at 37% and panic disorder without agoraphobia the highest recovery rate at 82% (Bruce *et al.*, 2005)). For those who recovered from their intake episode, the 12-year probability of recurrence varied from 39% to 58% (Bruce *et al.*, 2005). Similarly, a sample of young adolescents in the general population revealed differential rates of homotypic continuity over two years depending on the anxiety disorder studied, with social phobia revealing the highest continuity (Ferdinand *et al.*, 2008). While the above studies reveal a range of rates in homotypic continuity for CMD, there is no doubt that those with persistent CMD in adolescence are at much greater risk of subsequent CMD (Colman *et al.*, 2007), underscoring the importance of early intervention.

Alongside this homotypic continuity there has also been substantial heterotypic continuity observed in young people, where one disorder predicts another disorder at a later timepoint. Following a substantial body of research on the subject (for a review see (Rutter, Kim-Cohen and Maughan, 2006)), recent cohort studies have assessed homotypic and heterotypic continuity across many mental disorders. Evidence for both has been demonstrated from childhood to early adolescence (Shevlin, McElroy and Murphy, 2017), and in adulthood (Lahey *et al.*, 2014). Findings indicate that heterotypic continuity increases with age. In the child-to-adolescent sample, within broad categories of mood, anxiety, and disinhibitory disorders, eight specific disorders predicted, and were predicted by each other (Shevlin, McElroy and Murphy, 2017). Across these broad categories, mood and disinhibitory disorders predicted anxiety disorders. In adulthood, even more heterotypic continuity was demonstrated across 10 specific disorders. A specific disorder from each category of mood, anxiety, and disinhibitory disorders predicted (and was predicted by) another specific disorder in each of the other categories. Less frequent disorders also exhibit heterotypic continuity: children with strong psychotic symptoms at age 11 had higher rates of PTSD, anxiety, depression, and substance dependence by age 38, and were more likely to have attempted or completed suicide (Fisher *et al.*, 2013). Not only are these troubling outcomes for people with mental disorders, but this heterotypic continuity underscores the overlapping nature of mental disorders, raising questions about their distinctness.

1.2 Are there better ways to construe mental health?

Mental disorders are now understood to be complex combinations of psychological problems which are inherently dimensional, as opposed to categorical, discrete entities (Clark *et al.*, 2017). In addition to the heterotypic continuity discussed above, this point is underscored by the ubiquity of individuals with one mental disorder who simultaneously meet criteria for another mental disorder. High rates of comorbidity in mental disorders have been demonstrated in adolescent (48% (Kessler *et al.*, 2012)), emerging adult (47%, (Newman *et al.*, 1996)), and adult (45%, (Kessler, Chiu, *et al.*, 2005)) populations. However, as alluded to in the heterotypic continuity findings discussed above,

some of these disorders are more systematically related than others. There has long been evidence that mental disorders form higher-order factors of internalizing and externalizing – with negative affectivity being common to the depressive and anxiety disorders which load onto the former factor, and disinhibition being common to the antisocial behaviour disorders, ADHD, and substance use disorders which load onto the latter factor (Achenbach, 1966). This dual structure has been confirmed through many subsequent studies and indeed through meta-analyses of the data from many of these international studies (Krueger and Markon, 2006; Kessler *et al.*, 2011). More recently, the importance of a distinct third and potentially more severe dimension, that of a thought disorder factor, has emerged when a broader array of disorders have been included alongside those which fall into internalizing and externalizing dimensions (Kotov, Chang, *et al.*, 2011). This factor has comprised diagnoses of mania, psychosis, as well as paranoid, schizoid, and schizotypal personality disorders (Kotov, Ruggero, *et al.*, 2011). The inclusion of this factor is particularly important given the young age of onset and disabling nature of such disorders (Anderson, 2019), as well as the finding that psychotic disorders exhibit a greater incidence than previously thought (Jongsma *et al.*, 2019).

However, there are further ways to extend how mental health can be modelled. Lahey et al made the important point that even these higher order factors remain correlated, which suggests an even higher-order factor overlays these dimensions (Lahey *et al.*, 2012). Their analyses revealed that the best-fitting model was a bifactor model, which included a general psychopathology factor on which all of the mental disorders loaded. These findings were swiftly confirmed in another sample, with the general factor being coined '*p'*, to represent a general liability for psychopathology (Caspi *et al.*, 2014). However, being based on DSM diagnoses, these models are grounded on assumptions made by clinical experts as to how symptoms are aligned with respect to diagnoses (Markon, 2010; Krueger and Markon, 2011). Circumventing this issue, an additional layer of dimensionality can be added by modelling symptoms directly. This also allows subthreshold symptoms, which may still be impairing (Rutter, Kim-Cohen and Maughan, 2006), to inform the model. Symptoms have indeed been shown to have greater ability to predict clinical outcomes than diagnoses alone (Brittain *et al.*, 2013). Further, models of mental health can extend beyond symptoms and include positive dimensions, acknowledging that mental health is more than the absence of mental illness (Keyes, 2002). Such an endeavour may capture the concept of 'enduring mental health' enjoyed by those who remain free of mental disorder from birth to mid-life (Schaefer *et al.*, 2017). Regardless of the specific details of how mental health is modelled, the resultant factors must be externally validated in order to ascertain their utility moving forward. It is to these predictors of mental health I turn to next.

1.3 Key predictors of mental health in young people

The deleterious effect of childhood trauma and poor family relationships on subsequent mental health have been well-documented. A systematic review of studies of family relationships assessed during childhood, with psychiatric outcomes obtained 10 years later and adjusted by confounding, helps clarify the aspects of parenting which have the most prolonged effect on young people (Weich *et al.*, 2009). Severe child abuse and/or neglect by family members predicted later depression and PTSD, whereas harsh maternal discipline and/or maternal unresponsiveness predicted anxiety, PTSD, and adolescent suicide attempts (other disorders were not measured). A more recent study has confirmed and extended these findings by prospectively measuring childhood victimization from ages 5-12 across many domains (Schaefer *et al.*, 2018). Victimization encompassed physical maltreatment, neglect, exposure to domestic violence, as well as sexual, peer/sibling, internet/mobile phone, and crime victimization. By age 18, each of these types of victimization individually predicted internalizing, externalizing, and thought disorder factors which were based on symptoms of mental disorders. There was a linear relationship with the extent of poly-victimization and the levels of symptoms on each of these outcomes.

Other aspects of environmental adversity also have a clear link with subsequent mental disorder. Lower socio-economic status in childhood has been related to a higher incidence of internalizing and externalizing behaviours and thought problems in adolescence (Wadsworth and Achenbach, 2005). Regardless of timing from infancy to early adolescence, childhood adversity has been prospectively associated with increased rates of mood, anxiety, substance misuse, psychotic disorders, and ADHD in older adolescents and emerging adults (Björkenstam *et al.*, 2016). Adversities were distinct from victimization, ranging from those which disrupted the family structure (familial death, parental separation), were based on parental dysfunction (parental psychiatric disorder, substance misuse, or severe criminality), or household factors (received public assistance, residential instability). Risk for psychiatric disorder increased twofold for those with three or more adversities. Stressful life events experienced after childhood (eg. death of a close family member or friend, breakup of a romantic relationship, contact with police) have also been associated with mental illness. Such events experienced by adolescents and emerging adults have been shown to predict both anxiety and depressive disorders ten years later (Asselmann *et al.*, 2015). Stressful life events have also been correlated with higher levels of adolescent psychotic experiences (Shakoor *et al.*, 2016) and externalizing behaviours (Amone-P'Olak *et al.*, 2009).

Social support has been shown to have a protective role in the development and progression of mental disorders in young people and may act to buffer the effect of stressful events on subsequent psychosocial functioning (Cohen and Wills, 1985). In adolescents, family support has been shown to mediate the effect of childhood family adversities on subsequent depressive symptoms, whilst friendships mediated the effect of relational bullying on depressive symptoms (van Harmelen *et al.*, 2016). In a sample of adolescents and emerging adults, social support from friends

but not family was related to improved subsequent resilient functioning – that is, better psychosocial functioning than expected based on recalled childhood family experiences (Van Harmelen *et al.*, 2017). However, parental support has been shown to be related to emerging adults' improved emotional adjustment over time (Wintre and Yaffe, 2000; Levitt, Silver and Santos, 2007). Thus, both sources of social support appear to be important in predicting the subsequent mental health of young people.

Personality factors may also predispose young people to developing a mental disorder. A meta-analysis of personality traits and depression, anxiety, and substance use disorders in adults revealed that neuroticism (the tendency to exhibit a diffuse negative emotionality) was strongly related to all mental disorders (Kotov *et al.*, 2010). Adolescent neuroticism has been related to externalizing disorders (Lahey *et al.*, 2008) and subsequent psychotic symptoms (Goodwin, Fergusson and Horwood, 2003). Impulsivity in adolescents has been shown to be predictive of drinking problems, marijuana and drug use, as well as conduct and hyperactivity problems 18 months later (Castellanos-Ryan *et al.*, 2013). However, a prosocial disposition (eg., being inclined towards helping and sharing, and showing concern for others) was protective of externalizing disorders and depression in adolescents (Lahey *et al.*, 2008).

1.4 Interventions to improve mental health

Considering these factors which contribute to young people's reduced mental health, intervention in this age group is imperative. In addition to the likelihood of continuity of a mental disorder into adulthood, there are broader consequences of an early onset to mental disorders. Young people with a mental disorder report interference in their work and daily activities (Newman *et al.*, 1996), which means they are less likely to be in employment, education, or training, and more likely to be on employment benefits (Knapp, 2014). According to peers, adolescents with a mental disorder exhibit disturbed behaviour, and official records reveal they have more contact with criminal justice services and higher rates of criminal convictions (Newman *et al.*, 1996; Knapp, 2014). However, systemic impairment extends well into mid-life. The presence of a moderate or severe psychological problem by the age of 16 has been shown to predict decreased probability of working and a decreased family income at ages 23, 33, 42, and 50, as well as a decreased likelihood of cohabiting or being married at the latter three ages (Goodman, Joyce and Smith, 2011). Further, childhood psychological problems predicted reduced cognitive functioning, agreeableness, conscientiousness, and emotional stability at age 50. In comparison, physical health problems by age 16 did not exert the same toll on income. This "long shadow" cast by early onset mental health problems highlights the large societal benefit of intervention in young people.

Great strides have been made over the past several decades in developing psychological treatments, which are now considered to be the frontline intervention for many mental disorders (Clark, 2018). Particularly for depression and anxiety disorders, cognitive behavioural therapies

(CBT) have been shown to have enduring effects on symptoms following the cessation of treatment (Hollon, Stewart and Strunk, 2006; Goodyer *et al.*, 2017), although for depression, other psychotherapies have been shown to be equally effective (Goodyer *et al.*, 2017). While psychotrophic medications are effective at reducing symptoms during the course of treatment, outcomes following cessation of treatment in randomized controlled trials have been shown to be poorer than psychotherapy (Hollon, Stewart and Strunk, 2006). CBT is thought to alter more stable cognitive dispositions and underlying information-processing styles in a way that medication cannot (ibid). As the most widely studied psychotherapy (Hofmann *et al.*, 2012), CBT enjoys broad support as a treatment, with specific forms of CBT being recommended for adolescents and adults with depression and anxiety disorders (NICE, 2014b, 2017, 2019a) and psychosis (NICE, 2014c, 2016), as well as antisocial and borderline personality disorders in adults (NICE, 2015).

The importance of psychotherapy is reflected in best-practice guidelines established by the UK's National Institute for Clinical Excellence (NICE) guidelines. Since 2004, this independent organization has been systematically reviewing the evidence for effectiveness of a variety of interventions for mental disorders (Clark, 2018). NICE recommends a stepped-care approach, where the least intrusive, most effective intervention should be offered first. For the more common mental disorders, lower-intensity psychological interventions (such a group therapy) are recommended for anxiety and mild manifestations of depression; for more moderate forms of these disorders, a higher-intensity psychological intervention (such as individual therapy) is recommended, potentially in combination with medication (NICE, 2014b, 2017, 2019a).

Launched in 2008, England's Individual Access to Psychological Therapies (IAPT) programme aims to substantially increase access to evidence-based psychological treatment for young people and adults ages 16 and above suffering from various anxiety and depressive disorders (Clark, 2018). The most recent adult psychiatric morbidity survey in the UK revealed that in 2014, treatment rates for those with CMD had indeed increased since 2007 (Mcmanus *et al.*, 2016). However, both surveys revealed that counselling was less likely to be received than medication, and rates of increase were greater for medication than counselling. Nonetheless, the number of people treated in the IAPT programme continues to rise, with approximately two-thirds showing reliable improvement in 2017, the most recent data available (Clark, 2018). While exceeding its current aim of treating 15% of adults with CMD each year (Baker, 2018; Clark, 2018), the aim is to increase this to at least 25% by 2020/21 (NHS England, 2015a). Wait times have also decreased over previous years and are better than targets: 87% waited less than 6 weeks to enter treatment (Baker, 2018). One of the advantages of the IAPT program is that it captures session-by-session data on clinical outcomes (Clark, 2018), and thus such progress is able to be monitored.

1.5 Access to services

While the above figures are encouraging, there is still substantial unmet need with respect to emerging adults' access to mental health treatment. Even if NHS England's goal of treating 25% of those ages 16 and above with CMD by 2020/21 is met, the remaining 75% still represent 4.5 million people with CMD who are not accessing services (NHS England, 2015a). There is also evidence that emerging adults may have greater difficulty with retention in therapy than their older counterparts. Using IAPT outcome data from 2010-2011, individuals in the third decade of life had the highest number of referrals into IAPT of any decade, but were the least likely to attend or complete therapy, and had the lowest levels of reliable improvement (Pettit *et al.*, 2017). These observed difficulties with treatment in emerging adults are consistent with prior literature (Pottick *et al.*, 2014), and may be due to their still-developing executive functioning skills (Giedd, 2004; Liston *et al.*, 2006), which could contribute to more missed sessions and dropout (Henin and Berman, 2016). Having often moved away from home, emerging adults typically have less frequent contact with parents (Arnett, 2000), key sources of support for engaging in and persisting with treatment. Further, the artificial boundary imposed by many mental health systems compelling young people to transition at age 18 from child to adult services can create a "yawning gap" in access to services (McGorry, 2013). Movement is being made to redefine services for young people up until age 25, but the examples are still few and far between (McGorry, Bates and Birchwood, 2013). In these aspirational and existing services there needs to be greater awareness of the particular challenges young people face in order to more fully meet their mental health needs.

Access to mental health services for adolescents needs considerable attention. A recent report by the Children's Commissioner for England highlights that nearly 15 times as many funds are spent on adult mental health as compared with children's mental health (Children's Commissioner, 2018). Child and adolescent mental health services (CAMHS) typically serve young people 18 and under (Singh *et al.*, 2010). While an IAPT programme for children and young people exists as part of CAMHS, the mandate is solely to increase the quality of and improve access to existing mental health services as opposed to developing a new tier of care, as is the case for adult IAPT (Wolpert *et al.*, 2016; Clark, 2018). Nonetheless, recently there has been an increase in the numbers accessing CAMHS (Children's Commissioner, 2018), in line with objectives to expand these services (NHS England, 2015a). However, clear barriers to treatment still exist. Based on two NHS datasets from 2017-18, less than a third of those referred to CAMHS were treated within the year (Children's Commissioner, 2018), after an average wait times of 9 weeks until assessment (NHS Benchworking Network, 2018). The above figures hide the unmet need in young people with mental health problems who do not make contact with services. Internationally, incidence rates of adolescents with a mental disorder range from 12-25% (Ford, Goodman and Meltzer, 2003; Canino *et al.*, 2004; Green *et al.*, 2005; Farbstein *et al.*, 2010; Merikangas *et al.*, 2010; Kessler *et al.*, 2012; Vicente *et al.*, 2012; Paula *et al.*, 2014; Johnson *et al.*, 2016; Sadler *et al.*, 2018), yet generally,

rates of mental health service contact fall well below these levels – representing 34-56% of those with a mental disorder (Canino *et al.*, 2004; Mansbach-Kleinfeld *et al.*, 2010; Merikangas *et al.*, 2010; Vicente *et al.*, 2012; Costello *et al.*, 2014; Johnson *et al.*, 2016).

Of those who do make contact with mental health services, one approach to determining service effectiveness is to assess outcomes using individual-level approaches. For example, in CYP IAPT, outcomes have been reported using the percentage of individuals who showed reliable improvement (Wolpert *et al.*, 2016; Edbrooke-Childs *et al.*, 2018), using Jacobson and Truax's reliable change index (RCI), which is purported to determine change above and beyond that due to measurement error (Jacobson and Truax, 1991). CYP IAPT outcomes have also been reported considering whether users who started above a clinical threshold moved below that threshold following treatment, and whether this clinical change was also reliable (Wolpert *et al.*, 2016). However, there are several limitations to RCI and other methods for calculating individual change. Firstly, these methods typically do not address the missing data which is typical in longitudinal data (Speer and Greenbaum, 1995). Analyses based on completer samples (eg. had at least three treatment sessions in addition to assessment (Wolpert *et al.*, 2016; Edbrooke-Childs *et al.*, 2018) inflate estimates of treatment response when compared against estimates of all who were assigned to the treatment (Bollini *et al.*, 1999). Taking the latter intent-to-treat approach is considered more relevant for ascertaining the public health benefit of a treatment since treatment non-adherence is included in the estimate (Ten Have *et al.*, 2008). Secondly, most of the methods to calculate individual change yield unacceptable levels of false positives unless the test-retest reliability coefficient of the outcome is used in calculating change (Ferrer and Pardo, 2014). For some outcomes, such as the general psychopathology factor developed in Chapter 3 and utilized as an outcome in Chapter 4, this information is unavailable, rendering these methods inappropriate. Thirdly, methods to calculate individual change have been shown to result in unacceptable levels of false negatives even at high effect sizes (eg. Cohen's d=1.8 yielded at false negative rate of 43% (Ferrer and Pardo, 2019)). This renders their estimates of treatment effectiveness questionable. Finally, often studies of individual change are comprised only of a treated group (eg. Wolpert *et al.*, 2017; Edbrooke-Childs *et al.*, 2018). Without a non-treated control group, any change following treatment cannot be unequivocally attributed to service utilization. Addressing whether standard care is better than no care is particularly important in mental illness, where regression to the mean and spontaneous remission of symptoms is common (Hengartner, 2019).

Randomized control trials (RCTs) can be employed to test the efficacy of mental health treatments. Such outcomes are typically estimated using an intent-to-treat approach (Ten Have *et al.*, 2008) and thus improve upon analyses of just completer samples. However, RCTs are highly controlled to demonstrate efficacy under ideal situations, and thus lack real-world generalizability (Burnam, 1996). Pragmatic or practical clinical trials (PCTs) are better equipped to inform decision making by enrolling a more diverse population, recruiting from a variety of settings, and measuring a

broad range of outcomes (Tunis, Stryer and Clancy, 2003). However, even PCTs still may be more restrictive than standard mental health care, by excluding based on comorbid mental health conditions, for example (Dowrick *et al.*, 2000). Further, the issue of a non-treated control group still bedevils PCTs. A non-treated control group is typically not allowed due to ethical considerations. Patients cannot be denied care and thus are still randomized into different treatments. Wait lists may be considered ethically appropriate to use as a no-treatment comparison group in clinical trials, although there is question as to how much a waitlist is equivalent to no treatment. A recent network meta-analyses performed on randomized controlled trials for depression treatments identified by NICE revealed that in 12-18-year olds, participants in the waitlist condition had no worse outcomes than those in many psychological interventions (NICE, 2019b). This finding either indicates that the treatments compared were ineffective or that a waitlist condition may not be synonymous with no intervention. Based on data from RCTs of psychotherapy treatments for depression in adults, improvements during the waitlist condition have been estimated at a pre- post-treatment effect size of 0.4 (Minami *et al.*, 2007). While waitlist patients may improve due to being part of a clinical trial (Munder *et al.*, 2019), this finding also does not rule out that the waitlist change is reflective of regression to the mean or spontaneous remission of symptoms (Hengartner, 2019). However, when compared against a no treatment condition in a network meta-analysis, waitlist participants performed significantly worse than no treatment (Furukawa *et al.*, 2014). The authors conjectured that perhaps the waitlist group was motivated to remain depressed to be able to receive treatment at the end of the waitlist. Nonetheless, the above findings highlight the limitations of using RCTs and PCTs to determine effectiveness of mental health services.

In light of the above limitations, observational community studies, where young people with mental health needs may or may not make contact with mental health services, represent an excellent opportunity to provide good quality evidence of mental health treatment effectiveness. Unlike the approaches outlined above, such studies provide a naturalistic non-treated comparison group coupled with a real-world representation of service usage. The obvious limitation of selection bias in such a sample (Altman and Bland, 1999) can be addressed through adjusting by the propensity score (Rosenbaum and Rubin, 1983). This statistical technique balances the treated and untreated groups based on all observed covariates, similar to the random distribution of covariates between groups in an RCT. However, groups cannot be balanced by unmeasured covariates as RCTs are. Propensity adjustment generally cannot be implemented alongside calculations of individual change; comparing individual improvement rates of a treated versus an untreated group would therefore be biased. Thus, propensity weighting is typically limited to group-level comparisons. This trade-off is reasonable, given the limitations in individual change approaches outlined above. Adjustments based on the propensity score greatly increase confidence in any associations made between young people's contact with mental health services and subsequent gains in mental health.

Acknowledging that all approaches have their own unique set of advantages and limitations, it is important that policy makers consider propensity-weighted findings from observational community studies in conjunction with findings from RCTs, PCTs, and routine data collected from CYP IAPT and other mental health services. Such triangulation of evidence will allow decision makers to have a broad base from which to propose improvements in CYP MHS.

1.6 The present body of work

The present work utilizes two epidemiological studies (Goodyer *et al.*, 2010; Kiddle *et al.*, 2018) to assess the effectiveness of services used by young people with mental health needs. As these data are longitudinal, they allow for the modelling of trajectories based on service utilization. Trajectories can be estimated at the group-level (controlling for individual differences) or calculated individually. However, given the above barriers and limitations regarding the calculation of individual change, the focus of change in the present body of work is on that of group-level changes.

I tackle this issue of mental health treatment effectiveness in Chapter 2 (Neufeld *et al.*, 2017) using a British cohort of adolescents followed from ages 14 to 17 (Goodyer *et al.*, 2010). Trajectories of self-reported depressive symptoms were modelled over three years, with rate of change compared by groups based on age 14 mental disorder and service contact. A wide range of sociodemographic, environmental, individual, mental health, and diagnostic variables were used to adjust the data for potential confounding. Further, in order to address lack of randomization of this community study, propensity weighting (Rosenbaum and Rubin, 1983) was performed on those with a mental disorder, adjusting for the initial likelihood to access services or not, based on the variables which had previously been used to adjust for confounding. Clinical relevance of findings was bolstered by implementing a clinical cut-off. This study is the first of its kind to relate community mental health service contact with subsequent improvement in adolescent mental health, while addressing non-randomisation of service contact and attrition. Findings are discussed in light of policy implications (Neufeld, Jones and Goodyer, 2017).

Given the limitations of existing transdiagnostic models of mental health, in Chapter 3 I comprehensively validated a bifactor model of mental illness and wellness using longitudinal data from a sample of adolescents and emerging adults (Kiddle *et al.*, 2018). Findings in this chapter are developmental and therefore independent of mental health treatment, but lay the groundwork for assessing the relationship of mental health treatment with a transdiagnostic factor of mental health in the next chapter. The bifactor model developed in this chapter has been previously reported using cross-sectional data only (St Clair *et al.*, 2017), and redresses a gap in the literature by focusing more extensively on the key developmental period when the majority of mental illnesses emerge (Kessler *et al.*, 2007). As scant few bifactor studies have done (only (Castellanos-Ryan *et al.*, 2016; Greene and Eaton, 2017)), I tested longitudinal invariance, enabling more valid interpretation of homotypic and heterotypic continuity coefficients than in studies where this

invariance was not tested (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018). I determined construct replicability (Hancock, 2001) of the latent factors over time, which the majority of bifactor models consistently omit (Rodriguez, Reise and Haviland, 2016a). Expanding on prior work, I used a suite of external variables to determine longitudinal construct validity and predictive validity of the factors, based on background, environmental, and personality variables, harmful behaviours, IQ, and BMI. A multivariate model including the first three domains of variables determined the relative importance of these variables in predicting the factors from the bifactor model. Finally, the accelerated longitudinal design of the present study (Kiddle *et al.*, 2018) allows accelerated growth modelling (Duncan and Duncan, 2004) to be performed to obtain growth curves over the full age range of 14-27, whilst taking cohort effects into account. This was performed on each of the factors from the bifactor model, to determine rate and shape of change during this sensitive period, as has not previously been done. Together these findings help increase understanding of the structure of mental health during adolescence and emerging adulthood.

I next used this validated distress factor (*p*) in Chapter 4 to ascertain whether community mental health treatment is related to reduced levels of distress over two years in adolescents and emerging adults. This chapter extends Chapter 2 by determining whether treatment effects are observed in an additional sample with a broader age range (Kiddle *et al.*, 2018) during the full key period when mental disorders emerge (Kessler *et al.*, 2007) and when transitions occur between mental health services (Singh *et al.*, 2010). A transdiagnostic outcome was used to ascertain whether general treatment-related improvements are seen across all disorders. Further, mechanisms of treatment action were explored by performing mediation analysis with measures of social functioning. As presence of mental disorder was not determined in this study, analyses comprised those above the population mean on distress. This acts as a threshold for treatment need, which in Chapter 2 I show is important when estimating treatment effects (Neufeld *et al.*, 2017). This also allowed the groups to be more comparable in order to propensity weight based on background, social environment, and harmful behaviour variables. Findings are discussed with respect to access to care in adolescents and emerging adults.

I conclude by discussing the broader implications of the findings in Chapter 5. Analyses in two separate community samples of young people robustly support the notion that treatment-as-usual is related to improvements in specific (ie: depression) and transdiagnostic (ie: *p)* measures of mental health. Such treatment-related improvements in mental health appear to also yield benefits in family functioning in adolescents. However, over the past decade in the UK there has been a striking increase in reported prevalence of young people with a long-standing mental health condition (Pitchforth *et al.*, 2019). While some improvements in services have been made (for CAMHS: (Children's Commissioner, 2018); for adults, IAPT: (Clark, 2018)), society needs to prioritize young people's mental health in order to keep pace with these needs. Not only is the moral

imperative to do so strong, but the long-term societal benefits of such actions are compelling (Goodman, Joyce and Smith, 2011; Knapp, 2014; Clark, 2018).

CHAPTER TWO

Reduction in adolescent depression following contact with mental health services

2.1 Introduction

This chapter is based on publications from a study assessing the effectiveness of service contact due to mental health needs. As first author, I performed all analyses and drafted the publications. Using a community sample, lack of randomization was addressed using propensity weighting to balance baseline covariates between those who were and were not referred. This statistical approach equalizes the groups in a manner similar to an RCT (Rosenbaum and Rubin, 1983). Propensity weighting has not previously been undertaken to assess the effectiveness of mental service contact in adolescents with a demonstrated need for services. Service need was defined in the present study by those who met criteria for a mental disorder. Attrition in this 3-year longitudinal study was addressed by multiple imputation, and a clinical cut-off was implemented in order to render findings more clinically relevant. Findings indicated that mental health service contact was related to a greater reduction in subsequent depression in adolescents compared with those who had not made service contact (Neufeld *et al.*, 2017). While regression to the mean in the untreated sample indicated that those without service contact were as a whole below MFQ clinical cut-off three years later, this could be due to the fact that not all had a depressive diagnoses at baseline – due to sample size restraints, all with any mental disorder were included in the analyses. The study acknowledges this as a limitation and that this inclusivity would, if anything, bias findings to the null.

The two related publications which followed the initial article in Lancet Psychiatry (Neufeld *et al.*, 2017) are also included in this chapter. Firstly, the editor invited us to submit a response based on a PubPeer review of the initial article. This correspondence allowed further detailing of the paper's statistical methods. In particular, I expounded on the appropriateness of the utilized propensity score method for small samples, and justified the approach taken for specification of the propensity score model. A minor correction to table 1 was also included; the version of the Lancet Psychiatry paper below is the corrected version. Secondly, I was invited to write a paper expanding on the policy implications of the Lancet Psychiatry paper, which is the third publication in this chapter (Neufeld, Jones and Goodyer, 2017). This article reviewed relevant academic and grey literature, which underscored the need for more studies to longitudinally assess adolescent mental health and service usage. Comparing the Neufeld et al, 2017 findings to grey literature indicated the deleterious effects of austerity on young people's access to mental health services, as well as the importance of school-based counselling and better training of primary providers who make mental health service referrals.

In sum, findings from this chapter support the notion that community-based service contacts for mental health problems is related to reduced adolescent depression. Evidence provided indicates that access to services appears to have been put at risk based on government funding cuts. Suggestions are given for ways to improve identification of mental health problems and how to ensure continued access to important services. Finally, the findings from this chapter provide impetus for testing the effectiveness of mental health services in another sample with altered parameters such as a different age range or outcome. A transdiagnostic outcome is developed in Chapter 3 which is then used in Chapter 4 as the outcome for assessing mental health treatment effectiveness in adolescents and emerging adults. Chapter 2's findings are further extended in Chapter 4 by testing social support as a potential mechanism of action of mental health treatment.

Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK

Sharon A S Neufeld, Valerie J Dunn, Peter B Jones, Tim J Croudace, Ian M Goodyer

Lancet Psychiatry 2017; 4: 120–27

Published Online January 10, 2017 http://dx.doi.org/10.1016/S2215-0366(17)30002-0

This online publication has been corrected. The corrected version first appeared at thelancet.com/psychiatry on March 27, 2017

See Comment page 86

Department of Psychiatry, University of Cambridge, Cambridge, UK (S A S Neufeld BSc, V J Dunn BEd, Prof P B Jones PhD, Prof I M Goodyer MD); **and School of Nursing and Midwifery and Social Dimensions of Health Institute, University of Dundee, Dundee, UK** (Prof T J Croudace PhD)

Correspondence to: Mrs Sharon AS Neufeld, Developmental Psychiatry, Department of Psychiatry, University of Cambridge, Cambridge CB2 8AH, UK **sasn2@medschl.cam.ac.uk**

Summary

Background Evidence regarding the association between service contact and subsequent mental health in adolescents is scarce, and previous findings are mixed. We aimed to longitudinally assess the extent to which depressive symptoms in adolescents change after contact with mental health services.

Methods As part of a longitudinal cohort study, between April 28, 2005, and March 17, 2010, we recruited 1238 14-year-old adolescents and their primary caregivers from 18 secondary schools in Cambridgeshire, UK. Participants underwent follow-up assessment at months 18 and 36. Trained researchers assessed the adolescents for current mental disorder using the Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime version (K-SADS-PL). Caregivers and adolescents reported contact with mental health services in the year before baseline. Adolescents self-reported depressive symptoms (Mood and Feelings Questionnaire [MFQ]) at each timepoint. We assessed change in MFQ sum scores from baseline contact with mental health services using multilevel mixed-effects regression adjusted for sociodemographic, environmental, individual, and mental health confounders, with multiple imputation of missing data. We used propensity score weighting to balance confounders between treatment (users of mental health services) and control (non-users of mental health services) groups. We implemented an MFQ clinical cutoff following the results of receiver operating characteristic analysis.

Findings 14-year-old adolescents who had contact with mental health services in the past year had a greater decrease in depressive symptoms than those without contact (adjusted coefficient –1·68, 95% CI –3·22 to –0·14; p=0·033). By age 17 years, the odds of reporting clinical depression were higher in individuals without contact than in service users who had been similarly depressed at baseline (adjusted odds ratio 7·38, 1·73–31·50; p=0·0069).

Interpretation Our findings show that contact with mental health services at age 14 years by adolescents with a mental disorder reduced the likelihood of depression by age 17 years. This finding supports the improvement of access to adolescent mental health services.

Funding Wellcome Trust, National Institute for Health Research.

Introduction

See Online for appendix

Many mental disorders emerge during adolescence and continue into adulthood.[1] In depressive disorders, younger onset is associated with more depressive episodes, longer episode duration, increased comorbidity, suicidality, and admission to hospital.[2] Among individuals with a diagnosed depressive disorder, adolescents are more likely than adults to delay contact with mental health services, thereby increasing episode duration and risk of recurrence. Clearly, early identification and treatment of mental disorders during adolescence would contribute to reduction and perhaps prevention of adverse sequelae.

Measurement of the treatment gap—the discrepancy between disorder prevalence and proportion treated—is a prerequisite to enable policy makers to prevent such adverse sequelae from arising. To predict service need, a clearly recognised cutoff for mental disorder, such as meeting DSM diagnostic criteria, is desirable. Our review of international studies that report DSM-IV disorder and past-year contact with mental health services for those with a disorder (appendix pp 1, 2), found that 12–25% of adolescents have a mental disorder, of whom only 34–56% access mental health services. Previous surveys[3,4] in the UK report much higher proportions of contact with mental health services (71% of children or adolescents with a mental disorder); however, unlike most studies, these estimates classify seeking help from a teacher as a mental health service contact. Other studies[5,6] report 12–19% lower service use rates for anxiety than for depression.

The association between adolescents' contact with mental health services and subsequent mental health remains unclear in community samples, but is vital to clarify if adolescent mental health services are to compete for health-care funding. Findings from studies[7,8] using broad definitions of mental health problems without a cutoff for service need have shown that use of mental health services had little effect on subsequent mental health problems. However, results are more promising if

Research in context

Evidence before this study

In 2015, a task force in the UK noted the paucity of good quality national information regarding Child and Adolescent Mental Health Services (CAMHS) outcomes. To identify previous published work, with no language restrictions, that assessed the association between CAMHS use and subsequent mental health, we searched PubMed (* denotes wildcard) for articles published in the past 16 years (from Jan 1, 2000, to July 5, 2016) for the terms (service* OR help-seek*) AND (psychopatholog* OR mental* OR psychiatric*) AND (observation* OR community OR survey OR cohort OR epidemiolog*) AND (longitudinal[Title] OR prospective[Title] OR change[Title] OR reduc*[Title] OR improve*[Title] OR effectiveness[Title] OR outcome[Title]) AND (adolescen*[Title] OR youth*[Title] OR young*[Title]). We required studies to reflect treatment-as-usual mental health service use, and have a non-service using comparison group. We identified additional papers by checking citations.

We identified six studies that yielded mixed findings regarding the association of service contact with subsequent mental health. Two studies that assessed change in all service users without a clearly recognised cutoff for service need, such as DSM, showed that mental health service use had little effect on subsequent total mental health problems over and above that to be expected from natural remission. The four remaining studies assessed adolescents at greater risk of a mental disorder or those with a DSM diagnosis. These studies showed an improvement in mental health following service contact, but none addressed non-randomisation of service contact or attrition, only one adequately addressed confounding variables, and only one showed significant effects that were clinically relevant. None of these studies were from the UK (three were from the USA and three were from Europe).

Added value of this study

To our knowledge, this study is the first of its kind in the UK, and the first to support the association of mental health service contact and the improvement of mental health by late adolescence, while addressing non-randomisation of service contact and attrition. In addition to propensity score weighting (which balances treatment and control groups on confounders, similar to a randomised control trial) to adjust for participants' initial likelihood to access services, and multiple imputation to deal with missing data, we used a clinically relevant cutoff and adjusted for a wide range of time-varying confounding variables. These adjustments give greater confidence than previous studies to the notion that mental health service contact is related to meaningful improvements in subsequent mental health. This study is also the first we are aware of that shows that the association of mental health with previous treatment is attenuated if that treatment was irrespective of service need.

Implications of all the available evidence

The spending of the UK National Health Service (NHS) on children's mental health services has fallen by 5·4% in real terms since 2010 (£41 million), despite an increase in demand. The present findings support the positive role played by mental health services in a cohort before these NHS cuts, illustrating to policy makers the validity of increasing the availability of child mental health services to at least 2010 levels. That positive findings became non-significant upon inclusion of all mental health service users irrespective of disorder underscores the importance of clinical assessment when making referral decisions. These findings support training of service referrers (eg, in primary care or schools) in detection of the presenting features of mental disorders, to increase the proportion of referrals of individuals with a clear need who could be more responsive to treatment.

adolescents are at greater risk of, or already have, a mental disorder. In adolescents who witnessed community violence, use of mental health services reduced depressive symptoms.[9] Adolescents with fearful spells or panic attacks were more likely to develop diagnosable panic disorder and depression if they had not used mental health services.[10] Patients treated for emotional disorders at Child and Adolescent Mental Health Services (CAMHS) showed significant improvement compared with controls,[11] yet this change was not clinically meaningful. Finally, among DSM-diagnosed adolescents, users of specialist mental health services had reduced symptoms compared with those who were untreated, but only if eight or more sessions were attended.[12] However, none of the studies that showed a positive association between service contact and mental health addressed non-randomisation or attrition. Only one study[9] adequately addressed confounding variables (ie, those associated with both predictor and outcome, which could bias the association between service use and subsequent mental health), and only one study[10] showed significant effects that were clinically relevant.

In the present study, we used a longitudinal repeated-measures design on a community ascertained cohort to assess change in adolescent depressive symptoms from ages 14 years to 17 years after contact with mental health services. For the outcome, we used depressive symptoms as a valid identifier of major depressive disorders,[13,14] which are highly prevalent[5] and predictive of future morbidity.[2] To extend this previous work, the design controls for differences in symptoms and background factors among service users and non-users at baseline and over time, in individuals with and without a DSM-defined mental disorder. We hypothesised that self-reported depression scores would be reduced to a greater extent in adolescents who contacted mental health services than in those with no contact, but that these effects would be stronger in the subsample with a clearly defined need for mental health services, based on the presence of a diagnosable mental disorder. We

hypothesised that these findings would remain when we addressed attrition, confounding variables, non-randomisation of mental health service contact, and clinical relevance.

Methods

Study design and participants

As part of the ROOTS longitudinal cohort study[15] of mental health, between April 28, 2005, and March 17, 2010, we recruited 1238 adolescents and primary caregivers (1134 [92%] were the biological mother of the adolescent) from 27 secondary schools in Cambridgeshire, UK. 18 secondary schools approached agreed to participate, with letters of invitation sent until the sample size reached a prespecified cutoff of 1000 participants. Of a possible 3762 students, 1238 agreed to participate. Participants were interviewed separately and completed questionnaires at mean ages 14·5 years (timepoint 1 [T1]), 16 years (timepoint 2 [T2]), and 17·5 (timepoint 3 [T3]) years (T1–3 means, SDs 0·3). Written informed consent was obtained from adolescents and caregivers before participation. Cambridgeshire 2 Research Ethics Committee local ethics committee provided ethics approval.

Procedures

At T1, trained researchers assessed adolescents' mental health status using the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime version (K-SADS-PL)[16] to establish DSM-IV[17] axis 1 diagnoses. Adolescents with a high clinical index (subthreshold for diagnosis, but exhibiting four symptoms and clinical impairment) were classified as diagnosed. Experienced psychiatrists (IMG, PBJ) trained interviewers and conducted consensus meetings regarding all K-SADS assessments. Inter-rater agreement for diagnosis was high (95%). Disagreements were settled by clinical consensus meetings between clinical psychiatry experts.

Mental health service contact was defined as an adolescent's assessment or treatment of a mental health problem by a primary care provider (ie, general practitioner) or a mental health specialist from any sector. Caregiver and adolescent responses were obtained by use of various measures (items in appendix pp 3–5). T1 past-year mental health service contact (no or yes) was generated as an exposure variable, and any mental health services after T1 (no or yes, post-T1–3) as a confounder. Caregivers reported contacts with adolescent mental health services at T1 from a semi-structured interview, with high inter-rater agreement on core indicators (κ=0·7–0·9; Cambridge Early Experiences Interview[18]) and from a self-reported questionnaire at T1 and T3. Adolescents were interviewed at T3 about mental health service contact before K-SADS-PL assessment. Adolescents also reported how often they had seen a doctor or other health professional regarding depressive symptoms in the past month (Kessler's Psychological Distress Scale[19]). We combined adolescent and caregiver responses with either response if one was missing, or with the positive response if sources disagreed (considering differential recall and caregivers potentially unaware of adolescent service use).

A combined variable was derived at T1 that defined participants with current mental disorder (yes or no) and past-year mental health service contact. This variable resulted in four levels: unaffected (no current disorder or past-year service contact), service contact only, disorder only, and disorder and service contact.

We assessed the Mood and Feelings Questionnaire (MFQ)[20] at T1–3. This 33-item adolescent self-report of current or past 2 weeks' depressive symptoms covers DSM criteria for major depressive disorders. The MFQ has shown prognostic validity in clinic and non-clinic samples,[13,14] yielding high internal consistency (α=0·92–0·94) in the present sample. Higher sum scores indicate more symptoms.

We chose 18 putative confounders that covered sociodemographic, environmental, individual, mental health, and diagnostic domains (appendix pp 7, 8) based on a previous association with mental health service contact, or depression. For example, family structure, functioning and mental health problems, peer support, maltreatment, stressful events, socioeconomic status, gender, past referrals for mental health problems, current diagnosis type, severity, and comorbidity have all been related to current mental health service contact.[21] We assessed seven confounders at multiple timepoints (appendix p 9).

Statistical analysis

We did primary analyses on an imputed dataset (appendix pp 7, 8) of individuals with complete data for T1 past-year mental health service contact and current mental disorder.

Imputed longitudinal MFQ scores were the outcome in multilevel mixed-effects linear regression models with maximum likelihood estimation, implemented in STATA 13.0. This analysis nests correlated data, thereby accounting for violations in the assumption of independence. For the present data, repeated assessments over time were nested within individuals (the random effect). Fixed effects (ie, predictors in the regression) included linear, quadratic, and categorical effects of age, and confounders (appendix p 7). We assessed categorical effects of T1 disorder and services (unaffected or disorder only or disorder and services) and this variable's interaction with age. We did not include the services-only group of individuals in the primary analysis because without a mental disorder their need for services was less clear. We explored the effects of nesting by school by adding school as a further random-effect.

We did receiver operating characteristic (ROC) analysis to determine the ability of MFQ to classify affective disorder. In ROC analysis the true positive rate (sensitivity) is plotted against the false positive rate (1–specificity). We estimated the area under the curve (AUC) and used it as

	Imputed sample			Complete case sample		
	n	Coefficient (95% CI)	p value	n	Coefficient (95% CI)	p value
MFQ all timepoints						
Main effects						
Disorder and services variable	3360	1·06 (0·42 to 1·69)	0·0014	2469	1·56 (0·95 to 2·17)	<0·0001
Age (linear)	3360	−0·11 (−0·34 to 0·12)	0·36	2469	−0·24 (−0·45 to −0·02)	0·032
Age^2 (quadratic)	3360	0·04 (−0·25 to 0·33)	0·77	2469	−0·36 (−0·62 to −0·10)	0·0075
Disorder and services variable × age						
Unaffected *vs* disorder only	3360	−1·01 (−1·91 to −0·10)	0·029	2469	−0·34 (−1·25 to 0·57)	0·46
Unaffected *vs* disorder and services	3360	−2·69 (−3·94 to −1·44)	0·0004	2469	−2·89 (−4·12 to −1·66)	<0·0001
Disorder only *vs* disorder and services	3360	−1·68 (−3·18 to −0·19)	0·028	2469	−2·54 (−4·04 to −1·04)	<0·0001
Disorder and services variable × age^2						
Unaffected *vs* disorder only	3360	−0·28 (−0·58 to 0·02)	0·070	2469	−0·08 (−0·38 to 0·22)	0·60
Unaffected *vs* disorder and services	3360	−0·83 (−1·24 to −0·42)	0·0004	2469	−0·99 (−1·39 to −0·58)	<0·0001
Disorder only *vs* disorder and services	3360	−0·56 (−1·06 to −0·05)	0·031	2469	−0·91 (−1·40 to −0·41)	<0·0001
Categorical analysis of age						
Unaffected						
T1–2	3015	0·05 (−0·70 to 0·79)	0·90	2257	0·59 (−0·07 to 1·24)	0·078
T2–3	3015	0·21 (−0·54 to 0·97)	0·58	2257	−0·95 (−1·63 to −0·27)	0·0063
T1–3	3015	0·26 (−0·47 to 0·99)	0·48	2257	−0·36 (−1·00 to 0·28)	0·27
Disorder only						
T1–2	213	−2·40 (−5·62 to 0·82)	0·14	140	−0·90 (−4·11 to 2·30)	0·58
T2–3	213	−0·83 (−4·33 to 2·67)	0·64	140	−1·42 (−4·75 to 1·89)	0·40
T1–3	213	−3·23 (−6·38 to −0·08)	0·044	140	−2·32 (−5·53 to 0·88)	0·15
Disorder and services						
T1–2	132	−4·08 (−9·07 to 0·91)	0·11	72	−1·29 (−6·76 to 4·21)	0·65
T2–3	132	−3·64 (−9·22 to 1·94)	0·20	72	−7·85 (−14·55 to −1·15)	0·022
T1–3	132	−7·72 (−12·98 to −2·46)	0·0044	72	−9·13 (−14·81 to −3·44)	0·0016
T1 MFQ						
Unaffected *vs* disorder only	1120	5·56 (3·58 to 7·53)	0·0004	983	5·03 (2·85 to 7·20)	<0·0001
Unaffected *vs* disorder and services	1120	5·56 (2·96 to 8·16)	0·0004	983	7·52 (4·63 to 10·42)	<0·0001
Disorder only *vs* disorder and services	1120	−0·01 (−3·10 to 3·09)	1·00	983	2·50 (−0·96 to 5·95)	0·16
T3 MFQ						
Unaffected *vs* disorder only	1120	2·73 (0·22 to 5·24)	0·033	769	4·20 (1·73 to 6·67)	<0·0001
Unaffected *vs* disorder and services	1120	−2·03 (−5·55 to 1·49)	0·26	769	−1·20 (−4·67 to 2·27)	0·50
Disorder only *vs* disorder and services	1120	−4·76 (−8·75 to −0·77)	0·020	769	−5·40 (−9·47 to −1·34)	0·0085

Data were adjusted as follows: gender, sociodemographics (ethnic origin, Index of Multiple Deprivation, adolescent living with biological parents), environmental factors (number of stressful life events in the past year, current family dysfunction and friendships, any family-focused adversities by T1), and mental health factors (any past Schedule for Affective Disorders and Schizophrenia for School-Age Children diagnosis, any mental health services after T1, any emotional problems in a family member [past 3 years or present], current antisocial traits). Variables not included were any mental health service referral age 0–13 years (p=0·19 in base model) and pubertal status (not a true confounder as p>0·10 and ρ<0·10 with predictor). MFQ=Mood and Feelings Questionnaire. T1=timepoint 1 (age 14·5 years). T2=timepoint 2 (age 16 years). T3=timepoint 3 (age 17·5 years).

Table 1: **Longitudinal change in MFQ by current mental disorder and past-year contact with mental health services at T1**

an index of diagnostic accuracy; a higher AUC reflects greater accuracy. The MFQ has previously been shown to have good-to-high diagnostic accuracy with this method.[13,14] Additionally, MFQ scores above the 75th percentile are an established behavioural marker for clinical diagnosis of major depression.[22] The Youdin Index was calculated to determine the clinical cutoff point, because it maximises sensitivity and specificity,[23] thereby increasing correct classification of individuals with and without depression.

To address the absence of randomisation of mental health service use, a propensity score was generated to weigh the outcome model. A propensity score is the individual probability of attending or receiving a service or treatment conditional on observed baseline covariates. The score is designed to balance confounders between a treatment and control group, as is done in a randomised control trial.[24] The primary propensity-adjusted analyses comprised data from adolescents with a mental disorder, because those in the disorder-only group were the most appropriate for comparison with the disorder-and-services group (appendix pp 7, 8 provide further details of propensity score). To reduce estimate bias, we first did analyses of the full sample with a disorder, then we restricted the sample to the region of common support—

the range of propensity scores which were observed in both treated and untreated individuals.[25] We estimated the propensity score weighted outcome models with generalised linear modelling (GLM) with a logit link, with adjustment for post-baseline confounding variables. A robust estimator accounted for the sample weighting.

To address the importance of use of a clearly defined need for mental health services based on the presence of a mental disorder, we reanalysed data including all service users, irrespective of disorder.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Of the 1238 participants recruited, 1190 adolescents had data for T1 current mental disorder and past-year mental health service contact (appendix p 6). The number of respondents with complete data for all outcomes and covariates at all timepoints was 983 (83%) for T1, 717 (60%) for T2, and 769 (65%) for T3. 64 (5%) adolescents made past-year contact with mental health services; 126 (11%) had a current mental disorder. Among individuals with a disorder, 48 (38%) reported past-year service contact and 46 (96%) of these contacts were based on T1 past-year recall; 36 (84%) of 43 of these adolescents attended five or more sessions (n=5 had missing data for treatment length). In the disorder-and-services group (n=48), disorders were affective (n=16 [33%]), anxiety (n=10 [21%]), behavioural (n=25 [52%]), and other (n=5 [10%]); 14 (29%) of these participants had a comorbid K-SADS diagnosis (appendix p 9).

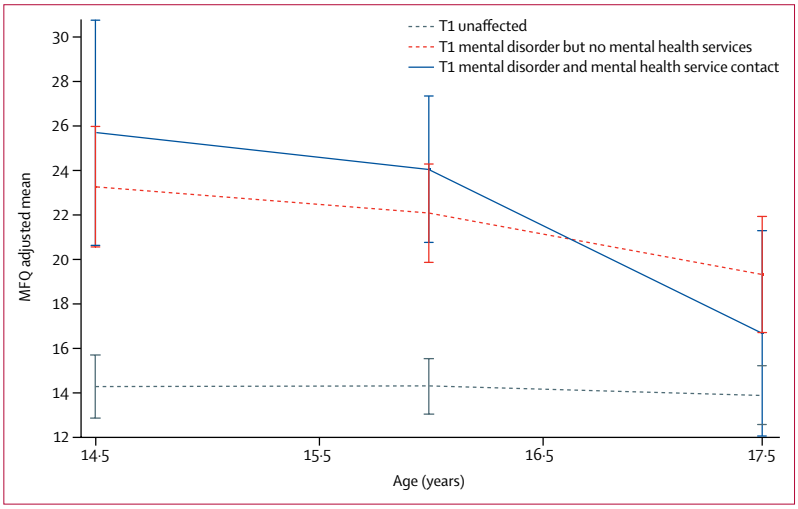


Figure: **Adolescent MFQ scores by T1 current mental disorder and past-year contact with mental health services**
Disorder and services variable; imputed and adjusted results. Error bars represent SDs. Adjustments made as for table 1. MFQ=Mood and Feelings Questionnaire. T1=timepoint 1 (age 14·5 years).

Overall, 16 (25%) of 64 service users had no disorder, and differed from the disorder-and-services group: baseline MFQ scores were lower in the no-disorder group, although with no significant difference between groups (coefficient –7·64, 95% CI –15·30 to 0·02; p=0·051), and MFQ scores did not change over time (coefficient 1·22, –1·01 to 3·44; p=0·28). Adolescents with a disorder predominantly accessed CAMHS, whereas unaffected adolescents mostly accessed a school counsellor (appendix p 11). Unaffected service users were less antisocial than service users with a disorder (coefficient –3·20, 95% CI 1·10 to 5·29; p=0·0034); remaining covariates p>0·062 (means in appendix p 9).

Adolescents with a disorder were substantially more impaired than unaffected adolescents across all domains of confounders (appendix p 9). When we compared adolescents with a disorder by mental health service contact, individuals varied mainly in diagnostic factors (appendix p 9).

1002 (84%) of 1190 service contacts were reported by both adolescents and caregivers, showing 98% agreement and high chance-corrected agreement (κ=0·78, 95% CI 0·71–0·84). The remaining service contacts were based on either adolescent or caregiver report.

Findings from adjusted multilevel mixed-effects regression analysis revealed that at T1, individuals in both the disorder only and disorder-and-services groups had significantly higher MFQ scores than did those in the unaffected group, but scores between the disorder only and disorder-and-services groups did not differ significantly (table 1, figure). MFQ scores in both these groups improved over time compared with the unaffected group, in which scores remained stable; however, scores improved more quickly among the disorder-and-services group than the disorder-only group (table 1, figure). By T3, scores in the disorder-and-services group had improved (reported reduced symptoms) to the levels of those in the unaffected group (table 1, figure). By contrast, at T3, patients in the disorder-only group reported significantly more symptoms than did those in both the disorder-and-services group and the unaffected group (table 1). Analyses repeated on complete case data yielded similar results (table 1; appendix p 12 shows imputed and complete-case analysis results from unadjusted models). Nesting by school did not affect complete-case results; thus, we did not do clustering during imputation. All data we present for comparability are non-nested results.

ROC analysis revealed MFQ as an excellent discriminator of affective disorder (AUC=0·93, 95% CI 0·90–0·96). The Youden Index indicated an MFQ clinical cutoff point of 22, with 94% sensitivity and 79% specificity, greater than previously obtained in a similar sample measured with differing cutoff point methodology.[14]

We included nine baseline covariates in the propensity score weighting (table 2). Propensity score weighted

	Propensity score weighted only*		Propensity score weighted and adjusted for post-baseline confounds		Post-baseline confounds
	OR (95% CI)	p value	OR (95% CI)	p value	
Adolescents with a T1 mental disorder: service contact *vs* none					
Full propensity score sample (n=119)	3·70 (1·40–9·82)	0·0086	5·23 (1·47–18·63)	0·011	T2 MFQ; T3 family dysfunction, stressful life events
Common support sample (n=98)	4·36 (1·41–13·47)	0·011	7·38 (1·73–31·50)	0·0069	T2 MFQ; T3 stressful life events , family dysfunction, living with biological parents
All with T1 mental health service contact *vs* T1 mental disorder but no services					
Full propensity score sample (n=134)	1·78 (0·81–3·92)	0·15	2·41 (0·92–6·32)	0·073	T1 MFQ;† mental health service contact after T1; T2 friendships; T3 stressful life events, living with biological parents
Common support sample (n=94)	2·36 (0·93–6·02)	0·072	2·65 (0·88–7·97)	0·085	T1 MFQ; mental health service contact after T1; T3 stressful life events, living with biological parents, family dysfunction

OR=odds ratio. T1=timepoint 1 (age 14·5 years). T2=timepoint 2 (age 16 years). MFQ=Mood and Feelings Questionnaire. T3=timepoint 3 (age 17·5 years). *Variables used in the propensity score model are ethnic origin, gender, pubertal status, mental health referrals aged 0–13 years, past Schedule for Affective Disorders and Schizophrenia for School-Age Children diagnosis, current behavioural diagnosis, and environmental factors (current friendships and family dysfunction, past-year stressful life events). †T1 MFQ was used if more strongly related to predictor and outcome than T2 MFQ.

Table 2: **MFQ clinical cutoff point at T3 predicted by propensity score weighted mental health service contact at T1**

GLM revealed that among adolescents with a mental disorder, those without contact with mental health services at T1 had nearly four times the odds of being depressed by T3 compared with those in the disorder-and-services group (table 2). Inclusion of post-baseline confounding variables increased odds by more than five times, and in the common support sample, to more than seven times (table 2). Data for propensity score covariates were missing for five (4%) of 124 adolescents with a disorder. To assess the effect of MFQ imputation and missing covariate data on findings, we did unweighted GLM with mental health service contact at T1 predicting T3 MFQ clinical cutoff (adjusted by T1 MFQ only) in three separate models: model A (raw MFQ [n=95]), model B (imputed MFQ [n=124]), and model C (imputed MFQ with missing data from propensity score weighted covariates [n=119]). Effect sizes (calculated from odds ratios[26]) for mental health service contact in these models were similar (0·44 for model A, 0·46 for model B, and 0·45 for model C), indicating no effect of imputation or missing data.

We repeated analyses by expanding the treatment group to include all adolescents who had made past-year contact with mental health services at T1, including 16 individuals with no T1 mental disorder. Comparison groups remained the same as before. The multilevel mixed-effects regression required the same confounding variables as the primary analyses, yielding equivalent results for the unaffected group compared with the other groups. Although this treatment group had the equivalent T1 MFQ to the disorder-only group (coefficient –0·94, 95% CI –3·81 to 1·93; p=0·52) as in the primary analyses, the two groups did not differ in their rate of change over time (linear coefficient –0·68, –2·07 to 0·70; p=0·33; quadratic coefficient –0·27, –0·74 to 0·20; p=0·26). Results did not differ significantly with propensity score weighted GLMs (table 2, appendix pp 7, 8).

Discussion

To our knowledge, this study is the first in adolescents to support the role of contact with mental health services in improving mental health by late adolescence, while addressing non-randomisation and attrition. Four similar studies[9–12] did not address these issues; only one study[9] adequately controlled for confounding variables, and one other study[10] showed significant effects that were clinically relevant. Two studies[11,12] only assessed specialist mental health services, and one study[12] reported effects of services only if eight or more sessions were attended. In the present study, we considered mental health services from all sectors irrespective of treatment length, we multiply imputed missing data, used propensity score weighting to adjust for participants' initial likelihood to access services, and data yielded clinically relevant results robust to a wide range of confounds. Contact with mental health services appeared to be of such value that after 3 years the levels of depressive symptoms of service users with a mental disorder were similar to those of unaffected individuals. Among adolescents with a mental disorder at age 14 years, the odds of those without past-year contact with mental health services having clinical depression by age 17 years were greater than for service users who had been similarly depressed at baseline. Recruitment of participants from the general population, who vary in diagnosis type, severity, and treatment type, and the absence of strict inclusion criteria as in randomised controlled trials also increases the external validity of our study, especially for public mental health and policy makers in the field of community and specialised youth services.

Our findings are in contrast with the null[8] or negative[7] association of mental health services reported with longitudinal total emotional and behavioural problems, with no diagnostic threshold. These studies defined mental health services in a similar manner to the present study; one study[7] implemented propensity matching to address the absence of randomisation. However, measurement of total problems irrespective of clinical typology might mask potential influences of mental health services on emotional or internalising symptoms. Previous null findings can also be explained by a disregard to service need. The present study's findings became non-significant when all users of mental health services were included in the treatment group irrespective of disorder. This outcome underscores the importance of assessment, and supports training of service referrers (eg, in primary care or schools) in the presenting features of mental disorders, to increase the proportion of referrals of adolescents with a clear need who could be more responsive to treatment. Our findings suggest that adolescents accessing mental health services without a mental disorder might be less antisocial, but with fewer symptoms they could be less likely to improve from treatment. Future work should further elucidate this group.

Our study has some limitations. First, verification of the self-report of mental health service use against medical records would have been beneficial; however, findings are supported by high caregiver–adolescent agreement and similar proportions reported in comparable studies in other countries—eg, in adolescents with a DSM-IV diagnosis, 34–56% had past-year contact with any mental health services and 19–25% had contact with specialist mental health services (appendix p 1); the proportions in our study were 38% and 22%, respectively (appendix p 11). Second, heterogeneous treatment makes speculation about a mechanism for improvement difficult. However, common features across treatments could have a role; for example, listening, advice giving, problem solving, being non-judgmental, and being supportive. Larger studies assessing service use separately by treatment type might reveal relative effectiveness, to aid policy makers in determining which services to support. Third, sample size prohibited a focus on participants with a depressive diagnosis; thus, we included adolescents with any DSM diagnosis. However, because adolescents without depression are less likely to show change in depression related to service contact, inclusion of all diagnoses biases the findings to the null. Furthermore, because of numbers of participants, we could not do analyses by varying treatment lengths. However, the intention-to-treat assumption also biases findings to the null; therefore, it is noteworthy that an effect of service use was found. Finally, although addition of covariates and propensity score weighting helped us to address confounding variables, our study had no pretreatment baseline. A larger study with more longitudinal assessments could allow analysis of adolescents initiating service use in a naturalistic setting.[12]

Although our findings are an encouragement to policy makers and commissioners that CAMHS helps to improve mental health, such findings cannot be cause for complacency. Figures published in 2015 show that National Health Service (NHS) spending on children's mental health services in the UK has fallen by 5·4% in real terms since 2010 (£41 million), despite an increase in demand.[27] Audits have shown a resultant increase in referrals and waiting times; providers report increasingly complex and severe presenting problems, associated with longer stays in inpatient facilities.[28] The present study occurred in a cohort before these NHS cuts, illustrating to UK policy makers the importance of increasing availability of CAMHS to at least the 2010 levels. Globally, in high-income countries, total mental health spending represents no more than 6% of governmental health expenditures; in many other countries, this figure is less than 1%,[29] despite mental disorders being one of the leading causes of non-communicable disease burden worldwide.[30] More studies assessing the effectiveness of CAMHS are needed[28] for children's mental health to compete for government funds.

When mental health services are ramped-up, care needs to be taken to reach individuals with mental health needs who would typically not access services, comprising more than 60% of those with a mental disorder in our sample. This approach could include increasing community-based services, and ensuring a clear access point to mental health services, such as a designated individual in every school and primary care practice.[28] Focused training of such individuals in identification of mental disorders could help to prioritise access to mental health services for young people with a clearly defined need, to the betterment of their mental health and wellbeing.

Contributors

IMG, PBJ, and TJC conceived and designed the ROOTS study. VJD, PBJ, TJC, and IMG organised the conduct of, and carried out the ROOTS study (including acquiring study data). SASN coded health services data, and SASN and VJD resolved any queries. SASN analysed the data, and SASN, IMG, PBJ, and VJD interpreted the data. SASN drafted the manuscript, and IMG, PBJ, and VJD criticised the manuscript for intellectual content. All authors have read and approved the final version of the manuscript. SASN and IMG are the guarantors for the study.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was funded by The Wellcome Trust (grant number 074296), and the National Institute for Health Research Collaboration for Leadership in Applied Health Research & Care for Cambridgeshire and Peterborough. We thank the participants, and the schools and research assistants who helped with recruitment and data collection.

References

1 Jones PB. Adult mental health disorders and their age at onset. *Br J Psychiatry* 2013; **54:** s5–10.

2 Korczak D, Goldstein B. Childhood onset major depressive disorder: course of illness and psychiatric comorbidity in a community sample. *J Pediatr* 2009; **155:** 118–23.

3 Green H, McGinnity A, Meltzer H, Ford T, Goodman R. Mental health of children and young people in Great Britain, 2004. Basingstoke: Office for National Statistics, 2005.

4 Meltzer H, Gatward R, Goodman R, Ford T. Mental health of children and adolescents in Great Britain. London: Office for National Statistics, 2000.

5 Merikangas KR, He J-P, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics* 2010; **125:** 75–81.

6 Costello EJ, He J-P, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with psychiatric disorders: 12-month data from the national comorbidity survey-adolescent. *Psychiatr Serv* 2014; **65:** 359–66.

7 Jörg F, Ormel J, Reijneveld S, Jansen DE, Verhulst FC, Oldehinkel AJ. Puzzling findings in studying the outcome of "real world" adolescent mental health services: the TRAILS study. *PLoS One* 2012; **7:** e44704.

8 Zwaanswijk M, Verhaak PF, van der Ende J, Bensing JM, Verhulst FC. Change in children's emotional and behavioural problems over a one-year period. *Eur Child Adolesc Psychiatry* 2006; **15:** 127–31.

9 Chen W-Y, Corvo K, Lee Y, Hahm HC. Longitudinal trajectory of adolescent exposure to community violence and depressive symptoms among adolescents and young adults: understanding the effect of mental health service usage. *Community Ment Health J* 2016; published online June 10. DOI:10.1007/s10597-016-0031-5.

10 Asselmann E, Wittchen H-U, Lieb R, Höfler M, Beesdo-Baum K. Does help-seeking alter the risk for incident psychopathology in adolescents and young adults with and without fearful spells or panic attacks? Findings from a 10-year prospective-longitudinal community study. *J Affect Disord* 2014; **169:** 221–27.

11 Nilsen TS, Handegård B-H, Eisemann M, Kvernmo S. Evaluating change in symptomatic and functional level of children and youth with emotional disorders: a naturalistic observation study. *Eur Child Adolesc Psychiatry* 2015; **24:** 1219–31.

12 Angold A, Costello J, Burns BJ, Erkanli A, Farmer E. Effectiveness of nonresidential specialty mental health services for children and adolescents in the "real world". *J Am Acad Child Adolesc Psychiatry* 2000; **39:** 154–60.

13 Wood A, Kroll L, Moore A, Harrington R. Properties of the mood and feelings questionnaire in adolescent psychiatric outpatients: a research note. *J Child Psychol Psychiatry* 1995; **36:** 327–34.

14 Daviss WB, 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 Allied Discip* 2006; **47:** 927–34.

15 Goodyer IM, Croudace T, Dunn V, Herbert J, Jones PB. Cohort profile: risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project. *Int J Epidemiol* 2010; **39:** 361–69.

16 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:** 980–88.

17 American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th edn. Washington, DC: American Psychiatric Association, 1994.

18 Dunn VJ, Abbott RA, Croudace TJ, et al. Profiles of family-focused adverse experiences through childhood and early adolescence: the ROOTS project a community investigation of adolescent mental health. *BMC Psychiatry* 2011; **11:** 1–16.

19 Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med* 2002; **32:** 959–76.

20 Costello EJ, Angold A. Scales to assess child and adolescent depression—checklists, screens, and nets. *J Am Acad Child Psychiatry* 1988; **27:** 726–37.

21 Zwaanswijk M, Verhaak PFM, Bensing JM, van der Ende J, Verhulst FC. Help seeking for emotional and behavioural problems in children and adolescents. *Eur Child Adolesc Psychiatry* 2003; **12:** 153–61.

22 Owens M, Herbert J, Jones PB, et al. Elevated morning cortisol is a stratified population-level biomarker for major depression in boys only with high depressive symptoms. *Proc Natl Acad Sci USA* 2014; **111:** 3638–43.

23 Perkins NJ, Schisterman EF. The inconsistency of 'optimal' cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006; **163:** 670–75.

24 Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70:** 41–55.

25 Heckman JJ, Ichimura H, Smith J, Todd P. Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proc Natl Acad Sci USA* 1996; **93:** 13416–20.

26 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000; **19:** 3127–31.

27 O'Brien S, Meek L, Adam S, et al. The mentally healthy society: the report of the taskforce on mental health in society. London: Labour Party, 2015.

28 McShane M, Rouse J, Allington-Smith P, et al. Future in mind: promoting, protecting and improving our children and young people's mental health and wellbeing. London: Department of Health, 2015.

29 WHO. Mental Health Atlas 2014. Geneva: World Health Organization; 2015.

30 GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388:** 1603–58.

Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK

Following recent comments on PubPeer regarding our paper,[1] we want to further clarify our statistical methods. To reiterate, we examined mental health service contacts in an adolescent cohort (n=1238), in which 1190 (96%) participants had useable data, but a small percentage (11%, n=126) of these participants had a mental disorder. Of the participants with a mental disorder, 48 (38%) had a past-year mental health service contact at baseline. Our investigation tested whether self-reported depression scores were lower 3 years later among those with a mental illness at baseline who received a mental health service contact compared with those who did not. The statistical methods chosen aimed to account for the confounding effects of a set of fixed covariates that might have accounted for the association of service contact with depression and absence of randomisation to mental health service use at recruitment. Much of the justification for our data analytic strategy is published online in the appendix.

We note in the paper that findings from the multilevel models unadjusted by covariates are presented in the appendix (p 12). These findings show that unadjusted imputed depression scores and unadjusted raw depression scores improved more quickly among the disorder-and-services group than the disorder-only group, consistent with both the imputed and raw adjusted findings. Our use of propensity analysis following the multilevel modelling was a way of more robustly checking the findings from the multilevel modelling. In our appendix (pp 7, 8), we have given the rationale for the appropriateness of using propensity scoring given the sample size. Although the appendix is fully available online, we are pleased to highlight, and expound upon, our rationale here.

We used a propensity score method appropriate for small sample sizes, as discussed and published in the literature: "The propensity score method used to check covariate balance between groups and weight the data was inverse probability of treatment weighting (IPTW). IPTW gives correct estimations of treatment effect in small sample sizes" (appendix p 7). The appropriateness of this method has been shown down to n=40 by Pirracchio and colleagues[2] who "conducted a series of Monte Carlo simulations to evaluate the influence of sample size, prevalence of treatment exposure, and strength of the association between the variables and the outcome and/or the treatment exposure, on the performance of [IPTW]". Their findings show that "Decreasing the sample size from 1000 to 40 subjects did not substantially alter the Type I error rate, and led to relative biases below 10%".[2]

We took additional steps to decrease bias in specifying our propensity score model. Pirracchio and colleagues[2] state that "Including variables unrelated to the exposure but related to the outcome in the PS [propensity score] model decreased the bias and the variance as compared to models omitting such variables". Other findings from Monte Carlo simulation experiments concur with Pirracchio and colleagues[2] and are cited in our appendix. We state (appendix p 7) that "baseline covariates correlated to the outcome (MFQ clinical cut-off age 17 [ρ]≥0·10) [were] used to predict baseline mental health service contact", and indeed, several of these covariates are unrelated to the exposure (appendix p 10).

Furthermore, we ensured we were not overparameterising the logistic regression model used in generating the propensity score. Although many might be familiar with the rule of thumb that logistic regression models must have a minimum of ten events per predictor variable (EPV), Vittinghoff and McCulloch[3] cite this rule as being based on simulation studies, which only vary the number of events. Their simulation study[3] not only varied the number of events, but also the number of predictor variables, sample sizes, values of the regression coefficient for the primary predictor, multiple correlation of the primary predictor with the model covariates, and prevalence of a binary primary predictor. After examining 9328 scenarios with binary primary predictors and 3392 scenarios with continuous primary predictors, problematic scenarios (ie, CI coverage <93%, type I error rate >7%, or relative bias >15%) were encountered in 7% or less of the models with 5–9 EPV, predominantly with different numbers of predictors than used in our study. Indeed, such problems were still observed in models with 10–16 EPV. This outcome led them to conclude that "systematic discounting of results…from any model with 5–9 EPV does not appear to be justified",[3] and that relaxing the rules to 5–9 EPV is appropriate. We have not gone below this figure in models used in our paper.

After we estimated the propensity score, we took further steps (appendix p 7) to reduce estimate bias: "Stabilized IPTWs were used to reduce impact of extreme weights, thus reducing estimate bias."[4] In our appendix we also indicate that we ensured the propensity model was correctly specified, by checking and ensuring the balance of all covariates (not just those in the propensity model) following weighting, as exhorted by Austin.[5] Finally, as explained and cited in our methods,[1] "we restricted the sample to the region of common support—the range of propensity scores which were observed in both treated and untreated individuals" to further reduce estimate bias.

In summary, by using appropriate methods, given the small sample size

Lancet Psychiatry 2017
Published Online
March 27, 2017
http://dx.doi.org/10.1016/S2215-0366(17)30140-2

For more on **PubPeer** see https://pubpeer.com/publications/CCAB01FD663B858B53CD2DEE251AC6

For our **appendix** see http://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(17)30002-0/supplemental

and taking additional steps to reduce estimate bias, we conclude that our findings can be defended. We did, however, cite limitations related to a small sample size in our discussion, and exhort that future research assessing the association of mental health service contact with subsequent mental health should employ larger sample sizes when possible.

We acknowledge that an odds ratio overestimates a relative risk when the disease is not rare in the population, typically >10%. While depression was uncommon (3%, n=31) in the total sample of 1190, it was not uncommon (25%, 31 of 126) in the propensity sample consisting of those only with a mental disorder. Therefore, for clarity, our findings from the propensity score analysis should read "the odds of reporting clinical depression were higher in individuals…" as opposed to "seven times higher…".

We also wish to take this opportunity to clarify our sample size. While there were 1238 participants recruited to the cohort, we explain in the results that: "1190 adolescents had data for T1 current mental disorder and past-year mental health service contact". However, there were additional missing data from the outcome and covariates, bringing numbers of respondents with complete data for outcome and all covariates to 983 (83%) instead of 995 for T1, 717 (60%) instead of 778 for T2, and 769 (65%) instead of 856 for T3 (sixth column shows baseline sample size separate by each covariate; appendix p 9). Therefore, the numbers listed in our article in the first paragraph of the Results are different and we acknowledge this error in reporting. Listed in column 5 of table 1 are the sample sizes for completed data for outcome and all covariates and they add up, as expected, to the total across all time points (2257 + 140 + 72=2469).

To expand upon this sample size with imputation, we required each self-report questionnaire to be at least partially completed at one timepoint to be used in imputation. These imputed measures were computed (explained in appendix p 7) across all three timepoints with other measures related to attrition. This process resulted in the imputed sample size being smaller than 1190. The imputed sample size with covariates was in fact 1120, and we should have noted that in the main paper.

In column 2 of table 1, involving imputed adjusted data, the sample sizes indeed do not add up (eg, 2965 + 202 + 126=3293 not 3302). This n=9 difference was due to differential loss of participants for each of the subsamples as a result of inexact merging of different imputation files for some covariates, resulting in some of the original raw data with missing values being randomly included in some imputations. When this error was rectified, n=3360 for the first set of analyses (by age), n=3015 for the second set (by unaffected group), n=213 for the third (the disorder-only group), and n=132 for the fourth (the disorder-and-services group).

We recomputed our findings in columns 3 and 4 of table 1 and these are virtually unaltered: if anything, they become slightly stronger. As before, 14-year-old adolescents who had contact with mental health services in the past year had a greater decrease in depressive symptoms than those without contact (linear adjusted coefficient –1·68, 95% CI –3·18 to –0·19; p=0·028; quadratic adjusted coefficient –0·56, 95% CI –1·06 to –0·05; p=0·031). By T3, participants in the disorder-and-services group reported significantly fewer symptoms than did those in the disorder-only group (adjusted coefficient –4·76, 95% CI –8·75 to –0·77; p=0·020).

We declare no competing interests.

Sharon A S Neufeld, Peter B Jones, Ian M Goodyer
sasn2@medschl.cam.ac.uk

Department of Psychiatry, University of Cambridge, Cambridge CB2 8AH, UK

1 Neufeld SAS, Dunn VJ, Jones PB, Croudace TJ, Goodyer IM. Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK. *Lancet Psychiatry* 2017; **4:** 120–27.
2 Pirracchio R, Resche-Rigon M, Chevret S. Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. *BMC Med Res Methodol* 2012; **12:** 1–10.
3 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007; **165:** 710–18.
4 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11:** 550–60.
5 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; **46:** 399–424.

Child and adolescent mental health services: longitudinal data sheds light on current policy for psychological interventions in the community

Sharon A.S. Neufeld, Peter B. Jones and Ian M. Goodyer

Sharon A.S. Neufeld is a Research Associate, Peter B. Jones is a Professor and Ian M. Goodyer is a Professor, all at the Department of Psychiatry, University of Cambridge, Cambridge, UK.

Abstract

Purpose – *The purpose of this paper is to expand upon policy implications of a recent study assessing adolescent mental health service contact and subsequent depression.*

Design/methodology/approach – *Review of related evidence from academic and grey literature.*

Findings – *Studies assessing the role of mental health services in reducing mental disorder during adolescence are sparse, and even prevalence figures for adolescent mental disorders are out-of-date. Adolescent mental health service contact rates are shown to fall concurrent with budgetary decreases. School-based counselling is highlighted as an important source of help that may be at risk of being cut. Increased training of General Practitioners and school counsellors is needed to improve efficiency in specialist Child and Adolescent Mental Health Services (CAMHS).*

Practical implications – *Longitudinal studies of young people's mental health should include mental health service usage and its relationship with subsequent mental health outcomes.*

Social implications – *Funding cuts to CAMHS must be avoided, school-based counselling must be protected, and service referrers should be better trained.*

Originality/value – *This paper highlights the need for increased CAMHS data, sustained funding, and improved training for this vital service.*

Keywords *Policy, Depression, Child and adolescent mental health services, Community interventions*

Paper type *Viewpoint*

Received 20 March 2017
Accepted 18 April 2017

This study was funded by the Wellcome Trust (Grant No. 074296), and the National Institute for Health Research Collaboration for Leadership in Applied Health Research & Care for Cambridgeshire and Peterborough.

Young people's mental health problems account for many adulthood adversities, including greater likelihood of mental disorder (Jones, 2013), decreased income, decreased probability of being employed or maintaining a stable cohabiting relationship (Goodman *et al.*, 2011), and greater contact with the criminal justice system (Knapp *et al.*, 2016). Increasing the effectiveness and numbers treated by Child and Adolescent Mental Health Services (CAMHS) would therefore yield personal, economic, and societal benefits over the lifespan. Recently, a paper was published supporting the association of treatment-as-usual mental health service contact with improved mental health by late adolescence (Neufeld *et al.*, 2017). While the baseline mental health service data were collected a decade ago (2005/2006), such data are rare and provide insights relevant to current CAMHS. Some policy implications for CAMHS arise from this study, pertaining to the evidence base, funding, continuity of services, and training of referrers.

Studies assessing the role of mental health services in reducing mental disorder during adolescence are sparse, an oversight that must be addressed. The literature review conducted by Neufeld *et al.* (2017) found only six studies internationally which assessed the relationship of adolescent mental health services and subsequent mental health; none were as rigorous in simultaneously addressing non-randomisation of service usage, attrition, and clinical relevance as Neufeld *et al.* (2017). Longitudinal studies of young people's mental health should without question include mental health service usage and its relationship with subsequent mental health outcomes.

Policy would be more greatly informed if larger samples (e.g. national surveys) assessed a variety of psychosocial and perhaps physiological outcomes by individual diagnoses and/or treatment sectors. In the UK, such data are sorely lacking. Even up-to-date prevalence figures for mental disorders are glaringly absent: the most recent survey was last carried out in 2004 (Green *et al.*, 2005). Whilst a new survey will be carried out in 2017 (HM Government, 2017), the lag in collecting information that so vitally instructs service provision and planning for young people is concerning. For comparison, the Adult Psychiatric Morbidity Survey has been conducted twice as often – every seven years (McManus *et al.*, 2016). National surveys on young people's mental health must keep pace with those performed in adults, and they must rigorously assess the impact of service contact.

The data from the Neufeld *et al.* (2017) study were obtained prior to funding cuts to CAMHS, and taken together with other evidence, can make the case for how deleterious such austerity is to mental health service access for young people. From 2005/2006, Neufeld *et al.* (2017) found that 38 per cent of 14-year olds with a mental disorder had made contact with mental health services in the past year; however, in 2014/2015 only 25 per cent of all children and young people with a mental disorder had made such service contact (NHS England, 2015). During this time, between 2008/2009 and 2012/2013, CAMHS funding dropped by 5.4 per cent in real terms (Lamb, 2015) so that in 2012/2013, only 6 per cent of NHS' total mental health budget was spent on CAMHS (McShane *et al.*, 2015). Services from data in 2005/2006 that Neufeld *et al.* (2017) showed were related to an improvement in subsequent depression in young people have been overstretched due to austerity. For example, the number of young people attending A&E due to a psychiatric condition had more than doubled in 2014/2015 compared with 2010/2011 (Frith, 2017), indicating a breakdown in access to primary mental health services. In contrast, funding for adult mental health services was less impacted during this period, with NHS funds falling for the first time in a decade in 2011/2012 by 1 per cent in real terms (The Kings Fund, 2015). Encouragingly, among adults with mental disorder, service contact rose from 24 per cent in 2007 to 37 per cent in 2014 (McManus *et al.*, 2016), indicating that more stable funding can facilitate service access. It is heartening that the NHS aims to increase rates of young people's mental health service contact back up to 35 per cent by 2020/2021 (NHS England, 2015). However, society must acknowledge the suffering in our young people that has not been alleviated due to austerity measures, and resolve to ever-increase connection of young people with mental health services which are effective.

Data from the Neufeld *et al.* (2017) paper indicate the importance of school-based counselling, yet this source of help for young people must be protected. Neufeld *et al.* found that for those with a mental disorder, after specialist CAMHS, the next most used service was school counselling, and for those without a mental disorder, school counselling was the most highly used service. The current government has promised to provide funding for mental health first aid training for teachers in secondary schools (HM Government, 2017), enabling them to better identify those with mental health issues and connect them to support services (Mental Health First Aid England, 2016). However, this is against a backdrop of freezing school budgets until 2020/2021, the very budgets which typically fund school-based counselling (Frith, 2016). Increased identification of young people's mental health problems is commendable; at the same time leaving key services at risk of being cut is highly counterproductive, potentially increasing pressure on more specialist CAMHS. In total, 90 per cent of the cost of young people's mental health problems falls on the education system (Frith, 2017). The fact that young people who do not meet diagnostic criteria are referred back from specialist CAMHS to counselling in schools and General Practitioner (GPs) surgeries (Frith, 2016) underscores the importance of such services in preventing more serious problems. Funding for school-based counselling must be ring-fenced, whether it be funded through the education sector or NHS, to ensure young people have adequate service access prior to specialist CAMHS.

Service referrers, such as those in primary care or schools, need to be better trained in identifying the presenting features of mental disorders, to help prioritise specialist CAMHS for more serious cases. Neufeld *et al.* (2017) showed that improvements related to mental health service contact were only seen in those who had a clear need for services, as defined by the presence of a mental disorder. The findings imply that those meeting a diagnostic threshold of mental disorder may be more responsive to treatment, and that prioritising more serious cases could make the system

more effective. Currently, specialist CAMHS turn away 23 per cent of the children and young people referred to them for treatment by GPs or teachers (Frith, 2017). This implies two things: specialist CAMHS cannot cope with the population needs, and/or referrals need to be more appropriately made. Both may be true. Regardless, increased training of GPs and school counsellors could improve efficiency in specialist CAMHS by minimising subthreshold cases that are assessed but not uptaken by CAMHS. Such efficiency is sorely needed particularly when services are overburdened, and could help improve waiting times, which have been found to be unacceptably long (Frith, 2017). Indeed, it is clear that GPs could use more training in identifying mental disorders. The Royal College of General Practitioners (2016) reports that 90 per cent of people with mental health problems are managed in primary care. However, even in the recent past, most GP training has not included a rotation in mental illness (The Centre for Economic Performance's Mental Health Policy Group, 2012). Such gaps in training do nothing to mitigate the treatment gap: a meta-analysis showed that GPs correctly identified only 47.3 per cent of depression cases (Mitchell *et al.*, 2009). A high rate of access to individuals with mental disorders coupled with insufficient background knowledge to appropriately identify such cases represents an egregious missed opportunity. In addition to improved training for GPs, there should also be a clear pathway for training and supervision of school-based psychological workers to facilitate appropriate referrals to specialist CAMHS from the education sector. Care needs to be taken to ensure strong connections with primary care and schools to specialist CAMHS for young people who need additional help.

In sum, while the association of mental health services with a subsequent decline in adolescent depression (Neufeld *et al.*, 2017) is heartening, more studies of this nature are needed. Larger samples could enable a better understanding of the relative roles of various sectors in reducing specific mental disorders, to more specifically inform service provision. Neufeld *et al.*'s (2017) data suggest funding cuts have drastically reduced rates of service contact, and that school-based counselling is a well-utilised service, which may be at risk of being cut when the data suggest it should not be. Ensuring this service and better training of service referrers could help ease the strain on specialist CAMHS, and help all CAMHS work in a more integrated fashion.

References

Frith, E. (2016), *Progress and Challenges in the Transformation of Children and Young People's Mental Health Care*, Education Policy Institute, London.

Frith, E. (2017), *Health Select Committee Inquiry on Children and Young People's Mental Health – The Role of Education*, Education Policy Institute, London.

Goodman, A., Joyce, R. and Smith, J.P. (2011), "The long shadow cast by childhood physical and mental problems on adult life", *Proceedings of the National Academy of Sciences USA*, Vol. 108 No. 15, pp. 6032-7.

Green, H., McGinnity, A., Meltzer, H., Ford, T. and Goodman, R. (2005), *Mental Health of Children and Young People in Great Britain, 2004*, Office for National Statistics, Basingstoke.

HM Government (2017), *The Government's Response to the Five Year Forward View for Mental Health*, Department of Health, London.

Jones, P.B. (2013), "Adult mental health disorders and their age at onset", *The British Journal of Psychiatry*, Vol. 202, pp. S5-10.

Knapp, M., Ardino, V., Brimblecombe, N., Evans-Lacko, S., Iemmi, V., King, D., Snell, T., Murguia, S., Mbeah-Bankas, H., Crane, S., Harris, A., Fowler, D., Hodgekins, J. and Wilson, J. (2016), *Youth Mental Health: New Economic Evidence*, London School of Economics and Political Science, London.

Lamb, N. (2015), "Mental health services: children – Department of Health written question", available at: www.theyworkforyou.com/wrans/?id=2014-12-16.218865.h&s=mental+health (accessed 20 March 2017).

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

McShane, M., Rouse, J., Allington-Smith, P. *et al.* (2015), *Future in Mind: Promoting, Protecting and Improving our Children and Young People's Mental Health and Wellbeing*, Department of Health, London.

Mental Health First Aid England (2016), *Summary of Evaluations: Mental Health First Aid Training Courses*, Mental Health First Aid England, London.

Mitchell, A.J., Vaze, A. and Rao, S. (2009), "Clinical diagnosis of depression in primary care: a meta-analysis", *The Lancet*, Vol. 374, pp. 609-19.

Neufeld, S.A.S., Dunn, V.J., Jones, P.B., Croudace, T.J. and Goodyer, I.M. (2017), "Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK", *Lancet Psychiatry*, Vol. 4, pp. 120-27.

NHS England (2015), *Implementing the Five Year Forward view for Mental Health*, NHS England, Redditch.

Royal College of General Practitioners (2016), *The RCGP Curriculum: Professional and Clinical Modules*, Royal College of General Practitioners, London.

The Centre for Economic Performance's Mental Health Policy Group (2012), *How Mental Illness Loses out in the NHS*, London School of Economics and Political Science, London.

The Kings Fund (2015), *Mental Health under Pressure*, The Kings Fund, London.

Corresponding author

Sharon A.S. Neufeld can be contacted at: sasn2@medschl.cam.ac.uk

CHAPTER THREE

Development of a longitudinal transdiagnostic model of mental health over adolescence and emerging adulthood

3.1 Introduction to the chapter

The current classification of mental disorders as distinct diagnostic categories is hampered by high rates of overlap between syndromes (Kessler, Chiu, *et al.*, 2005). When various disorders or symptoms are modelled together, the common variance between these can be modelled with an overarching latent factor. All items in the model therefore load onto this general factor. A 'bifactor model' also allows additional distinct specific factors to be formed which share only a subset of items in common. Item loadings therefore bifurcate, with items indicating both the general factor and a specific factor. Reflecting the general overarching factor, a "common liability to all forms of psychopathology", or *p* has been proposed (for review see (Caspi and Moffitt, 2018)). The present paper does not attempt to equate all bifactor models of psychopathology with the model from which *p* was coined (Caspi *et al.*, 2014). The *p*-factor will instead be used as a short form to denote common mental distress, as the general factor has previously been construed (Stochl *et al.*, 2015; St Clair *et al.*, 2017). What is important to note is that this concept of shared variance between mental disorders arises from investigation of different datasets and analytic approaches.

Consistently, bifactor models of psychopathology have yielded a better fit compared with other less-hierarchical (ie: first-order) approaches such as a single factor model and/or correlated-factors model. Initially supported by nine studies reviewed in Lahey et al (Lahey *et al.*, 2017), there are now more studies supporting this (Böhnke and Croudace, 2016; Stochl *et al.*, 2015; Neumann *et al.*, 2016; Waldman *et al.*, 2016; Castellanos-Ryan *et al.*, 2016; Deutz *et al.*, 2016, 2018; Martel *et al.*, 2016; Snyder, Young and Hankin, 2017; St Clair *et al.*, 2017; Afzali *et al.*, 2018; Olino *et al.*, 2018; Pettersson *et al.*, 2018; Schaefer *et al.*, 2018; Lahey *et al.*, 2018a; McElroy *et al.*, 2018; Watts, Poore and Waldman, 2019; Constantinou *et al.*, 2019; Gluschkoff, Jokela and Rosenström, 2019). However, there have been criticisms of the reliance on model fit to favour the bifactor model over competing models (Murray and Johnson, 2013; Gignac, 2016; Reise *et al.*, 2016; Bonifay, Lane and Reise, 2017; Watts, Poore and Waldman, 2019). Consequently, stronger tests of the validity of the bifactor model have been advocated (Watts, Poore and Waldman, 2019), such as assessing the representation of factors by their indicators and construct replicability (Rodriguez, Reise and Haviland, 2016b). Such tests are crucial if one is to infer psychological meaning from *p* and the specific factors.

Initially modelled in adults (Lahey *et al.*, 2012; Caspi *et al.*, 2014), *p* has been replicated in studies of children (Olino *et al.*, 2014, 2018; Martel *et al.*, 2016; Neumann *et al.*, 2016; Pettersson *et*

al., 2018), adolescents (Tackett *et al.*, 2013; Laceulle, Vollebergh and Ormel, 2015; Noordhof *et al.*, 2015; Patalay *et al.*, 2015; Carragher *et al.*, 2016; Castellanos-Ryan *et al.*, 2016; Snyder, Young and Hankin, 2017; Afzali *et al.*, 2018; Schaefer *et al.*, 2018; Constantinou *et al.*, 2019), across childhood and adolescence (Deutz *et al.*, 2016, 2018; Murray, Eisner and Ribeaud, 2016; Waldman *et al.*, 2016; McElroy *et al.*, 2018; Watts, Poore and Waldman, 2019), across adolescence and young adulthood (St Clair *et al.*, 2017; Class *et al.*, 2019), and in young adulthood (Lahey *et al.*, 2018a). It is particularly important to deepen our understanding of *p* during adolescence and emerging adulthood, which are rife with social, physiological, and neurological changes (Arnett, 2000; Giedd, 2004; Liston *et al.*, 2006), and a concomitant rise in mental illness which has strong continuity into adulthood (Jones, 2013).

Initial work focused on modelling mental illness based on the presence or absence of a disorder (Lahey *et al.*, 2012; Caspi *et al.*, 2014), despite accumulating evidence that psychopathological constructs tend to be more continuous than discrete (Krueger *et al.*, 2018). Many subsequent models have captured a greater level of nuance by adding a degree of dimensionality to diagnostic groupings (Tackett *et al.*, 2013; Olino *et al.*, 2014, 2018; Laceulle, Vollebergh and Ormel, 2015; Castellanos-Ryan *et al.*, 2016; Martel *et al.*, 2016; Neumann *et al.*, 2016; Waldman *et al.*, 2016; Snyder, Young and Hankin, 2017; Schaefer *et al.*, 2018; Watts, Poore and Waldman, 2019). Some studies have dispensed with the hierarchy of diagnostic groupings altogether by modelling the underlying symptoms of mental illness (Böhnke and Croudace, 2016; Stochl *et al.*, 2015; Patalay *et al.*, 2015; Carragher *et al.*, 2016; Deutz *et al.*, 2016, 2018; Murray, Eisner and Ribeaud, 2016; St Clair *et al.*, 2017; Afzali *et al.*, 2018; McElroy *et al.*, 2018; Pettersson *et al.*, 2018; Constantinou *et al.*, 2019). Diagnostic-based models erroneously presume homogeneity within a diagnosis, whereas symptom-level models capture this heterogeneity and are freed of structural assumptions (Markon, 2010). As the latter models are based on symptoms, high scores on *p* and the specific factors are still reflective of putative mental illness. Despite typically covering a large degree of breadth (eg. internalizing, externalizing, and psychotic symptoms), most of the above *p*-factor models are focussed purely on mental illness. Yet mental health consists of more than simply the absence of mental illness (Keyes, 2002). Only two of the above studies addressed this issue by including aspects of mental illness and wellness (Murray, Eisner and Ribeaud, 2016; St Clair *et al.*, 2017), plus an additional study in adults (Böhnke and Croudace, 2016). Findings support the uniqueness of a positive factor, even after accounting for the valence of wording (St Clair *et al.*, 2017), which helps alleviate concern that the general or specific factors merely reflect a response style (Caspi and Moffitt, 2018).

Once a multidimensional model of mental health of sufficient breadth and detail has been devised, assessing such a model longitudinally is crucial, yet largely missing from the current body of research on *p* (Caspi and Moffitt, 2018). This must be rectified, given that *p* encompasses a liability to mental disorder with relevance across the lifespan. Strong longitudinal measurement

invariance - enabling valid interpretations of latent means and correlations over time (van de Schoot, Lugtig and Hox, 2012) - has rarely been established in models of *p* (Deutz *et al.*, 2016, 2018; Gluschkoff, Jokela and Rosenström, 2019), and criteria for determining this are inconsistent. Longitudinal association of *p* and the specific factors with external validation measures can help elucidate their meaning, yet only a few models of *p* have undertaken this (Lahey *et al.*, 2012; Caspi *et al.*, 2014; Castellanos-Ryan *et al.*, 2016; Pettersson *et al.*, 2018; Schaefer *et al.*, 2018; Class *et al.*, 2019; Constantinou *et al.*, 2019). Finally, our understanding of developmental change in *p* and the specific factors is limited, as phenotypic stability of these factors has not been ascertained in older adolescents and young adults, nor have growth curves of these factors been modelled at any developmental stage. The present chapter aims to redress these gaps in the literature in two sections. The first section (3A) focuses on measurement invariance, reviewing relevant cut-offs for bifactor models and developments regarding missing data. An empirical application is employed, where longitudinal measurement invariance, construct replicability, as well as factor representation and strength consistency is demonstrated. The second section (3B) aims to elucidate longitudinal construct validity and predictive validity, homotypic and heterotypic continuity, and developmental trajectories of the latent factors in the bifactor model of mental health presented in section 3A.

Section 3A: Measurement invariance testing in bifactor models with categorical indicators: review and application

3A.1 Introduction

In latent models such as the bifactor model, establishing strong longitudinal measurement invariance (MI) is vital to enable valid interpretations of latent means and factor correlations over time (van de Schoot, Lugtig and Hox, 2012). Making such comparisons in non-invariant models is biased and can result in artefactual differences emerging based on inconsistent structure or construct measurement (Brown, 2006; Chen, 2008). Establishing longitudinal invariance also supports the notion that *p* and specific factors are consistently measurable over time, arguing for the model's developmental applicability. However, to date, few existing bifactor models of mental health have established strong MI. Typically, when generating a bifactor model longitudinally, a model is generated separately at each time wave (Castellanos-Ryan *et al.*, 2016; Deutz *et al.*, 2016, 2018; Murray, Eisner and Ribeaud, 2016; Greene and Eaton, 2017; Snyder, Young and Hankin, 2017; McElroy *et al.*, 2018; Olino *et al.*, 2018; Gluschkoff, Jokela and Rosenström, 2019). If the same model yields acceptable fit across all waves, this shows that the number of factors and patterns of loadings are the same at each wave and the model demonstrates configural invariance over time (van de Schoot, Lugtig and Hox, 2012). However, having established this configural invariance, many longitudinal bifactor studies of mental health do not go on to more robustly test MI (Murray, Eisner and Ribeaud, 2016; Snyder, Young and Hankin, 2017; McElroy *et al.*, 2018; Olino *et al.*,

2018). In studies where strong MI has not been demonstrated over time, such as the above longitudinal bifactor models, the meaning of the construct and the levels of the underlying items are not known to be equal across time points (Muthén and Asparouhov, 2002). Therefore, interpretation of the longitudinal associations in such models is questionable.

MI is typically tested using multiple group confirmatory factor analysis with groups defined in a between-subjects manner (Van De Schoot *et al.*, 2015). Groups compared could be time waves (as in longitudinal invariance) or gender. Longitudinal MI can also be performed in a single-group model (with wide formatted data). While such an approach better accounts for the dependant nature of the data, the much larger array of data can result in improper solutions, particularly for models with many items (Vandenberg and Lance, 2000). Regardless of approach taken, invariance testing proceeds in a hierarchical manner using nested models: when constraints are added to a model, invariance is supported if the model fit does not substantially deteriorate compared to the less constrained model. If the model fits well in each group (or fits well in a multigroup or multiwave model without equity constraints), **configural invariance** is established (van de Schoot, Lugtig and Hox, 2012). Factor loadings can then be fixed to equity between groups to test for **metric (or weak factorial) invariance**. If the metric model does not exhibit substantially worse fit than the configural model, this indicates questions represent the same construct at each assessment, and therefore the same meaning can be attributed to these constructs (van de Schoot, Lugtig and Hox, 2012). If metric invariance is established, then **strong factorial (ie: scalar) invariance** can be tested by additionally making item intercepts or thresholds equivalent over time. This tests whether the levels of the underlying items (thresholds, for categorical indicators) or item intercepts (for continuous indicators) are also equivalent between groups (Muthén and Asparouhov, 2002). With categorical indicators, Muthén and Muthén argue that it is most appropriate to simultaneously constrain loadings and thresholds given that they jointly define item functioning (Muthén and Muthén, 2017). **Strict (or residual) invariance** (called **unique factor invariance** in categorical data (Liu *et al.*, 2017) can be determined by also fixing residual (ie: error) variances to be equal, although this is less common, and some argue is unnecessary (Vandenberg and Lance, 2000; Brown, 2006; Pendergast *et al.*, 2017; Seddig and Leitgöb, 2018).

There are a handful of bifactor models of mental health which have been tested for longitudinal MI beyond configural invariance, yet inconsistent criteria has been used to establish invariance. Three bifactor studies tested for metric invariance (weak factorial invariance) by constraining the factor loadings to be equal over time (Castellanos-Ryan *et al.*, 2016; Greene and Eaton, 2017; Olino *et al.*, 2018). These studies rejected metric invariance, but for different reasons: the metric model yielded unacceptable fit (Olino *et al.*, 2018), or the chi-square difference test between nested models ($\Delta\chi^2$), indicated that constrained models had significantly worse fit than unconstrained models (Castellanos-Ryan *et al.*, 2016; Greene and Eaton, 2017). While lack of metric invariance is undeniable in the former study, the latter studies had large sample sizes

(>2,000), and under these conditions, $\Delta\chi^2$ is highly sensitive to inconsequential differences, and thus may not be an accurate indicator of invariance (Putnick and Bornstein, 2016). When a more widely accepted indicator of invariance has been utilized (change in the comparative fit index, ΔCFI ≤.01 (Putnick and Bornstein, 2016)), strong longitudinal MI has been established in bifactor models (Deutz *et al.*, 2016, 2018; Gluschkoff, Jokela and Rosenström, 2019). These findings are consistent with a review which found that use of ΔCFI was related to higher levels of MI being established as compared to when $\Delta\chi^2$ was utilized (Putnick and Bornstein, 2016). Indeed, one of the above studies which rejected metric invariance based on $\Delta\chi^2$ reported their CFI values; from these values it is apparent that ΔCFI did in fact support metric invariance (Greene and Eaton, 2017). Having attained metric invariance, it is possible their model could have also attained strong MI. This exemplifies how conclusions regarding MI change depending on the criteria used.

An important caveat is that in all the preceding studies indicators were categorical or dichotomous, yet none tested strict invariance, as has been argued is required in this type of data (Millsap and Yun-Tein, 2004). Further, none of these studies discussed the impact that this non-normality may have on MI cut-offs, or how the dual factor loadings in the bifactor model may influence these cut-offs. Given that most bifactor models of mental health use categorical indicators (ie: diagnoses or Likert response items) this is an important issue to address. It is clear that a review of MI cut-offs in categorical data and in bifactor models is necessary in order to provide guidelines for applied researchers who utilize such models.

The replicability of the *p* model is also largely lacking in the current literature, despite being advocated as an important test of the validity of such models (Watts, Poore and Waldman, 2019). A review of bifactor models encompassing personality or mental health domains revealed that construct replicability as measured by coefficient *H* is rarely reported (Rodriguez, Reise and Haviland, 2016a), and was only reported in one of the above longitudinal studies (McElroy *et al.*, 2018). Coefficient *H* is important as it denotes how well a latent factor is represented by its contributing items, and therefore the factor's replicability across studies (Hancock, 2001). Coefficient *H* was originally termed "construct reliability" (Hancock, 2001; Hancock and Mueller, 2001); however, "construct replicability" more recently appears to be the preferred terminology (Rodriguez, Reise and Haviland, 2016b). Such terminology also differentiates coefficient *H* from omega. While some have used coefficient omega hierarchical as a measure of reliability for *p* and the specific factors (Böhnke and Croudace, 2016; Martel *et al.*, 2016; Afzali *et al.*, 2018; Constantinou *et al.*, 2019), coefficient *H* is more appropriate, as it is based on the optimal weighting of indicators for a given construct (Rodriguez, Reise and Haviland, 2016b). If *p* and the specific factors are to have relevancy in future work, these factors must be reproducible and thus demonstrate adequate levels of coefficient *H*.

Having established longitudinal MI and construct replicability, longitudinal studies can also help reveal the stability of *p* and the specific factors, and whether *p* decreases in strength over time as psychopathology becomes more specific (*p*-differentiation) or increases in strength as comorbidities grow (dynamic mutualism) (Murray, Eisner and Ribeaud, 2016). A handful of studies give bearing on the question of *p*-differentiation versus dynamic mutualism by assessing factor strength consistency. Studies spanning childhood to mid-adolescence (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018) revealed that both general and specific factors were consistent in their strength over time, with *p* consistently explaining the most model variance. However, the stability of these factors in a sample spanning mid-adolescence to young adulthood is unknown. Given the explosion of changes young people experience during this period on the social, physiological, and neurological level (Arnett, 2000; Giedd, 2004; Liston *et al.*, 2006), the stability of psychopathology factors is important to assess during this time.

Consistent factor representation – that *p* should be represented to an equivalent level by the model's constituent items – has recently emerged as an additional way to test the appropriateness of the *p*-model (Watts, Poore and Waldman, 2019). With *p* having been described as a general liability for psychopathology (Caspi and Moffitt, 2018), some have noted differential loading levels of various diagnoses on *p* across studies*,* citing this as an argument against the truly general nature of *p* (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019)*.* While *p* should indeed be well-represented by various forms of psychopathology, it seems improbable and heavily restrictive that *p* in every sample, across a diverse range of ages and settings, should be influenced equivalently by all measured dimensions of psychopathology. Further, what defines whether the magnitude of indicators loading on the general factor is "relatively uniform" (Krueger *et al.*, 2018)? Studies have defaulted to the rather simplistic idea that higher loading diagnoses define the general factor (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019). Others have argued against interpreting the general factor based on the highest loadings on that factor, because this in fact complicates interpretation (Oltmanns *et al.*, 2018). For example, if psychosis represents *p,* what can be made of all the other diagnoses which load onto *p* but are unrelated to psychosis? When loadings are considered more broadly, studies reveal *p* with good representation across psychopathology domains. Contrary to *p* in Lahey et al's study being defined by distress (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019), fears also loaded onto *p* to a large effect (an average of .65), with illegal drug dependence close behind (an average loading of .61, reflecting some externalizing behaviours (Lahey *et al.*, 2012)). Caspi et al's *p* is defined not only by thought disorder as some claim (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019), but had strong loadings (>.8) for major depression and generalized anxiety in addition to thought disorder diagnoses, and an average medium loading (.47) across externalizing diagnoses (Caspi *et al.*, 2014). It appears that the general nature of *p* is more apparent when substantial (moderate-to-large) loadings are considered across diagnoses.

The present study sought to extend prior work by firstly reviewing existing MI cut-offs under a variety of conditions, to create relevant invariance guidelines for a bifactor model with categorical indicators. Recent developments regarding missing data under these conditions also reviewed. This will then be followed by application of these invariance cut-offs in a bifactor model of self-reported items covering mental illness and wellness (St Clair *et al.*, 2017), assessed three times in a cohort spanning adolescence and emerging adulthood (Kiddle *et al.*, 2018). Following tests of gender and longitudinal invariance, construct replicability as well as strength consistency and factor representation of the general and specific factors shall be assessed. Adequate factor representation will be more broadly construed as moderate-to-high average factor loadings, as opposed to the more strict equal loadings applied elsewhere (Watts, Poore and Waldman, 2019).

I theorize that analyses shall support invariance of the present bifactor model, similar to what has been found in bifactor models of symptoms which yielded specific internalizing and externalizing factors (gender invariance (Deutz *et al.*, 2016, 2018; Greene and Eaton, 2017; Lahey *et al.*, 2018a)) and longitudinal invariance (Castellanos-Ryan *et al.*, 2016; Deutz *et al.*, 2016, 2018; Greene and Eaton, 2017; Gluschkoff, Jokela and Rosenström, 2019)). I predict that construct replicability will be higher in the general factor compared with the specific factors, as is typical (Rodriguez, Reise and Haviland, 2016a). It is possible that some of the specific factors will not show adequate construct replicability, especially those which are based on a smaller item pool. Consistent with prior work on younger samples, I hypothesize that both general and specific factors shall explain similar proportions of variance over time, with *p* consistently explaining the majority of variance (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018). I anticipate that at least moderate average loadings will be seen for the specific factors, with similar average loadings on *p* from original measures of mental illness and wellness.

3A.2 Review of existing measurement invariance cut-offs

To date, there have been no simulation studies which have examined goodness-of-fit indices of a bifactor model with categorical indicators; therefore the related literature will be reviewed. (Findings below are based on simulating multiple levels of invariance to at least the strong level unless specified.) To assess MI, fit indices are assessed to see how much worse fitting the more constrained model is to the less constrained model. A simultation study using continuous indicators recommended ΔCFI as the most appropriate goodness-of-fit index for MI (Cheung and Rensvold, 2002). In this and further simulation studies, non-invariance has been demonstrated when CFI worsens in the more constrained model by .01 or more (Cheung and Rensvold, 2002), .005 to .01 or more (Chen, 2007) or most strictly, by .002 or more (Meade, Johnson and Braddy, 2008). Despite the greater complexity in a bifactor model, with multiple loadings on most items (all must load onto the general factor and most also load onto a specific factor), metric invariance cut-offs for ΔCFI in bifactor models indicated by continuous variables fall within this range (non-invariance if

ΔCFI≥0.003 to 0.004 (Khojasteh & Lo, 2015)). However, when using ordered categorical indicators it is inappropriate to test MI assuming multivariate normality, as the factor structure may be differentially distorted across groups (Lubke & Muthen, 2004). Lubke and Muthen argue that use of the WLSMV estimator may detect variation, particularly in thresholds, which could otherwise be unapparent using an MLR estimator. In categorical data, the appropriateness of the above invariance cut-offs for CFI has been found to be acceptable in first-order models, particularly when models are correctly specified, sample sizes are large (≥1000), and if a small degree of non-invariance is acceptable (Sass, Schmitt, & Marsh, 2014). Specifically, Sass showed that Chen's 2007 invariance cut-offs for ΔCFI and ΔRMSEA adequately identified invariant models (non-invariance if ΔCFI≥.01, ΔRMSEA≥.01). Using ΔCFI cut-offs based on Meade et al's (2008) stricter criteria there was enough power to detect large amounts of non-invariance (if ΔCFI≥.002), and also small levels of invariance if the sample size was 1000 or more (Sass, Schmitt, & Marsh, 2014). Further, with sample sizes of 1000 or more, large non-invariance was able to be detected using Meade et al's cut-off of ΔRMSEA≥0.007.

An important caveat is that model fit criteria will differ based on the model parameters tested in each Monte Carlo simulation (Putnick and Bornstein, 2016). As it is not possible to test all model parameters, the results of each simulation only apply to the conditions tested. Therefore, a Bayesian approximate MI approach may be the most appropriate (Seddig and Leitgöb, 2018). In contrast with strong MI's requirement of factor loadings and intercepts to be exactly equal over time, in Bayesian approximate MI, these differences are assumed to be almost, but not exactly zero. The degree of flexibility in equivalence is specified by the prior distribution of the model, typically a normal distribution with a mean of zero and a small variance (<1). While a promising approach, it is not yet possible to test Bayesian approximate threshold invariance with polytomous items (http://www.statmodel.com). Further, the Deviance Information Criteria, a useful tool for Bayesian model comparison, is also not available with models utilizing categorical variables (ibid). Finally, the posterior predictive p-value which indicates a good-fitting model when non-significant (similar to the chi-square test of model fit) also is overly sensitive to large sample sizes, rejecting models with minor deviations from the hypothesized model (Hoofs *et al.*, 2018). Therefore, the statistics used to compare models as well as the influence of various priors (Seddig and Leitgöb, 2018) are not available with categorical variables or relevant in large sample sizes.

In sum, there are several arguments in support of determining invariance of a bifactor model with categorical indicators based on the more lenient ΔCFI<0.01 (Cheung and Rensvold, 2002; Chen, 2007). Firstly, models used to devise Meade et al's cut-offs have been criticised as being too strict (Little, 2013). Secondly, cut-offs generated in bifactor models are more liberal than Meade et al's cut-offs (Khojasteh & Lo, 2015). Thirdly, these invariance cut-offs appear to adequately identify invariant models in first-order categorical data (Sass et al., 2014). Fourthly, large sample sizes

(≥1000, typical of bifactor models) make ΔCFI more prone to rejecting invariance (Chen, 2007), indicating that a smaller ΔCFI may be overly strict. Fifthly, ΔCFI<0.01 has also been used to demonstrate invariance previously in bifactor models with categorical indicators (but based on Cheung and Rensvold's 2002 simulations only (Li *et al.*, 2018; Gluschkoff, Jokela and Rosenström, 2019)). Finally, it is doubtful that a small degree of non-invariance will influence conclusions related to the means of factor scores over time or across genders, and so using the above criteria is acceptable (Sass, Schmitt and Marsh, 2014). It is also noted that in large sample sizes, ΔRMSEA<0.007 rules out large deviations from non-invariance (Sass et al., 2014), and thus should be considered alongside ΔCFI.

The chi-square difference test should not be used to demonstrate MI in bifactor models, as this test has the power to detect inconsequential differences between groups in large sample sizes and in models of high complexity (Cheung and Rensvold, 2002), both typical of bifactor models. While the goodness-of-fit indices (eg. CFI, RMSEA) have also been shown to result in increased rejection of invariance as sample size increases, this was to a lesser degree than for the chi-square difference test (Chen, 2007). In simulations of MI in bifactor models, the magnitude of factor loading differences was shown to contribute most to change in goodness-of-fit indices, but sample size contributed most to change in the chi-square difference test (Khojasteh & Lo, 2015). Further, Yuan and Chan demonstrate that a significant chi-squared difference test does not imply that groups are not comparable, nor does a non-significant finding guarantee the model is not misspecified (Yuan & Chan, 2016). Any misspecification is problematic given that an acceptable base model is required before testing a further restricted model. Thus it is important to achieve correct specification of each increasingly restrictive model by assessing the model fit statistics of each model (Yuan & Chan, 2016). RMSEA<.05 should be obtained for all increasingly restrictive models, as done in categorical data (Millsap & Yun-Tein, 2004).

Some argue that in categorical data strict invariance must also be met in order to demonstrate invariance (Millsap and Yun-Tein, 2004; Liu *et al.*, 2017). Strict invariance (loadings, thresholds, and residual variances are all invariant) shows that the amount of explained variance for each item is the same across groups (van de Schoot, Lugtig, & Hox, 2012). With categorical indicators, this means that changes in the means, variances, and within-group covariances of the continuous latent responses are attributable to changes in the latent factor (Liu *et al.*, 2017). It has been recommended that the strict invariant model be assessed for acceptability of overall fit but not change in goodness of fit statistics (Millsap and Yun-Tein, 2004). While no goodness of fit cut-offs have been developed for this level of invariance in categorical data (let alone for higher-order models), ΔCFI<0.01 has been shown to be supportive of strict MI in first-order models with continuous indicators (Cheung and Rensvold, 2002; Chen, 2007). In absence of more relevant simulation studies, this cut-off could be cautiously applied to higher-order models with categorical indicators.

3A.3 Missing data in measurement invariance

The missing data inherent in longitudinal studies is important to properly address in MI testing. The weighted least squares mean and variance adjusted (WLSMV) estimator is ideal for latent modelling using categorical indicators (Beauducel and Herzberg, 2009). However, in the presence of missing data WLSMV is a limited information estimator as opposed to a full-information estimator such as maximum likelihood (FIML), which means that estimates will be biased due to missing data (Liu *et al.*, 2017). While multiple imputation of items might be considered as an approach to address missing data, the pooling of model fit statistics across multiply imputed datasets in order to evaluate MI has not yet been addressed in the literature ((Liu *et al.*, 2017); no studies have been found subsequent to this).

A recent study has shed light on the appropriateness of various estimators in testing multigroup MI with incomplete categorical items and asymmetric thresholds (Chen *et al.*, 2019). With incomplete categorical items and asymmetric thresholds, sample sizes of 1000, and missing data rates of 30% and 50%, WLSMV resulted in acceptably small levels of mean relative bias in loading estimates and their standard errors (Chen *et al.*, 2019). When Chen et al instead used FIML as the estimator for these categorical items, there were unacceptable levels of bias even at a sample size of 1000. While WLSMV did result in high type I error rates in the chi-squared difference test (Chen *et al.*, 2019), this test is problematic to use for MI testing as outlined above. Thus, WLSMV is an acceptable estimator for MI testing with missing data rates up to 50% in categorical indicators in large sample sizes. The key latent modelling programs, Mplus (Muthén and Muthén, 2017) and lavaan (Rosseel, 2012), both have this estimator available, with equivalent options for model parameterization, model fit statistics, and the handling of missing data (Liu *et al.*, 2017).

Differential levels of missing data over time can result in imbalanced sizes of groups being compared, which also has implications for MI. Simulations have shown that severely unbalanced groups can mask non-invariance (Yoon and Lai, 2018). However, when testing scalar invariance when one group was half the size of the other, ΔRMSEA was not adversely affected. Under the above conditions, scalar non-invariance was found to improve ΔCFI to a small degree (.005), but this was only tested in very small sample sizes (groups of 200 and 400 (Yoon and Lai, 2018). These findings underscore the importance of assessing several fit indices when drawing conclusions regarding invariance between unbalanced groups.

3A.4 An empirical example

The present example longitudinally extends a model published on baseline data of 118 items from measures of depressive, anxiety, obsessive, antisocial behavioural, and psychotic-like symptoms, as well self-esteem and well-being (for details, see Table 3A.1 and (St Clair *et al.*, 2017)). At baseline, compared with other models such as a single-factor or correlated factor model, the best-fitting and most theoretically plausible model was a Schmid-Leiman (S-L) bifactor

transformation (Brown, 2006) of a five-factor Confirmatory Factor Analytic model (ie: a general factor with five specific factors). A S-L transformation models the specific factors while estimating a general factor across all the items. This is computationally equivalent to adding a second-order factor over the first-order factors (Brown, 2006). All factors were set to be orthogonal (uncorrelated), congruent with the original theory of bifactor models that shared variance is captured in the general factor (Holzinger and Swineford, 1937). The final baseline model included 106 items as 12 items were removed due to very sparse endorsement or low loading on the general factor (<.30 (St Clair *et al.*, 2017). Based on item loadings, specific factors were construed as self-confidence, antisocial behaviour, worry, aberrant thinking, and mood – all independent of distress (*p*), which comprised the general factor. Additionally, a positive methods factor was specified to account for whether items were positively or negatively worded. This factor has not been accounted for in some prior bifactor models which simply recoded positively worded items (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018).

3A.4.1 Method

Sample: 2,403 young adults aged 14-24 were recruited from Cambridgeshire and London via general practitioners, schools, colleges, and purposive advertisements, comprising the NeuroScience in Psychiatry Network (NSPN) cohort (Kiddle *et al.*, 2018). Following an accelerated longitudinal design, multiple age-adjacent cohorts were recruited and assessed three times annually. Age cohorts (ages 14-15, 16-17, 18-19, 20-21, 22-24.99) were sex-stratified to ensure equal numbers of males and females in each cohort. The items in the bifactor model were obtained from a home questionnaire pack (HQP) mailed to participants' home. Three HQPs were collected annually (on average, HQP2 was collected 13.5 months (SD=3.64) after HQP1, and HQP3 27.2 months (SD=3.20) after HQP1). The self-reported HQPs consisted of questions relating to sociodemographics, personality, mood, behaviour, and environment.

Measures

The provenance of items used in the bifactor model are described on Table 3A.1.

Statistical analyses

All analyses were performed in Mplus Version 8 (Muthén and Muthén, 2017).

Table 3A.1: Bifactor model: original measures, constructs, item loadings[a]

		Number of items loading onto each factor						
Self-report measure[b]	**Key constructs assessed**	***P* (general factor)**	**Self-confid-ence**	**Anti-social**	**Worry**	**Aberrant thoughts**	**Mood**	**Positive worded items[c]**
Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)[1], 14 items	Mental well-being: evidence of current happiness, personal activity and personal achievement	13	13				8	13
Antisocial Behaviour Questionnaire (ABQ)[d], 11 items	antisocial behaviour: violating social norms, destructive behaviours, violence to people, lying and stealing	8		8				
Revised Children's Manifest Anxiety Scale (RCMAS)[2], 28 items	Generalized anxiety: physiological change, worry / oversensitivity, social concerns	25			7		1	
Revised Leyton Obsessional Inventory[3], 11 items	Obsessions and compulsions	11				8		
Schizotypal Personality Questionnaire (SPQ)[4], 11 items related to psychotic-like experiences[e]	total symptoms, hallucinations, delusions, and unusual perceptual experiences	10				9		
Rosenberg Self-Esteem Scale (RSES)[5], 10 items	self-esteem	10					9	5
Moods and Feelings Questionnaire (MFQ)[6], 33 items	depression	29					12	
TOTAL ITEMS		**106**	**13**	**8**	**7**	**17**	**30**	**18**

[a] model is based on St Clair et al, 2017, with MFQ24 dropped from the antisocial specific factor as done for the gender-invariant model (see results section for details). From original measures, 12 items were removed from the model due to very sparse endorsement or low loading on the general factor.

[b] all measures had a categorical response set. Participants were asked to rate their thoughts, behaviours, and emotions over the past two weeks, for all but SPQ (see below)

[c] method factor

[d] developed for use in a prior adolescent cohort study[7]

[e] SPQ items exhibiting a highly significant relationship (medium to large effect size) and face validity with psychotic-like experiences on the semi-structured PLIKS interview (PLIKSi[8]) from participants the In-Unit-Assessment subsample. See St Clair et al, 2017 for details.

References (Table 3A.1)

1. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Dinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health Qual Life Outcomes*. 2007;5:1-13. doi:10.1186/1477-7525-5-63
2. Reynolds CR RB. What I think and feel: a revised measure of children's manifest anxiety. *J Abnorm Child Psychol*. 1978;6(2):271-280.
3. 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
4. Raine A. The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophr Bull*. 1991;17(4):555-564.
5. Rosenberg M. *Society and the Adolescent Self-Image*. Princeton, N.J.: Princeton University Press; 1965.
6. Costello EJ, Angold A. Scales to assess child and adolescent depression – checklists, screens, and nets. *J Am Acad Child Psychiatry*. 1988;27:726–737.
7. Goodyer IM, Croudace T, Dunn V, Herbert J, Jones PB. Cohort Profile: Risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project. *Int J Epidemiol*. 2010;39(2):361-369. doi:10.1093/ije/dyp173
8. Horwood J, Salvi G, Thomas K, et al. IQ and non-clinical psychotic symptoms in 12-year-olds: Results from the ALSPAC birth cohort. *Br J Psychiatry*. 2008;193(3):185-191. doi:10.1192/bjp.bp.108.051904

Bifactor modelling*:* A Schmid-Leiman transformation was performed on all the HQP1 data as specified in the baseline model (St Clair *et al.*, 2017), since there were some additional participants recruited after publishing that paper.

There are different ways to specify a factor model to enable identification. In the present study, the metric of the factors was set by freeing all factor loadings and fixing factor variances to one (Muthén and Muthén, 2017). The other approach to setting the metric of a factor fixes a factor loading to one. However, if during MI testing a non-invariant loading is fixed to one, this can erroneously affect invariance tests for other loadings (Johnson, Meade and DuVernet, 2009). Further, discussions of how to choose this referent indicator are based on continuous but not categorical data (Bowen and Masa, 2015). Therefore, to avoid such problems the former approach to setting the metric of the factor was chosen.

Invariance testing**:** A single-group configural model (ie: wide formatted data) with HQP1-2 data was first tested. This resulted in a latent variable covariance matrix which was not positive definite. This error typically indicates a model is too complex for the data, so this approach was not pursued. Improper solutions in single-group models with many items have been noted by others (Vandenberg and Lance, 2000). Therefore, MI was tested using a multigroup confirmatory factor analytic framework to assess whether the latent factors were stable across gender and over time (van de Schoot, Lugtig and Hox, 2012).

To test gender invariance, the optimal Schmid-Leiman transformation reported at HQP1 (St Clair *et al.*, 2017) was performed separately in each gender at HQP1 to test appropriateness of fit in each group (Meade, Johnson and Braddy, 2008). Consistent with the original modelling, theoretically relevant modifications which were applicable to both genders were considered for low loadings (<.30 for the general factor and <0.15 for specific factors) and high modification indices (>100, due to the large number of participants in the present sample (St Clair *et al.*, 2017)). Multigroup modelling was then performed on both genders together (with gender modifications if criteria above was met) following the increasingly restrictive invariance tests outlined below. This model was then performed separately at each wave, to test for appropriateness of fit over time (Meade, Johnson and Braddy, 2008). Multigroup invariance testing was then performed on all three waves together. Metric invariance was not tested given the recommendation that it is most relevant to constrain loadings and thresholds simultaneously (Muthén and Muthén, 2017). Indeed, in the latent modelling package designed by these researchers (Mplus), it is not possible to constrain factor loadings without thresholds when an indicator loads onto more than one factor, as is the case in a bifactor model.

Specification of the increasingly constrained models was as follows. All models used the WLSMV estimator with theta parameterization, as appropriate for invariance testing of categorical indicators (Millsap and Yun-Tein, 2004). For the configural model, thresholds and factor loadings were free across groups, factor variances and residual variances were fixed at one in all groups,

and factor means fixed at zero in all groups (Muthén and Muthén, 2017). Correlated residuals were included in the model for indicators which were related but distinct (St Clair *et al.*, 2017). These residuals were fixed across groups for the configural model since group differences in correlated residuals would violate configural invariance (Joo & Kim, 2018) .For scalar and strict invariance models, the factor variances were fixed to one in one group and freed in the other group(s), and factor means were fixed to zero in one group and freed in the other group(s) (Muthén and Muthén, 2017). Both models had factor loadings and thresholds constrained to be equal across groups. In the scalar model, residual variances were fixed to one in one group and freed in the other group(s), but the strict invariance model had residual variances fixed to one in all groups. Further, as heterogeneous error covariances are a violation of strict invariance (Joo and Kim, 2018), the strict invariance model also fixed error covariances to be equal across groups.

To assess invariance, a CFI difference between the scalar and configural model of less than 0.01 was required (Cheung and Rensvold, 2002; Chen, 2007), as justified in the preceding review. Given the large sample size, I also considered ΔRMSEA<0.007 supportive of invariance, as indicated in categorical data (Sass, Schmitt and Marsh, 2014). The chi-squared difference test was not used to demonstrate MI, as this is biased by a large sample sizes (Meade, Johnson and Braddy, 2008) and high model complexity (Khojasteh and Lo, 2015), both features of the current data. Correct specification of each increasingly restrictive model was ensured by requiring RMSEA<.05 for all models, consistent with conventions for categorical data (Yuan & Chan 2016; Millsap & Yun-Tein 2004). As no goodness of fit cut-offs have been developed for strict invariance in categorical data (let alone for higher-order models), the strict invariant model was firstly assessed for acceptability of overall fit (Millsap and Yun-Tein, 2004), with a ΔCFI<0.01 compared to the scalar model additionally supportive of strict invariance (Cheung and Rensvold, 2002; Chen, 2007).

Construct replicability: Coefficient *H* measures reliability similar to Cronbach's alpha but is able to be calculated when data has a multilevel structure (McNeish, 2018). Just as Cronbach's alpha indicates how reliably a scale is measuring a unified construct, H denotes how well a latent construct is represented by its contributing items, and therefore how replicable across studies a factor might be (Hancock, 2001; McNeish, 2018). However, unlike Cronbach's alpha, the calculation of *H* allows for each item to contribute different amounts of information to the overall coefficient, resulting in more accurate reliability estimates (McNeish, 2018). Construct replicability was ascertained by calculating *H*, the ratio of variance explained by a latent variable relative to the unexplained variance (Rodriguez, Reise and Haviland, 2016b). Adequate construct replicability was reflected by *H*>.70 (Hancock and Mueller, 2001; Rodriguez, Reise and Haviland, 2016a), a cut-off which has been used in other bifactor studies (Gluschkoff, Jokela and Rosenström, 2019; Watts, Poore and Waldman, 2019).

Relative strength of the factors was measured over time using ***explained common variance (ECV)*** (Rodriguez, Reise and Haviland, 2016b). ECV is calculated for each factor by

dividing the variance explained by the factor by the total variance explained (ie: variance explained by the general and specific factors combined).

3A.4.2 Results

***Missing data*:** As for the baseline model (St Clair *et al.*, 2017), and consistent with other work (Brodbeck *et al.*, 2011), participants' data was included in the bifactor model at each time wave if they completed 85% of the original 118 items (100 items), and 85% of each original measure (Table 3A.1). Of the 2,403 participants, 99% (n=2,372), 69% (n=1,659) and 46% (n=1096) completed enough data to compute the bifactor model at HQP1-3 respectively. All available data was used for MI testing, resulting in unequal sample sizes between gender groups and over time (14% fewer males than females [1099 and 1273 respectively]; 30% fewer at HQP2 versus HQP1; 45% fewer at HQP3 versus HQP1). Simulation studies have shown that this level of imbalance between groups is not highly problematic for the ΔCFI or ΔRMSEA used in invariance testing (Yoon and Lai, 2018). At this level of missingness, WLSMV yields acceptably small levels of mean relative bias in loading estimates and their standard errors (Chen *et al.*, 2019).

***Initial bifactor modelling**:* As more participants were recruited after developing the initial bifactor model (St Clair *et al.*, 2017), the HQP1 model was re-run with these additional participants. This resulted in a model with nearly identical fit statistics to those originally reported (Table 3A.2), with all loadings significant and above cut-off (≥.30 for the general factor and ≥0.15 for specific factors). Five modification indices (MI)>100 indicated additional loadings onto specific factors. The highest MI, for a theoretically appropriate loading, was added to the model. However, as the chi-square did not improve by 10% in an already good-fitting model (Little, 2013), it was not necessary to make this or further modifications.

Table 3A.2: Mental health bifactor model fit indices

Model	n	Chi Square (χ^2)	df	# of para-meters	CFI	TLI	RMSEA	WRMR
HQP1[a] model (St Clair *et al.*, 2017)	2228	15859	5350	510	0.956	0.955	0.030	1.795
HQP1 present data	2372	16718	5350	510	0.955	0.954	0.030	1.846
HQP1 modified, females	1273	10935	5351	509	0.960	0.959	0.029	1.491
HQP1 modified, males	1099	8802	5351	509	0.968	0.967	0.024	1.314
Full HQP1 modified for gender	2372	16785	5351	509	0.955	0.953	0.030	1.851
Gender invariance (HQP1 data):	2372							
configural		19594	10717	1003	0.964	0.963	0.026	1.989
scalar		20293	11083	637	0.963	0.963	0.026	2.129
strict		18149	11204	516	0.972	0.972	0.023	2.257
HQP2 gender modified model	1659	12062	5351	509	0.970	0.969	0.027	1.549
HQP3 gender modified model	1096	9229	5351	509	0.974	0.973	0.026	1.386
HQP1-3 invariance	5127							
configural		36471	16083	1497	0.967	0.966	0.027	2.786
scalar		35876	16808	772	0.969	0.970	0.026	2.917
strict		29201	17030	530	0.980	0.981	0.020	3.066
Without ABQ11 (due to empty cells)								
HQP3 gender modified model	1096	9368	5248	506	0.972	0.971	0.027	1.385
HQP1-3 invariance	5127							
configural		37006	15774	1488	0.966	0.964	0.028	2.796
scalar		36238	16495	767	0.968	0.968	0.026	2.927
strict		28981	16735	527	0.980	0.981	0.021	3.075

[a] HQP=home questionnaire pack
*CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; WRMR = Weighted Root Mean Square Residual

***Measurement Invariance*:** When the original HQP1 model was run separately for each gender to identify any sources of misfit, all loadings were above cut-offs in the male model, but the female model had a low loading (0.126) of MFQ24 ("I was a bad person") on the antisocial behaviour specific factor. Also, ABQ11 ("I have deliberately hurt or been cruel to an animal") had a low loading on the general factor (0.24), and SPQ13 ("…had the sense that some person or force is around you, even though you cannot see anyone?") loaded slightly below cut-off (0.29). As some level of differential item function is acceptable, and the main goal in demonstrating invariance is to achieve convergence of increasingly constrained models, MFQ24 with the lowest loading was dropped from the antisocial specific factor (with loading maintained on the general factor). This modified model yielded excellent fit in separate gender models (Table 3A.2), with no additional loadings below cut-offs, and no items having modification indices above threshold.

The gender modified model also yielded excellent fit in the whole sample at each timepoint (Table 3A.2), with all loadings significant and above cut-offs. There were no theoretically-plausible item loadings with modification indices>100 at all three timepoints. However, only eight participants

endorsed ABQ11 (see above) at HQP3, which resulted in empty cells when correlated with two other items. Such sparsity could lead to inaccurate model estimation (Flora and Curran, 2004). The items in question were already dichotomized, so there were no response categories which could be collapsed. Therefore, in order to assess the potential influence of this sparsity in the data, sensitivity analysis was performed on longitudinal invariance models and HQP3 by dropping this item and comparing model fit with and without this item. Model fit for all invariance testing models and HQP3 remained excellent without this item (Table 3A.2). Without ABQ11, ΔCFI and ΔRMSEA differed by no more than .001 for MI models, and thus invariance conclusions were not altered (Table 3A.2). Further, using generated factor scores at HQP3, all factor autocorrelations with and without this item were 1.00. Together, these findings support minimal impact of ABQ11 sparsity at HQP3 on the model, and therefore inclusion of this item in the model is warranted.

Model fit was excellent for configural, scalar, and strict models across gender and time (Table 3A.2). Across gender and time, comparing scalar and strict models with the next least constrained model (configural and scalar, respectively), changes in CFI and RMSEA were less than even the strictest criteria (Meade, Johnson and Braddy, 2008): CFI declined at most by .001 and at worst RMSEA stayed the same. Thus strict MI was supported.

***Construct replicability, variance explained, factor representation*:** Across repeated assessments, all but anxiety specific factor demonstrated adequate construct replicability (*H*>.70 (Hancock and Mueller, 2001; Rodriguez, Reise and Haviland, 2016a), Table 3A.3). Indeed, antisocial and aberrant thoughts specific factors yielded *H*>.80, indicative of a well-defined latent variable, and *p* was very well defined at H=.99. Given that *H* is sensitive to high factor loadings (>.80 (Rodriguez, Reise and Haviland, 2016a)), I note that *H*-indices for the specific factors were not driven by a few high-loading items, as all specific factor loadings were <.78 (Table 3A.4). Three-quarters of the model variance was explained by the general factor at all three timepoints (Table 3A.3). Little variance was explained by the specific factors, which each exhibited stable rates of ECV over time (Table 3A.3).

Table 3A.3: Construct replicability and Explained Common Variance of general and specific factors from mental health bifactor model

	Construct Replicability (*H*)			**Explained Common Variance (ECV)**		
Latent Factor	**HQP1[a]**	**HQP2**	**HQP3**	**HQP1**	**HQP2**	**HQP3**
***P* (general factor)**	0.99	0.99	0.99	0.73	0.75	0.73
Self-Confidence	0.72	0.74	0.76	0.04	0.04	0.04
Antisocial	0.82	0.83	0.89	0.05	0.05	0.06
Worry	0.54	0.51	0.52	0.02	0.01	0.01
Aberrant thoughts	0.84	0.86	0.88	0.07	0.06	0.07
Mood	0.79	0.77	0.79	0.06	0.05	0.05
Positive worded items (ECV only)	n/a	n/a	n/a	0.04	0.04	0.04

[a] HQP=home questionnaire pack

P was well-represented by various domains of mental illness and wellness, with average item loadings from original measures ranging from .40-.69 for antisocial through to anxiety symptoms at HQP1 and similar values at subsequent waves (Table 3A.4). Effects for factor loadings on the general factor of .3, .5, and .7 have been considered to be small, medium, and large respectively (Gonzalez and MacKinnon, 2018), although others have considered values as low as .25 to reflect a moderate loading, and as low as .55 a large loading (Watts, Poore and Waldman, 2019). Considering these criteria, all original domains had at least a moderate average loading on the general factor, with the majority of domains loading strongly across items. Specific factors, which were allowed to have lower loadings for retention of items, had average loadings no lower than .30 across all waves. The .30 minimum is notable, as any item set with average loadings below this has been deemed "not worth considering" (Reise *et al.*, 2013). Each specific factor had items load on average at a moderate level, with the exception of antisocial, whose items loaded strongly on average (Table 3A.4). Specific factors were generally well-represented by their original measures, with the exception of the worry and mood specific factors, which were represented by less than half of the items from their original measure, with many of these items loading exclusively on the general factor (Table 3A.1). This likely explains why items from these measures loaded highest on *p* (Table 3A.4).

Table 3A.4: Factor representation by indicators

	# of items	Item loadings[a], mean (range)		
Latent Factor		**HQP1[b]**	**HQP2**	**HQP3**
***P* (general factor)**				
Symptoms by original measures				
Anxiety	25	.69 (.49-.81)	.74 (.59-.85)	.75 (.56-.84)
Depression	29	.68 (.47-.84)	.71 (.49-.86)	.72 (.49-.88)
Self-esteem	10	.66 (.54-.83)	.68 (.55-.84)	.68 (.53-.85)
Well-being	13	.53 (.36-.65)	.59 (.45-.72)	.60 (.47-.72)
Obsessions	11	.50 (.34-.72)	.56 (.40-.77)	.55 (.37-.71)
Psychotic-like	10	.42 (.33-.59)	.49 (.32-.68)	.48 (.36-.59)
Antisocial	8	.40 (.34-.53)	.43 (.33-.59)	.46 (.34-.61)
Specific factors[c]				
Self-Confidence	13	.40 (.31-.48)	.42 (.35-.50)	.44 (.38-.52)
Antisocial	8	.59 (.46-.71)	.59 (.41-.71)	.70 (.57-.76)
Worry	7	.36 (.24-.51)	.34 (.23-.49)	.35 (.23-.47)
Aberrant thoughts	17	.44 (.21-.65)	.44 (.21-.76)	.47 (.23-.77)
Mood	30	.33 (.22-.47)	.30 (.19-.41)	.32 (.17-.42)
Positive worded items[d]	18	.30 (.05-.60)	.31 (.10-.61)	.31 (.07-.64)

[a] absolute value of loadings
[b] HQP=home questionnaire pack
[c] see Table 3.A.1 for which items by original measure load onto specific factors
[d] for this methods factor, items were not required to load >.15

3A.5 Discussion

The present study moves forward the current body of work on the *p* factor by making the case for the importance of MI testing in bifactor models. Based on a review of the literature,

recommendations are made for MI cut-offs for bifactor models using categorical indicators. However, this review also highlights the need for more simulation studies to directly address MI in these models: only one bifactor study with continuous indicators has simulated one type of MI (Khojasteh and Lo, 2015), and no such studies exist with categorical indicators. As an alternative to more simulation studies, Bayesian approximate MI is a promising approach (Seddig and Leitgöb, 2018), yet the statistics used to compare models as well as the influence of various priors are not available with categorical variables or relevant in large sample sizes (Hoofs *et al.*, 2018).The appropriateness of MI testing with missing data was also reviewed. The WLSMV estimator is able to detect non-invariance with missing data rates up to 50% in categorical indicators in large sample sizes (Chen *et al.*, 2019). When testing MI in unbalanced groups, as often occur in longitudinal studies due to attrition, the present proposed cut-offs appear to be acceptable if the degree of imbalance is not severe (Yoon and Lai, 2018).

An empirical example was provided through the analysis of three repeated assessments of a bifactor model of mental illness and wellness in young people aged 14-24 (St Clair *et al.*, 2017). This model demonstrated strict gender and longitudinal invariance, and factors exhibited consistent construct replicability, strength, and item representation over time. These findings will now be discussed in more detail..

Measurement Invariance: The present study extends the limited prior work on gender and longitudinal invariance of bifactor models. Of the existing studies of bifactor models of psychopathology, only a small proportion have assessed gender invariance (Deutz *et al.*, 2016; Greene and Eaton, 2017; Snyder, Young and Hankin, 2017; Lahey *et al.*, 2018a) and longitudinal invariance (Castellanos-Ryan *et al.*, 2016; Deutz *et al.*, 2016; Greene and Eaton, 2017). In addition to fitting the data well at each of the three timepoints, the proposed bifactor model appears to be equivalent across males and females, and over three years of measurement. Demonstrating strict longitudinal MI indicates that changes in the self-reported items over time are attributable to changes in the latent factors over time (Liu *et al.*, 2017). This strengthens the validity of any longitudinal associations to be made more than prior longitudinal studies of *p* which have not addressed longitudinal invariance (Murray, Eisner and Ribeaud, 2016; Snyder, Young and Hankin, 2017; McElroy *et al.*, 2018). The same applies to any conclusions to be made about gender differences, given the demonstrated strict gender invariance of the proposed model.

Construct replicability**:** As anticipated, the general factor exhibited very high construct replicability (*H*) and *H* was higher than the specific factors(Rodriguez, Reise and Haviland, 2016a; McElroy *et al.*, 2018; Gluschkoff, Jokela and Rosenström, 2019). However, antisocial and aberrant thoughts specific factors also demonstrated *H* values indicative of well-defined latent variables likely to exhibit strong replicability across studies (Rodriguez, Reise and Haviland, 2016b, 2016a). This antisocial specific factor was better defined over time (average *H*=.85) than an externalizing specific factor found in children and adolescents (average *H*=.77 (McElroy *et al.*, 2018; Watts, Poore and

Waldman, 2019)) and adults (average *H*=.79 (Gluschkoff, Jokela and Rosenström, 2019)). Self-confidence and mood specific factors demonstrated adequate construct replicability (*H*≥.70 (Rodriguez, Reise and Haviland, 2016a). Only the anxiety specific factor was subpar, exhibiting comparable construct replicability to previously published a worry specific factor (average *H*=.52 versus .55 (Watts, Poore and Waldman, 2019) . Thus, interpretations with respect to the anxiety specific factor should be viewed with caution. However, the RCMAS measure used in the present bifactor model principally assesses social anxiety (Reynolds and Richmond, 1978). The anxiety specific factor could therefore be bolstered by including more varied measures of anxiety. Items loading on this specific factor pertaining to other constructs such as fearfulness and phobia could improve the definition of this construct.

Explained common variance: The general and specific factors explained consistent rates of common variance in the model over time, with the general factor explaining the majority of the common variance. Both these findings are consistent with prior work in children and young adolescents (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018). The demonstrated consistency in strength of the factors over time in this older cohort does not support either theory of dynamic mutualism (that *p* strength will increase over time as comorbidities increase) or *p*-differentiation (that liabilities to psychopathology become more specific with time (Murray, Eisner and Ribeaud, 2016)). The general factor explained approximately three-quarters of the common variance in the model at each timepoint, only slightly higher than the .7 reported in 14-year-olds (McElroy *et al.*, 2018). It has been suggested that bifactor models with ECV > .70 can be considered essentially unidimensional (Rodriguez, Reise and Haviland, 2016b). However, given adequate-to-high levels of latent construct replicability in all but one of the factors, further tests of the specific factors as well as the general factor are important to pursue.

Factor representation by items: In contrast to prior work (Watts, Poore and Waldman, 2019), *p* in the present study was consistently represented by many items of mental illness and wellness. Of the seven original domains, average loadings ranged from moderate-to-large (.40-.69) at baseline, with similar ranges found at other timepoints. Similarly, adolescent bifactor models of internalizing, externalizing, and disordered thought items had moderate-to-large average loadings on *p* (orthogonal specific factor models:.50-.55 (Carragher *et al.*, 2016); .43-.55 (Afzali *et al.*, 2018)). Moderate-to-large average loadings on *p* were also found for internalizing and externalizing items when adolescent models did not include disordered thought items (internalizing and externalizing respectively: .49, .37 (Patalay *et al.*, 2015)); .41, .61 for age 14 model (McElroy *et al.*, 2018) .53, .65 (Pettersson *et al.*, 2018); .34, .55 for adolescent self-report model (Deutz *et al.*, 2018)). While the literature reveals some variation in which items contribute more to *p* (eg: externalizing items loaded slightly higher on *p* than internalizing items when attention items were also included in the model on a separate specific factor (Deutz *et al.*, 2018; McElroy *et al.*, 2018; Pettersson *et al.*, 2018)), I argue that the level of variation is relatively trivial and does not support the notion that the defining

components of *p* varies considerably across studies (Watts, Poore and Waldman, 2019). It is likely that less variation in loadings is apparent when item loadings are averaged across domains, as opposed to measuring each diagnosis by one item; this could be an advantage of modelling *p* at the item level.

The specific factors all had moderate-to-large average loadings, although loadings below .15 were not retained on specific factors. Compared with other domains of mental illness and wellness, the original anxiety and mood scales had more items which singularly loaded on *p* and not also on the comparable specific factor. This may call into question the acceptability of these specific factors. However, in other adolescent studies modelling *p* at the item level, some internalizing items also only loaded significantly on *p* (Patalay *et al.*, 2015; Carragher *et al.*, 2016; Afzali *et al.*, 2018; McElroy *et al.*, 2018). This is consistent with the observation that complex bifactor structures (eg. with pure indicators of the general factor, specific factor cross-loadings, correlated errors) are common in practice (Abad *et al.*, 2017). Given the variety of symptoms reflected within a diagnostic category, it seems less problematic to a model of psychopathology if some symptoms do not load onto specific factors as anticipated, as opposed to whole diagnoses not loading as anticipated. This highlights another potential advantage to modelling psychopathology at the item level as opposed to the diagnostic level.

***Conclusion*:** The present study underscores the importance of demonstrating longitudinal measurement invariance in bifactor models and provides recommendations for cut-offs when such models utilize categorical indicators. Appropriately demonstrating MI will enable valid interpretations of latent means and correlations over time (van de Schoot, Lugtig and Hox, 2012), aiding clarification of the longitudinal role of *p* and the specific factors. However, further work is needed to clarify guidelines and expand practice of MI testing of bifactor models, especially with categorical indicators.

Based on guidelines justified herein, an empirical example of a bifactor model of mental illness and wellness exhibited strict gender and longitudinal invariance. The model largely withstood further tests of its appropriateness. *P* and all but one of the specific factors demonstrated adequate to excellent construct replicability at all timepoints. Over time, explained common variance was consistent, and therefore did not support either theory of dynamic mutualism or *p*-differentiation (Murray, Eisner and Ribeaud, 2016). Factor representation by items was also consistent over time, with average moderate-to-large item loadings on *p* from seven domains of mental illness and wellness, and moderate-to-large average loadings of items on specific factors. The present findings indicate that when factor representation is more expansively defined (ie: considering moderate-to-large loadings as opposed to the highest loadings only (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019)), the general factor is well represented by multiple forms of mental illness and wellness.

Findings from the empirical example highlight the issue of how to interpret the combined results from a variety of tests of model appropriateness. In the first study to discuss both construct replicability and factor representation by domains of mental illness, findings converged from both these tests for the uncorrelated specific factors in the bifactor model (Watts, Poore and Waldman, 2019). In the present example, these findings are not entirely convergent. The observation that many items from the original anxiety and mood scales exclusively loaded on *p* could raise questions of the appropriateness of their related specific factors. However, as the coefficient *H* values for the mood specific factor indicated acceptable construct replicability, this could indicate that only the worry specific factor is particularly questionable. Construct replicability has been clearly mathematically defined (Hancock, 2001; Hancock and Mueller, 2001), whereas no cut-off to denote an appropriate level of factor representation by indicators has been proposed (Krueger *et al.*, 2018; Watts, Poore and Waldman, 2019). In light of this, it seems advisable that observations based on factor representation by items be ancillary to construct replicability; if these are in conflict, construct replicability findings should be favoured.

Having established the above support for the present bifactor model, findings from further analyses can more confidently be interpreted with knowledge of the strengths and caveats of the model. Taking these steps enables individuals to better assess the utility of bifactor models moving forward. Such increased clarity should help further understanding of a psychological structure of mental health and illness over the youth period of the lifecourse. The next section of this chapter turns to establishing the putative meaning of the general and specific factors through longitudinal external validation. Homotypic and heterotypic continuity of these factors will also be assessed, and developmental trajectories probed.

Section 3B: Longitudinal validation and change of a transdiagnostic model of mental health over adolescence and emerging adulthood

3B.1 Introduction

The need for longitudinal validation of *p* and, particularly, the specific factors via theoretically relevant external variables has been highlighted (Caspi and Moffitt, 2018). It is apparent that individuals tend to move in and out of disorder categories over time (Rutter, Kim-Cohen and Maughan, 2006); longitudinally robust associations with external variables therefore appear to withstand these effects of sequential comorbidity (Caspi and Moffitt, 2018). Further, as developing a comprehensive structural model of psychopathology across the life course is important (Lahey *et al.*, 2017), validation of such a model based on a more enduring relationship with external variables is advantageous. However, only a handful of studies of children (Lahey *et al.*, 2015; Sallis *et al.*, 2019), adolescents (Patalay *et al.*, 2015; Castellanos-Ryan *et al.*, 2016; Pettersson *et al.*, 2018; Schaefer *et al.*, 2018; Laceulle *et al.*, 2019) and adults (Lahey *et al.*, 2012; Caspi *et al.*, 2014; Class

et al., 2019) have used longitudinal data to externally validate a bifactor model of mental illness. These studies generally did not assess the social environment; therefore the role of stressful life events, supportive friendships, and parenting styles, in predicting *p* and the specific factors over time is unknown. Further, none of the above studies ascertained multivariate predictors of the bifactor model across multiple domains, such as personality, social environment, and background measures. By revealing the relative importance of external validation measures, such analyses could help clarify the substantive meaning of the factors. External variable associations with the specific factors above and beyond the general factor would support their interpretation as common subfactors as opposed to uninterpretable residuals (Chen, West and Sousa, 2006). This combined with the previously established longitudinal invariance, construct replicability, and adequate item representation of the factors (see Section A) will aid interpretation of the factors over time.

After externally validating the bifactor model of psychopathology, it is important to ascertain phenotypic stability of this model in order to discern how these factors may change over development. Phenotypic stability has been assessed over childhood and adolescence (ages 2 to 16 (Castellanos-Ryan *et al.*, 2016; Murray, Eisner and Ribeaud, 2016; Snyder, Young and Hankin, 2017; McElroy *et al.*, 2018) and adulthood (Greene and Eaton, 2017). A further study has reported phenotypic stability of *p* and the specific factors from adolescence to early adulthood, although these longitudinal models were defined by different informants and different symptom dimensions (Class *et al.*, 2019). In the two studies of mid-adolescence (ages 13.5-16), homotypic continuity (measured by autocorrelations) in the general and specific factors was strong over 18 to 24 months (Castellanos-Ryan *et al.*, 2016; Snyder, Young and Hankin, 2017). However, homotypic continuity was more variable in early adolescence (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018) and from adolescence to early adulthood (Class *et al.*, 2019). Low levels (r~=0.1) of heterotypic continuity (correlations over time across factors) have been demonstrated during childhood as well as early and mid-adolescence (Castellanos-Ryan *et al.*, 2016; McElroy *et al.*, 2018; Olino *et al.*, 2018). While unsupported in another mid-adolescent sample (Snyder, Young and Hankin, 2017), statistical power may have limited the ability to discern heterotypic continuity in this study . Heterotypic continuity was not supported across adolescents to young adults nor in adults (Greene and Eaton, 2017; Class *et al.*, 2019). Given the heterogeneity of these findings, more work needs to be done to ascertain the role of these dynamic processes in young people. A large limitation of the above longitudinal studies is that none established strong measurement invariance of the latent constructs over time (see Chapter 3A); thus the validity of the reported longitudinal associations in the p- and specific factors is largely unknown. Such an oversight calls into question any conclusions made regarding dynamic mutualism or p-differentiation.

We cannot ascertain how liabilities to mental disorder change across the lifespan without a comprehensive assessment of their trajectories. Longitudinally characterizing the shape of *p* and the specific factors could help focus prevention efforts in key developmental periods of risk (ie:

when mean levels of *p* and/or specific factors are higher). However, the two epidemiologicalal studies of bifactor models of psychopathology with more than two timepoints (McElroy et al. 2018; Murray et al. 2016) made no attempt to characterize the growth curves of the bifactor scores (ie: how mean levels change over time and by age). Recently, two studies have reported on the bifactor model of psychopathology over the course of approximately 1.5 years during adolescent psychotherapy interventions and follow-up (Constantinou *et al.*, 2019; Aitken *et al.*, 2020). While these help elucidate the relationship of mental health treatment on *p* and the specific factors, our understanding of the developmental course of these psychological phenomena is still limited.

The present section seeks to extend prior work by determining, in a bifactor model of self-reported items of mental illness and wellness (St Clair *et al.*, 2017), using three repeated assessments from a cohort of adolescents and emerging adults (Kiddle *et al.*, 2018): (1) longitudinal construct validity and predictive validity using background, environmental, and personality variables, harmful behaviours, IQ, and BMI, (2) homotypic and heterotypic continuity, and (3) developmental trajectories of the latent factors. Having previously established gender and longitudinal invariance and adequate replicability of most of the factors (section 3A) aids interpretation of the factor scores over time.

I predict that the longitudinal data shall support the notion that the "distress independent" specific factors (St Clair *et al.*, 2017) do indeed index specific unique characteristics of mental wellness and illness, based on longitudinal relationships with background variables, measures of social environment, personality, cognition, harmful behaviours and BMI (Table 3B.1). Comparing this study's longitudinal findings with the current literature will help answer the question as to whether the additional specificity gained from the bifactor model is believable with respect to the interpretation of the specific factors. I predict homotypic continuity shall exist in the presence of a lesser degree of heterotypic continuity; that is, individuals will change over time differentially across the factors to a lesser extent than being consistent over time within factors (Castellanos-Ryan *et al.*, 2016; McElroy *et al.*, 2018; Olino *et al.*, 2018).

Finally, I hypothesize that the general and specific factors will show gender-differentiated developmental trajectories largely consistent with previously reported findings of their associated mental disorders (Kessler *et al.*, 2007). Such an assertion is supported by the validation of a self-reported symptom-level bifactor model by interview-based psychiatric diagnoses three years later (Brodbeck *et al.*, 2014). Given the present study's accelerated longitudinal design, I utilized accelerated growth modelling (Duncan and Duncan, 2004) to obtain growth curves of the general and specific factors over the full age range, revealing rate and shape of change of the factors from ages 14-27. Modelling the p-factor in such a way elegantly summarizes how mean levels of risk for mental disorder change over time.

Table 3B.1: Hypothesized relationship of validity measures with general and specific symptom factor scores[a]

A) Baseline[b] predictors of HQP3 symptom factors	*P* (general factor)	Self-confidence	Antisocial	Internalising	Aberrant thoughts
Background:					
Index of Multiple Deprivation	+[1]		+[1,2]	-[1], +[2]	+[3]
Ethnicity (non-white minority=1)			+[4,5]		+[6]
Social Environment:					
Stressful life events	+[c]		+[2]	+[2]	+[7]
Friendships	-[1]	+[8]	-[1]	-[1]	-[9]
Parenting					
Positive	-[c]	+[10]	0[11]	-[12]	-[9]
Inconsistent discipline	+[c]		+[13]	+[12] (only mood)	
Poor supervision	+[c]		+[13]	+[12] (only mood)	
Physical punishment	+[14]		+[14]	0[14]	+[15]
Personality[d]:					
Prosocial	-[1,16,17]	+[10e]	-[1,16,17]	+[1], 0[16,17]	0[18]
Negative emotionality	+[16,19–23], 0[17]	-[10]	+[16,17,23], 0[19–22]	+[17,19,20,23], 0[16,21,22]	+[24]
Daring	+[25], 0[16,17]	+[10e]	+[16,17,25]	-[16,25], 0[17]	+[25]
Impulsivity	+[20–23,25]		+[21–23,25], 0[20]	-[21], 0[20,22,23,25]	+[25]
In-Unit Assessment Measures					
Intelligence Quotient total	-[19]		0[19]	0[19]	-[26]
Childhood Trauma					
Total trauma	+[15,19]		0[19]	0[19]	+[15]
Physical abuse	+[14,15]		+[14]	0[14]	+[15]
Sexual abuse	+[14,15]		0[14]	0[14]	+[15]
Neglect	+[14,15]		0[14]	0[14]	+[15]
B) HQP1 symptom factors predicting outcomes[f]					
Harmful behaviours					
Alcohol use	0[27]		+[27]	+[27]	
Cannabis use	+[14,27,28]		+[14,27,28]	0[14,27,28]	
Other illicit substance use	+[14,27,28]		+[14,27,28]	0[14,27,28]	
Non-suicidal self-injury	+[14,29]		+[14], 0[29]	0[14,29]	
Body Mass Index	+[14,27]		-[14], 0[27]	0[14,27]	

[a] Based on cross-sectional data with a few exceptions[14,15,17,19,20,27,28]. In red if findings are from a bifactor model. "0" reflects no association reported, blank reflects no findings reported in the literature. All bifactor models but reference 14 had anxiety and mood combined as one internalizing specific factor. Any unique findings are specified.
[b] HQP=Home Questionnaire Pack. Baseline measures are from HQP1 unless specified as in-unit assessment.
[c] in the absence of prior association with a bifactor model, association with the general factor was hypothesized if an association was found in multiple facets of mental health
[d] the first three personality measures are from the CADS[16,17]. Reported associations with the NEO Five-Factor Inventory[30] enables comparison of CADS and NEO subscales, allowing hypothesis generation based on studies which have used the NEO[10,19,20]. The remaining studies have used alternate, but related measures[1,21–23,25]. Self-reported personality findings are cited unless only parent reports were available, as is the case with the studies of children[21,22].
[e] being the only study on this subject, a consistent pattern across NEO conscientiousness, agreeableness and extraversion subscales was deemed indicative of prosociality, as supported by a factor model of NEO and CADS[30], and similarly extraversion was deemed indicative of daring disposition.
[f] all measures are from wave 3 except BMI, which is from in-unit assessment wave 2

References (Table 3B.1)

1. Patalay P, Fonagy P, Deighton J, Belsky J, Vostanis P, Wolpert M. A general psychopathology factor in early adolescence. *Br J Psychiatry*. 2015;207(1):15-22. doi:10.1192/bjp.bp.114.149591
2. Amone-P'Olak K, Ormel J, Huisman M, Verhulst FC, Oldehinkel AJ, Burger H. Life Stressors as Mediators of the Relation Between Socioeconomic Position and Mental Health Problems in Early Adolescence: The TRAILS Study. *J Am Acad Child Adolesc Psychiatry*. 2009;48(10):1031-1038. doi:10.1097/CHI.0b013e3181b39595
3. Kirkbride JB, Jones PB, Ullrich S, Coid JW. Social deprivation, inequality, and the neighborhood-level incidence of

psychotic syndromes in East London. *Schizophr Bull*. 2014;40(1):169-180. doi:10.1093/schbul/sbs151
4. Backman H, Laajasalo T, Jokela M, Aronen ET. Interpersonal Relationships as Protective and Risk Factors for Psychopathy: A Follow-up Study in Adolescent Offenders. *J Youth Adolesc*. 2018;47(5):1022-1036. doi:10.1007/s10964-017-0745-x
5. Paget A, Parker C, Heron J, et al. Which children and young people are excluded from school? Findings from a large British birth cohort study, the Avon Longitudinal Study of Parents and Children (ALSPAC). *Child Care Health Dev*. 2018;44(2):285-296. doi:10.1111/cch.12525
6. Kirkbride JB, Hameed Y, Ioannidis K, et al. Ethnic minority status, age-at-immigration and psychosis risk in rural environments: Evidence from the SEPEA study. *Schizophr Bull*. 2017;43(6):1251-1261. doi:10.1093/schbul/sbx010
7. Shakoor S, Zavos HMS, Haworth CMA, et al. Association between stressful life events and psychotic experiences in adolescence: Evidence for gene-environment correlations. *Br J Psychiatry*. 2016;208(6):532-538. doi:10.1192/bjp.bp.114.159079
8. O'Brien SF, Bierman KL. Conceptions and perceived influence of peer groups: Interviews with preadolescents and adolescents. *Child Dev*. 1988;59(5):1360-1365.
9. Crush E, Arseneault L, Danese A, Jaffee SR, Fisher HL. Using discordant twin methods to investigate an environmentally mediated pathway between social support and the reduced likelihood of adolescent psychotic experiences. *Psychol Med*. 2019:1-8. doi:10.1017/s0033291719001983
10. Lampropoulou A. Personality, school, and family: What is their role in adolescents' subjective well-being. *J Adolesc*. 2018;67:12-21. doi:10.1016/j.adolescence.2018.05.013
11. Waller R, Hyde LW, Klump KL, Burt SA. Parenting is an environmental predictor of callous-unemotional traits and aggression: A monozygotic twin differences study. *J Am Acad Child Adolesc Psychiatry*. 2018;57(12):955-963. doi:10.1016/j.jaac.2018.07.882
12. Yap MBH, Pilkington PD, Ryan SM, Jorm AF. Parental factors associated with depression and anxiety in young people: A systematic review and meta-analysis. *J Affect Disord*. 2014;156:8-23. doi:10.1016/j.jad.2013.11.007
13. Burke JD, Loeber R, Birmaher B. Oppositional Defiant Disorder and Conduct Disorder: A review of the past 10 years, part II. *J Am Acad Child Adolesc Psychiatry*. 2002;41(11):1275-1293. doi:10.1097/00004583-200211000-00009
14. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is there a general factor of prevalent psychopathology during adulthood? *J Abnorm Psychol*. 2012;121(4):971-977. doi:10.1037/a0028355
15. Schaefer JD, Moffitt TE, Arseneault L, et al. Adolescent victimization and early-adult psychopathology: Approaching causal inference using a longitudinal twin study to rule out noncausal explanations. *Clin Psychol Sci*. 2018;6(3):352-371. doi:10.1177/2167702617741381
16. Tackett JL, Lahey BB, van Hulle C, Waldman I, Krueger RF, Rathouz PJ. Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *J Abnorm Psychol*. 2013;122(4):1142-1153. doi:10.1037/a0034151
17. Class QA, Van Hulle CA, Rathouz PJ, Applegate B, Zald DH, Lahey BB. Socioemotional dispositions of children and adolescents predict general and specific second-order factors of psychopathology in early adulthood: A 12-year prospective study. *J Abnorm Psychol*. 2019;128(6):574-584. doi:10.1037/abn0000433
18. Sullivan S, Lewis G, Wiles N, Thompson A, Evans J. Psychotic experiences and social functioning: A longitudinal study. *Soc Psychiatry Psychiatr Epidemiol*. 2013;48(7):1053-1065. doi:10.1007/s00127-013-0715-x
19. Caspi A, Houts RM, Belsky DW, et al. The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci*. 2014;2(2):119-137. doi:10.1177/2167702613497473
20. Castellanos-Ryan N, Briere FN, O'Leary-Barrett M, et al. The structure of psychopathology in adolescence and its common personality and cognitive correlates. *J Abnorm Psychol*. 2016;125(8):1039-1052. doi:10.1037/abn0000193
21. Neumann A, Pappa I, Lahey BB, et al. Single nucleotide polymorphism heritability of a general psychopathology factor in children. *J Am Acad Child Adolesc Psychiatry*. 2016;55(12):1038-1045.e4. doi:10.1016/j.jaac.2016.09.498
22. Olino TM, Dougherty LR, Bufferd SJ, Carlson GA, Klein DN. Testing models of psychopathology in preschool-aged children using a structured interview-based assessment. *J Abnorm Child Psychol*. 2014;42(7):1201-1211. doi:10.1007/s10802-014-9865-x
23. Hankin BL, Davis EP, Snyder H, Young JF, Glynn LM, Sandman CA. Temperament factors and dimensional, latent bifactor models of child psychopathology: Transdiagnostic and specific associations in two youth samples. *Psychiatry Res*. 2017;252:139-146. doi:10.1016/j.psychres.2017.02.061
24. Goodwin RD, Fergusson DM, Horwood LJ. Neuroticism in adolescence and psychotic symptoms in adulthood. *Psychol Med*. 2003;33(6):1089-1097. doi:10.1017/S0033291703007888
25. Carragher N, Teesson M, Sunderland M, et al. The structure of adolescent psychopathology: A symptom-level analysis. *Psychol Med*. 2016;46(5):981-994. doi:10.1017/S0033291715002470
26. Arseneault L, Cannon M, Fisher HL, Polanczyk G, Moffi TE, Caspi A. Childhood trauma and children' s emerging psychotic symptoms : A genetically sensitive longitudinal cohort study. *Am J Psychiatry*. 2011;168:65-72. doi:10.1176/appi.ajp.2010.10040567
27. Laceulle OM, Chung JM, Vollebergh WAM, Ormel J. The wide-ranging life outcome correlates of a general psychopathology factor in adolescent psychopathology. *Personal Ment Health*. 2019. doi:10.1002/pmh
28. Pettersson E, Lahey BB, Larsson H, Lichtenstein P. Criterion validity and utility of the general factor of psychopathology in childhood: Predictive associations with independently measured severe adverse mental health outcomes in adolescence. *J Am Acad Child Adolesc Psychiatry*. 2018;57(6):372-383. doi:10.1016/j.jaac.2017.12.016
29. Deutz MHF, Geeraerts SB, van Baar AL, Dekovice M, Prinzie P. The Dysregulation Profile in middle childhood and adolescence across reporters: factor structure, measurement invariance, and links with self-harm and suicidal ideation. *Eur Child Adolesc Psychiatry*. 2016;25(4):431-442. doi:10.1007/s00787-015-0745-x
30. Lahey BB, Rathouz PJ, Applegate B, Tackett JL, Waldman ID. Psychometrics of a self-report version of the Child and Adolescent Dispositions Scale. *J Clin Child Adolesc Psychol*. 2010;39(3):351-361. doi:10.1080/15374411003691784.

3B.2 Method

3B.2.1 Sample

The sample is predominantly drawn from the NSPN self-report HQP data, as in section 3A. For external validation, some additional measures were obtained from an in-unit assessment (IUA) of a subsample of participants (n=785) drawn from the original 2,403. Equally by age- and sex-strata, participants were invited to take part in the IUA using the order in which they been enrolled in NSPN (assumed to be random; (Kiddle *et al.*, 2018)). IUA1 was obtained approximately 6 months after HQP1 (mean=5.10, SD=4.60) and nearly two years before HQP3 (mean=22.16 months, SD=6.05), and IUA2 was obtained nearly two years after HQP1 (mean=22.78 months, SD=4.61).

3B.2.2 Measures

Factor scores from the longitudinal bifactor model developed in Section 3A were utilized in all models. External validation measures are described on Table 3B.2. All external validation measures have acceptable to excellent internal consistency (Cronbach's alpha $\geq$.70).

3B.2.3 Statistical analyses

Multiple imputation of HQP data: Due to a large amount of missing data at HQP3 (54% for some variables), and the categorical nature of the indicators in the present study and some outcome variables, multiple imputation of the data across all three timepoints was performed. Missing data patterns were first ascertained, as multiple imputation is appropriate under conditions of Missing Completely at Random (MCAR) or Missing at Random (MAR; (White, Royston and Wood, 2011). A significant Little's MCAR test (χ^2 (8, n=2,389)=25.01, p=.002) indicated that data were not MCAR (Little, 1988). However, a wealth of baseline variables were related to attrition (Table 3B.3), indicating that data are MAR and therefore appropriate for imputation. Bifactor scores were generated separately at each timepoint using the gender modified model (section 3A.4.2). All three waves of these factor scores and external validation variables (Table 3B.2) were imputed together in wide format, thereby incorporating within-person autocorrelations into the model (Gottfredson, Sterba and Jackson, 2017). 54 imputations were performed (White, Royston and Wood, 2011) using Bayesian estimation of an unrestricted model in Mplus, specifying categorical and continuous variables (Asparouhov and Muthén, 2010). All HQP analyses below are performed on this imputed data in Mplus version 8 (Muthén and Muthén, 2017).

Table 3B.2: Measures used for external validation

A) Home Questionnaire Pack measures	Timeframe	items	Notes
Background:			
Index of Multiple Deprivation[1]	current	1	Relative deprivation for small areas in England, based on post codes
Ethnicity (white versus non-white minority)	n/a	1	Recoded from the following categories: White, Asian, Black, Mixed, Other
Life Events Questionnaire[2,3]	past 18 months	12	Potentially stressful changes, disasters, accidents, and/or illnesses to self and close others. Sum score is used of events that were quite or very unpleasant, sad, painful and were upsetting for more than 2 weeks.
Cambridge Friendships Questionnaire	current, or not specified	8	Assesses number, availability, and quality of friendships; developed in adolescents[4]; shown invariance, reliability, and validity[5]
Alabama Parenting Questionnaire	Present (*if living at home)* or prior *(if not living at home)*		Typical occurrence of interactions are rated. The first three subscales are from the short form[6]; the physical punishment subscale is from the original measure[7]. Subscales in the HPQ1 data are correlated to a small-to-moderate degree (r=.10-.27, p<.0001).
Positive Parenting[6]		3	
Inconsistent discipline[6]		3	
Poor supervision[6]		3	
Physical punishment[7]		3	
Personality:			
Child and Adolescent Disposition Scale[8,9]	past year		Developed to test the relationship between dispositions and symptoms. Thus, items exclude anything linguistically similar to symptoms, making associations in the present study as unconfounded as possible.
Prosocial		13	
Negative emotionality		7	
Daring		5	
Barratt Impulsivity Scale[10]	not specified	30	Assesses attention, motor impulsiveness, self-control, cognitive complexity, perseverance, and cognitive instability.
Harmful behaviours:			
Drugs Alcohol and Self Injury			Developed for use in an adolescent cohort study[4]. Past year NSSI was ascertained by combining past month responses with those queried for the remaining 11 months. Reliability and validity of this question has been previously demonstrated[11].
Drunkenness	past month	1	
Cannabis Use		1	
Other illicit substance use		1	
Non-suicidal self-injury (NSSI)	Past year	2	
B) In-Unit Assessment Measures			
Wechsler Abbreviated Scale of Intelligence[12]	current		Full-scale IQ based on vocabulary and matrix reasoning.
Childhood Trauma Questionnaire	experiences from ages 0-16		Physical and emotional neglect subscales were combined to be comparable with prior studies[13,14]. The emotional abuse subscale was not reported on, being colinear with the total subscale (r=.83).
Total trauma			
Physical abuse			
Sexual abuse			
Neglect (physical + emotional)			
Body Mass Index (BMI)	current	2 (weight and height)	Age-adjusted BMI for ages 14-19 was calculated based on World Health Organization guidelines[15]. For older participants, BMI=weight/height^2.

References (Table 3B.2)

1. Smith T, Noble M, Noble S, Wright G, McLennan D, Plunkett E. *The English Indices of Deprivation 2015: Research Report*.; 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464485/English_Indices_of_Deprivation_2015_-_Technical-Report.pdf.
2. Goodyer IM, Herbert J, Tamplin A, Altham PME. Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents. *Br J Psychiatry*. 2000;177:499 –504. doi:10.1192/bjp.177.6.499
3. Goodyer I, Kolvin I, Gatzanis S. Recent undesirable life events and psychiatric disorder in childhood and adolescence. *Br J Psychiatry*. 1985;147:517-523.
4. Goodyer IM, Croudace T, Dunn V, Herbert J, Jones PB. Cohort Profile: Risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project. *Int J Epidemiol*. 2010;39(2):361-369. doi:10.1093/ije/dyp173
5. van Harmelen AL, Gibson JL, St Clair MC, et al. Friendships and Family Support Reduce Subsequent Depressive Symptoms in At-Risk Adolescents. *PLoS One*. 2016;11(5):e0153715. doi:10.1371/journal.pone.0153715
6. Elgar FJ, Waschbusch DA, Dadds MR, Sigvaldason N. Development and Validation of a Short Form of the Alabama

Parenting Questionnaire. *J Child Fam Stud*. 2007;16(2):243-259. doi:10.1007/s10826-006-9082-5
7. Shelton KK, Frick PJ, Wootton J. Assessment of parenting practices in families of elementary school-age children. *J Clin Child Psychol*. 1996;25(3):317-329. doi:10.1207/s15374424jccp2503_8
8. Lahey BB, Applegate B, Chronis AM, et al. Psychometric characteristics of a measure of emotional dispositions developed to test a developmental propensity model of conduct disorder. *J Clin Child Adolesc Psychol*. 2008;37(4):794-807. doi:10.1080/15374410802359635
9. Lahey BB, Rathouz PJ, Applegate B, Tackett JL, Waldman ID. Psychometrics of a Self-Report Version of the Child and Adolescent Dispositions Scale. 2010;39(3):351-361. doi:10.1080/15374411003691784.
10. Patton JH, Stanford MS, Barratt ES. Factor structure of the barratt impulsiveness scale. *J Clin Psychol*. 1995;51(6):768-774. doi:10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1
11. Wilkinson PO, Qiu T, Neufeld S, Jones PB, Goodyer IM. Sporadic and recurrent non-suicidal self-injury before age 14 and incident onset of psychiatric disorders by 17 years: Prospective cohort study. *Br J Psychiatry*. 2018;212(4). doi:10.1192/bjp.2017.45
12. Axelrod BN. Validity of the Wechsler Abbreviated Scale of Intelligence and other very short forms of estimating intellectual functioning. *Assessment*. 2002;9(1):17-23. doi:10.1177/1073191102009001003
13. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is there a general factor of prevalent psychopathology during adulthood? *J Abnorm Psychol*. 2012;121(4):971-977. doi:10.1037/a0028355
14. Schaefer JD, Moffitt TE, Arseneault L, et al. Adolescent Victimization and Early-Adult Psychopathology: Approaching Causal Inference Using a Longitudinal Twin Study to Rule Out Noncausal Explanations. *Clin Psychol Sci*. 2018;6(3):352-371. doi:10.1177/2167702617741381
15. de Onis M, Onyango A, Borghi E, Siyam A, Nishida C, Siekmanna J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ*. 2007;85:660-667. doi:10.2471/BLT

Table 3B.3: Relationship (rho[a]) of baseline measures with missingness in HQP2-3 [b]

Baseline measures (HQP1)	HQP2 missing	HQP3 missing
Background:		
Age	.05+	.03
Sex (1=female)	-.19*****	-.22*****
Centre (1=Cambridge)	.07+	.00
Index of Multiple Deprivation	.04	.02
Ethnicity (non-white minority=1)	.06	.06
Social Environment:		
Stressful life events	-.01	.00
Friendships	-.04	-.05*
Parenting		
Positive	-.04	-.03
Inconsistent discipline	.06**	.07***
Poor supervision	.14*****	.15*****
Physical punishment	.06**	.04
Personality:		
Prosociality	-.11*****	-.09*****
Negative emotionality	.01	.00
Daring	.07****	.09*****
Impulsivity	.11*****	.13*****
Harmful behaviours		
Drunkenness	.10***	.09***
Cannabis Use	.19*****	.16*****
Other illicit substance use	.17**	.17**
Non-suicidal self-injury	.02	.02
Bifactor model		
P (general factor)	.05+	.04
Self-Confidence	-.02	-.02
Antisocial	.11*****	.13*****
Worry	-.08****	-.07***
Aberrant thoughts	.01	.00
Mood	.03	.02
Positive worded items	-.02	-.02

[a] point biserial, polychoric, or tetrachoric rho for relationship of missingness with continuous, categorical, and dichotomous baseline measures respectively.
[b] HQP=Home Questionnaire Pack.
+ p<.05; *p<=.01; **p<.005; ***p<=.001; ****p<.0005; *****p<=.0001

FIML analyses of IUA data: Outcomes from the In-Unit Assessment (IUA) were continuous and thus Full Information Maximum Likelihood (FIML) with the inclusion of auxiliary variables was an appropriate way to deal with missing data (Collins, Schafer and Kam, 2001). For models predicting HPQ3 factor scores, HPQ1 and 2 factor scores were included as auxiliary variables; for the model predicting BMI at IUA2, BMI from IUA1 was the auxiliary variable. Auxiliary variables were included as extra dependent variables in the model, specified as recommended (Graham, 2003). Factor scores were not MCAR (for each sample, varying from n=759-780 due to some missingness in predictors, each Little's MCAR test had a p<.01 (Little, 1988)). However, for each model, factor scores exhibited covariate-dependent missingness (CDM), indicating lack of bias in these patterns of missingness (Little, 1995). As the case for MCAR data, when data exhibit CDM, employing FIML to analyse the data does not result in biased estimates. CDM of factor scores was ascertained with inclusion of the predictor (in bold) and confounds in the models (**IQ**, sex, age, ethnicity, centre: χ^2 (336, n=779)=370.50, p=.10; **total trauma**, minimization (to control for positive response bias in the Childhood Trauma Questionnaire (MacDonald *et al.*, 2016)), age: χ^2 (224, n=770)=231.39, p=.35; **neglect**, minimization, age, sex: χ^2 (280, n=758)=308.05, p=.12; **physical abuse**, minimization, age: χ^2 (224, n=767)=238.10, p=.25; **sexual abuse**, minimization, sex: χ^2 (224, n=767)=239.73, p=.22). Finally, BMI at IUA1 and 2 was not MCAR (χ^2 (2, n=711)=7.93, p=.02) but CDM was supported (**HPQ1 factor scores**, age, sex: χ^2 (20, n=711)=20.40, p=.43). IUA data was modelled in Stata 14.2 (StataCorp, 2015).

Validity: **Concurrent and discriminant validity** were assessed by concordance with the pattern of hypothesised associations (and lack thereof) with external variables as denoted in Table 3B.1. Evidence of these together can be taken to establish **construct validity**. All models were age, sex, and centre adjusted if these variables confounded the association (related to predictor and outcome with a standardized estimate>=.1, or p<.05 for HQP1 cohort or p<.10 for IUA cohort (due to IUA smaller sample size)). Next, a multivariate model was analysed, which included any validation measures from HQP1 cohort which predicted each factor score with a standardized estimate>=.1, or p<.05. Different predictors were allowed for each factor score.

Predictive validity of HPQ1 bifactor scores was determined in models where these factor scores predicted HPQ3 harmful behaviours and BMI. In addition to addressing confounding as above, these models also controlled for the positive methods specific factor. While the worry specific factor exhibited inadequate construct replicability (see Section 3A.4 and Table 3A.3), the validity of this factor will still be explored.

Phenotypic stability of the general and specific factors was assessed using cross-lagged panel analysis on imputed data (note: models utilizing FIML in raw data resulted in poor fit (TLI<.74)).

***Age analysis*:** Given the accelerated longitudinal design of the present study, accelerated growth modelling was performed on imputed data to determine the individual growth curves of the general and specific factors across the full age range studied (Duncan and Duncan, 2004). (Note: models did not converge when performed on raw factor scores). There were two adjacent cohorts in which there were not enough participants to model the data, (even if the cohorts were merged): those born in 2000 (n=12) and 2001 (n=2). Thus, these participants were removed, making the sample size 2,389 for the accelerated growth analyses. There was one participant born in 1987 who was grouped with the cohort born in 1988 to create a cohort size of 76. The remaining cohorts, born from 1988 to 1999, ranged in sample size from 100 to 361. A common growth model was specified across the age cohorts (separately for the general and specific factors), simultaneously analysed to determine a growth curve over the full 14-year-period from age 14 to 27. This sequential cohort multi-group analysis links adjacent waves of data, with each cohort having a different pattern of "planned missing data" (Duncan and Duncan, 2004). As data is missing from all cohorts due to the design of the study and not participant characteristics, these data can be presumed to be missing completely at random and therefore parameters can be appropriately estimated (Enders, 2010). Models are fit in each cohort, with invariance constraints across cohorts with overlapping time points. Growth curves were assessed for significant linear and quadratic change, with the higher-order term being retained if significant. HQP1 analysis revealed that in all factors there was either a main effect of gender or significant gender differences across age cohorts (St Clair *et al.*, 2017). Therefore, in each accelerated growth model I assessed the relationship of gender with intercept, slope, and quadratic term (the latter if relevant). As for the bifactor models, model fit was assessed using the Comparative Fit Index (CFI), Tucker–Lewis index (TLI), and the Root-Mean-Square Error of Approximation (RMSEA). Models with CFI and TLI > 0.95 and RMSEA < 0.05 are considered to exhibit good fit (Hu and Bentler, 1999; van de Schoot, Lugtig and Hox, 2012), with fit of CFI > 0.90 (Tucker and Lewis, 1973; Bentler, 1990, 1992) and RSMEA < 0.08 (MacCallum, Browne and Sugawara, 1996) being acceptable.

Despite having previously established measurement invariance from HQP1-3 (Chapter 3A), it would be ideal to assess measurement invariance of the bifactor model across all cohorts in the accelerated growth model (Marsh, Parker and Morin, 2016). However, this was not possible due to the complexity of the model, as well as the presence of missing at random data due to attrition in HQP2&3 coupled with relatively small cohorts. Cohorts consisted of 361 participants or less, with some items in the bifactor model at HQP3 only being present for 46% of participants – conditions which yield unacceptable levels of bias with FIML or the WLSMV estimator (Chen *et al.*, 2019). FIML alternatives which attempt to deal with the categorical nature of the data (maximum likelihood robust, or with a logit or probit link; (Chen *et al.*, 2019)) were not appropriate for the present model as they do not allow covariances for categorical variables in the model. Thus, attrition bias was best accounted for by using the multiply imputed data.

3B.3 Results

3B.3.1 Validity results are shown on Table 3B.4, with multivariate findings presented in Table 3B.4. For models in Table 3B.5, intercorrelations between predictors was <=.40. Model fit was excellent (RMSEA=.011, CFI=.987, TLI=.963). Worth noting is that the large sample size of the present study results in very significant *p*-values for some small effects. Thus, congruency of effects with the hypotheses in Table 3B.1 will be the primary focus.

Table 3B.4: External validation of general and specific mental health bifactor scores (standardized coefficients, age, sex, and centre adjusted[a])

A. Construct validation[b] of HQP3 bifactor scores						
	***P* (general factor)**	**Self-confidence**	**Antisocial**	**Worry**	**Aberrant thoughts**	**Mood**
Adjustments						
Age	.07*	.03	-.14***	.15***	-.03	-.00
Sex (female)	.24***	.11	-.32***	.32***	.03	-.05
Centre	.06	.05	.14+	.07	-.01	.02
Background						
Deprivation	.03	.00	.11***	-.05	.04	.04
Ethnicity (minority=1)	.05	.01	.19**	.06	.22**	.12
Social Environment						
Stressful life events	.25***	.00	.02	.02	.05	-.05
Friendships	-.36***	.14***	-.00	.11***	-.10***	-.16***
Parenting						
Positive	-.24***	.16***	-.04	.14***	.01	-.18***
Inconsistent discipline	.13***	-.00	.10***	-.10***	.10***	-.01
Poor supervision	.16***	-.02	.17***	-.10***	.05	.07+
Physical punishment	.12***	-.03	.09***	-.07+	.06+	-.01
Personality						
Prosocial	-.07*	.21***	-.13***	.13***	-.06+	-.17***
Negative emotionality	.41***	-.04	.02	-.02	.09***	.03
Daring	-.04	.12***	.20***	-.11***	.03	-.07+
Impulsivity	.28***	-.07+	.10***	-.15***	.08**	.07+
IUA1 Measures						
Intelligence Quotient	-.08	.01	-.04	.04	-.17***	-.04
Childhood Trauma						
Total trauma	.26***	.09	.02	-.05	.08	-.01
Physical abuse	.12**	-.01	.01	-.05	.04	-.00
Sexual abuse	.01	.04	-.07	.10	.01	-.12+
Neglect	.16*	.04	.03	-.09	.06	.06
B. Predictive validity[c] of HQP1 bifactor scores						
Harmful behaviours						
Alcohol use	.02	-.07+	.12***	.06+	-.06	-.09**
Cannabis use	.14***	.03	.22***	.06	-.06	.03
Other illegal drugs	.16**	-.00	.15**	.07	-.02	-.00
Non-suicidal self-injury	.44***	-.07	.00	-.02	.05	.10+
Body Mass Index (BMI)	.12**	.06	.06	-.07	-.01	.06

[a] adjusted if standardized estimate>=.1 with predictor and outcome or p<.05 for HQP1 cohort or p<.10 for IUA cohort (due to IUA smaller sample size). Estimates based on imputed data unless IUA1 measures or BMI, where full information maximum likelihood was used. HQP=Home Questionnaire Pack. IUA=In-Unit Assessment.
[b] validation measures from HQP1 (n=2403) unless specified as IUA1 measures (n=757-778). HQP3 is an average of 27 months after HQP1 and 22 months after IUA1.
[c] all but BMI obtained from HQP3 (n=2372, based on HQP1 data). BMI obtained at IUA2 (n=711), an average of 23 months after HQP1.

standardized estimate= .10-.19; .20-.29; .30+
+p<.05, *p<.01, **p<=.005, ***p<=.001

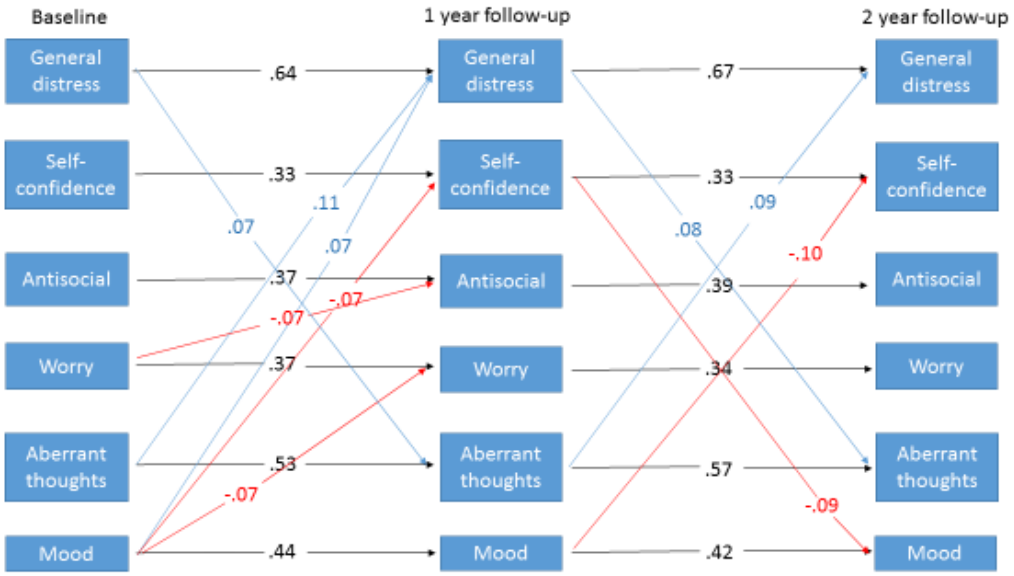
Table 3B.5: Multivariate model of HQP1 external validation measures[a] predicting bifactor scores two years later

	P (general factor)	Self-confidence	Antisocial	Worry	Aberrant thoughts	Mood
Background						
Age	.05^{+}		-.13***	.12***		
Sex (female)	.15**	.07	-.14*	.21***		
Centre			.03			
Deprivation			.07^{+}			
Ethnicity (minority=1)			.11		.19*	.08
Social Environment						
Stressful life events	.11***					
Friendships	-.20***	.07^{+}		.06^{+}	-.05	-.10***
Parenting						
Positive	-.09***	.08*		.09**		-.11***
Inconsistent	.02		.02	-.05	.05	
Poor supervision	.01		.08^{+}	.02		.01
Physical punishment	.03		.05	-.02	.01	
Personality						
Prosocial	.03	.17***	-.13***	.09*	-.03	-.12***
Negative emotionality	.21***				.05	
Daring		.09***	.19***	-.12***		-.05
Impulsivity	.11***	-.01	-.00	-.06	.03	-.00
Model R^2	.26***	.07***	.13***	.11***	.03**	.06***

[a] HQP=Home Questionnaire Pack. Variables from Table 3B.4 section A (excluding IUA1 measures) were entered into the model if standardized estimate>=.1 (or p<.05) with predictor and each HQP3 factor. Different predictors were allowed for each factor score. Intercorrelations between predictors was <=.40. RMSEA=.011, CFI=.987, TLI=.963.
standardized estimate= .10-.19; .20-.29; .30+
$^{+}$p<.05, *p<=.01, **p<=.005, ***p<=.001

3B.3.2 Homotypic continuity was demonstrated by significant autoregressive effects for all factors in the cross-lagged panel model (Figure 3B.1), with the highest stability coefficient for *p*. (The model was adjusted by age and sex, as centre was unrelated to all HPQ2 factors, p>=.26. Model fit was acceptable on all indices but TLI: RMSEA=.056, CFI=.908, TLI=.716). Of the specific factors, coefficients ranged from a medium autocorrelation for self-confidence to a large autocorrelation for aberrant thoughts. **Heterotypic continuity** is demonstrated by significant cross-lagged effects among the factors, albeit to small effect size or less. The reciprocal associations of *p* with aberrant thoughts, and self-confidence with mood were consistent across HQP1-HQP2 and HQP2-HQP3, with no other heterotypic associations consistent over both time periods.

Figure 3B.1: Cross-lagged associations[a] of general and specific mental health bifactor scores at repeated assessments, adjusted for age, sex, and centre.



[a] Standardized path coefficients of imputed data, significant at p< .01. Methods factor controlled for in model (excluded from figure for simplicity). Effect sizes as per correlations: 0.1=small, 0.3=medium, 0.5=large.

3B.3.3 Developmental trajectories were best modelled linearly (quadratic terms p>0.10) with the exception of antisocial, which exhibited a significant quadratic decline (p=.042; Table 3B.6; Figure 3B.2). All models adjusted by sex yielded acceptable or good fit (criteria in methods). *P* and worry increased over the developmental period while antisocial and aberrant thoughts decreased. At age 14 (model intercept), females exhibited higher levels of *p*, but over the developmental period males increased more than females on this factor, as evidenced by the significant relationship of slope with sex (Table 3B.6). Males were higher on the antisocial factor at age 14 but did not improve faster than females. There were no mood gender differences at age 14, but males increased more than females over the developmental period. Females were higher on worry at age 14 and worsened over the developmental period compared to males.

Table 3B.6: Accelerated growth models of mental health bifactor scores (sex-adjusted, imputed data). A: Unstandardised parameter estimates (male=1). B: Goodness-of-Fit Indices.

A.

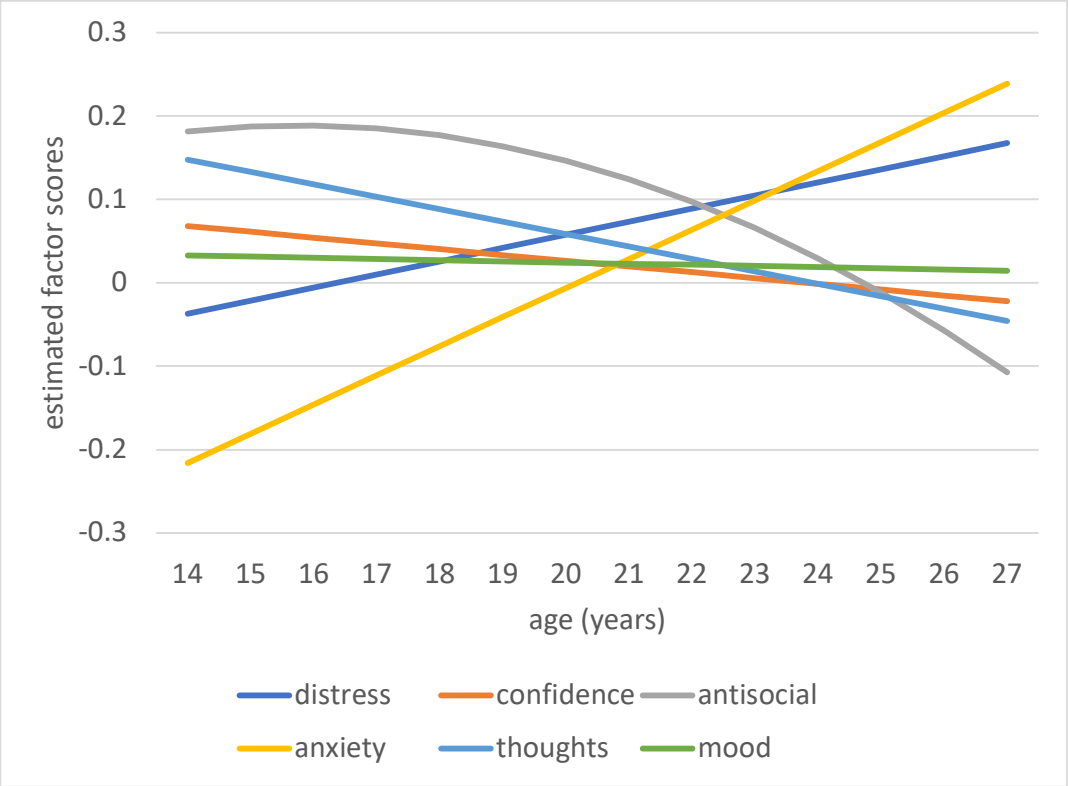
Mental health factors	Intercept mean	Slope[a] mean	Intercept with Slope	Intercept with Sex	Slope with Sex	Intercept variance	Slope variance
P (general factor)	0.04*	0.13*	-0.04	-.17***	0.24***	0.54***	0.31
Self-Confidence	0.02	-0.09	0.02	-0.04	0.02	0.26***	-0.15
Antisocial (quadratic)	0.16***	-0.21*	-0.13	0.15***	0.02	0.19***	0.63
Worry	-0.05***	0.34***	0.04	-0.13***	-0.12*	0.20***	-0.10
Aberrant thoughts	0.08***	-0.16***	-0.01	0.01	-0.06	0.29***	0.20
Mood	0.02	0.01	0.03	-0.01	0.19***	0.25***	0.07

B.

	X2 M (SD)[b]	CFI *M (SD)*	TLI *M (SD)*	RMSEA *M (SD)*	WRMR *M (SD)*
P (general factor)	158.87 (9.29)	.977 (.006)	.987 (.004)	.037 (.005)	3.60 (0.15)
Self-confidence	180.69 (14.69)	.926 (.019)	.957 (.011)	.048 (.006)	3.77 (0.17)
Antisocial	204.15 (14.13)	.925 (.013)	.954 (.008)	.060 (.005)	4.16 (0.17)
Anxiety	172.21 (13.53)	.942 (.017)	.966 (.010)	.044 (.006)	3.70 (0.17)
Aberrant thoughts	162.59 (11.16)	.965 (.010)	.979 (.006)	.040 (.006)	3.61 (0.16)
Mood	148.29 (10.14)	.974 (.010)	.985 (.006)	.031 (.007)	3.38 (0.15)

[a] all slopes linear except antisocial (all other quadratic means p>0.10)
[b] means and standard deviations for the 54 imputed data sets
*p<.05; **p<.005, ***p<=.001

Figure 3B.2: Estimated general and specific mental health bifactor scores, ages 14-27, adjusting for cohort effects using accelerated growth models on imputed data. Sex adjusted.



3B.4 Discussion

The present study longitudinally validates general and specific factors from a bifactor model of mental health in adolescents and young adults using a comprehensive suite of external variables. Further, this study clarifies phenotypic stability and heterotypic continuity which has been previously modelled only in younger cohorts (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018; Olino *et al.*, 2018), and is the first to elucidate developmental trajectories of the latent factors.

3B.4.1 Longitudinal construct validity: Analyses revealed longitudinal associations with relevant underpinnings of *p* largely in accordance with prior literature (Table 3B.1). Negative emotionality, impulsivity, lower prosociality, poor parenting, trauma history, stressful life events, fewer friendships, being female and older all predicted *p* two years later. The present study adds to the constellation of factors related to *p* by being the first to assess stressful life events and parenting styles (positive, inconsistent discipline, and poor supervision) alongside such a bifactor model. As these measures have been related to internalizing and externalizing symptoms (Table 3B.1), it is unsurprising to see all related to *p* two years after measurement.

Deprivation has previously been shown to have a small cross-sectional association with *p* (Patalay *et al.*, 2015) which was negligible two years later in the present study. Indeed, examination of cross-sectional associations in the present study reveal a comparable significant effect to the effect Patalay et al reported (standardized estimate=.09, p=.001). One possible reason for this attenuation could be changes in the social environment over time. As deprivation is related to social environment, and both are related to mental health, a positive change in this environment may attenuate the effects of deprivation on *p* over time so that direct effects are no longer seen (Tracy *et al.*, 2008). Regardless, this null prospective association underscores the importance of longitudinal associations when attempting to elucidate predictors of *p* and specific factors.

Of all the personality measures, negative emotionality most strongly predicted subsequent *p*, supporting the contention that *p* represents this affective state (Caspi and Moffitt, 2018). This finding has been demonstrated cross-sectionally in a sample of children and adolescents using the same personality measure as used presently, and was consistent whether personality was reported by parents or the child (Tackett *et al.*, 2013). Using the same sample but predicting *p* approximately ten years later in young adulthood, only parent-reported negative emotionality predicted *p* (Class *et al.*, 2019). However, a large body of literature using comparable measures also supports this relationship. Negative emotionality has been associated with *p* in samples of children (Olino *et al.*, 2014; Neumann *et al.*, 2016; Hankin *et al.*, 2017) and adolescents (Hankin *et al.*, 2017), and longitudinally predictive of *p* in adolescents (Castellanos-Ryan *et al.*, 2016) and adults (Caspi *et al.*, 2014). The present study included impulsivity in the suite of personality measures, which proved to be the next most predictive aspect of personality – a finding seen in children (Olino *et al.*, 2014) and adolescents (Hankin *et al.*, 2017) when both negative emotionality and impulsivity (or the inversely related effortful control) have been assessed. Impulsivity has been associated with *p* in children

(Olino *et al.*, 2014; Neumann *et al.*, 2016; Hankin *et al.*, 2017) and adolescents (Carragher *et al.*, 2016; Hankin *et al.*, 2017) and predictive of *p* two years later in adolescents (Castellanos-Ryan *et al.*, 2016). Less extensively studied, prosociality was predictive of *p* to a small degree as found in young people cross-sectionally (Tackett *et al.*, 2013; Patalay *et al.*, 2015) and longitudinally (Class *et al.*, 2019). Finally, daring disposition was not associated with *p* two years later, as anticipated based on cross-sectional (Tackett *et al.*, 2013), and follow-up data (Class *et al.*, 2019), although in contrast to cross-sectional findings with the related sensation-seeking (Carragher *et al.*, 2016). Despite some differences in models, informants, and follow-up periods, the literature is largely supportive of the present study's findings of personality measures with *p.*

This is the first study to have a multivariate model of cross-domain measures predict a bifactor model of psychopathology. The multivariate model consisted of background, social environment, and personality variables, which predicted 26% of the variance in *p*. (Trauma was not able to be included as a predictor in this model due to being assessed only in a subsample of participants). The only other study which has reported R^2 based on multivariate predictors of *p* found that personality measures predicted 14% of the variance in *p*, and cognitive measures, just 6% (Castellanos-Ryan *et al.*, 2016). The present findings help clarify the relative importance of variables from separate domains, where personality and social environment emerge equivalent in their effects. Negative emotionality and friendships had the greatest effect at predicting subsequent *p*, followed by impulsivity and stressful life events being the next most predictive. (Relationship with age and sex to be discussed with accelerated growth model findings.)

***Aberrant thoughts specific factor*:** The present study clarifies and extends the role a disordered thoughts specific factor in models of psychopathology. To date, this specific factor has emerged in only three bifactor models of psychopathology (Carragher *et al.*, 2016; Afzali *et al.*, 2018; Schaefer *et al.*, 2018). As in the present study, two used item-level data to model thought disorder (Carragher *et al.*, 2016; Afzali *et al.*, 2018). This specific factor did not emerge when symptom totals by diagnoses were modelled instead (schizophrenia and/or obsessive-compulsive (Caspi *et al.*, 2014; Castellanos-Ryan *et al.*, 2016), or thought disorder subscales (Tackett *et al.*, 2013; Laceulle, Vollebergh and Ormel, 2015; Neumann *et al.*, 2016)). The implication is that more detailed item-level data is required in order to model this dimension. The only study breaking with this pattern used a total score from an extensive interview of thought disorder assessing delusions, halluncinations, unusual thoughts and feelings (Schaefer *et al.*, 2018). Such a measure could have had greater sensitivity than other studies where subscales or diagnoses were modelled.

The present study is the first to relate a disordered thought specific factor to IQ. IQ negatively predicted aberrant thoughts specific factor two years later, consistent with findings that IQ is protective of subsequent psychotic symptoms in children (Arseneault *et al.*, 2011). What is particularly striking from the present findings is that IQ was more predictive of subsequent aberrant thoughts specific factor than *p*, to a comparable effect size reported with adult IQ and *p* in Caspi et

al's paper (Caspi *et al.*, 2014), where disordered thoughts based on symptom totals of schizophrenia and obsessive-compulsive disorders were subsumed into *p*. These findings indicate that, when modelled with enough specificity to emerge distinct from *p*, a disordered thoughts specific factor is indeed meaningful.

Further validation of the aberrant thoughts specific factor is supported by association with personality variables. A disordered thought specific factor has been cross-sectionally related to impulsivity and sensation-seeking (Carragher *et al.*, 2016). While impulsivity still predicted this specific factor two years later in the present study, daring disposition - characterized in part by sensation seeking (Lahey and Waldman, 2003) - did not. However, a small cross-sectional association with daring disposition in the present study (standardized estimate=.048, p=.016) supports the positive contemporaneous association with sensation-seeking reported by Carragher et al. Thus, this association does not appear to confer longer-term risk of disordered thoughts. The present study adds to Carragher et al's findings by also assessing negative emotionality and prosociality. The former was related to subsequent aberrant thoughts specific factor, consistent with longitudinal findings in young people when *p* was not extracted from psychotic symptoms (Goodwin, Fergusson and Horwood, 2003). The small negative association with prosociality and subsequent aberrant thoughts specific factor is not supported by previous longitudinal null findings from childhood to early adolescence ((Sullivan *et al.*, 2013) *p* was not extracted). Nonetheless, the present findings may indicate that prosociality in older adolescents and emerging adults is protective of disordered thoughts. Future studies are needed to confirm this finding.

Social environment was also related to subsequent aberrant thoughts specific factor in the present bifactor model, as has not been previously done. Of the parenting variables, inconsistent discipline and physical punishment (to a small degree) were related to subsequent aberrant thoughts. While the latter association was anticipated from meta-analytic findings of psychotic symptoms and diagnoses (Varese *et al.*, 2012), there are no studies which relate inconsistent discipline with the psychosis spectrum symptoms or diagnoses. However, this parental inconsistency may be related to other forms of parental dysfunction such as contradictory and fragmented parental communication, which is highly prevalent in the parents of psychotic children across diagnoses (De Sousa *et al.*, 2014). The present null findings of the aberrant thoughts specific factor with the trauma measure and stressful life events are most likely due to these relationships being subsumed by *p*, as studies of thought disorder have not assessed these relationships independent of *p* (Shakoor *et al.*, 2016; Schaefer *et al.*, 2018). Finally, supportive friendships but not positive parenting were protective against subsequent aberrant thoughts specific factor (note: there were no cross-sectional associations with positive parenting, p=.71). Both have been negatively cross-sectionally associated with psychotic experiences (to a similar degree) in adolescents in a measure which did not extract *p* (Crush *et al.*, 2019). Compared with friendships, the parenting environment of children with thought disorder is more confounded by features which

predict thought disorder, such as deviant parental communication ((De Sousa *et al.*, 2014) and note small-to-moderate correlations of the present parenting scales). Thus, any protective effects of positive parenting on the thought disorder specific factor may be drowned out by these features. This heterogeneity may be particularly evident when thought disorder is independent of *p*, as in the aberrant thoughts specific factor.

The present study also adds to the literature by being the first to relate a thought disorder specific factor to background variables. Deprivation was not related to increased aberrant thoughts as Kirkbride et al (Kirkbride *et al.*, 2014) found using the English Index of Multiple Deprivation, the measure employed in the present study. However, Kirkbride's findings were cross-sectional. As was the case with deprivation and *p,* the present study revealed a small significant cross-sectional association of deprivation and the aberrant thoughts specific factor (standardized estimate=.061, p=.025) which attenuated with time. Being of a non-white ethnic minority (22% of the sample) was related to subsequent increased aberrant thoughts. This was anticipated based on findings that non-white ethnic minorities are at elevated psychosis risk compared to those of white ethnicity (Kirkbride *et al.*, 2017). Minority ethnicity was the key factor to emerge from the multivariate model as predictive of the aberrant thoughts specific factor and may represent exposure to greater social disadvantage (Sharpley *et al.*, 2001). These findings, along with the prospective association with IQ (in a subsample so not able to be included in the multivariate model) increase confidence that this specific factor does indeed index disordered thought.

Self-confidence specific factor: this is the first bifactor study to externally validate such a factor and only the second to have such a factor emerge (Murray, Eisner and Ribeaud, 2016). This factor was predicted by external variables two years prior largely as anticipated based on cross-sectional associations with measures where a general factor was not extracted (Table 3B.1). One measure was a higher-order factor of subjective well-being comprising items of self-esteem and a positive attitude (Lampropoulou, 2018), similar to the well-being and self-esteem items which load on to our self-confidence specific factor. The consistency of this specific factor's associations with measures which have not extracted *p* suggests that self-confidence is not purely the inverse of mental illness but a unique and important construct which should be included in subsequent models of *p.* In the present study, the only departure from theorized expectations was that negative emotionality did not predict subsequent decreased levels of the self-confidence specific factor. However, cross sectionally this relationship was apparent (standardize estimate=-.076, p<.001). Thus, it appears that the effects of negative emotionality on the self-confidence specific factor attenuate with time. The multivariate model revealed prosocial disposition to most strongly predict this specific factor.

***Worry and mood specific factors*:** In addition to the present study's baseline findings (St Clair *et al.*, 2017) only a handful of bifactor models reported best fit with anxiety and mood as separate factors (Lahey *et al.*, 2012; Martel *et al.*, 2016; Greene and Eaton, 2017; Constantinou *et*

al., 2019; Watts, Poore and Waldman, 2019), and a further study where mood further bifurcated into the specific factors of depressive cognitions and melancholic features (Aitken *et al.*, 2020)). It is not yet clear whether the internalizing-externalizing structure is clearly superior to the worry-mood-externalizing structure of psychopathology and thus considering both models has been advocated (Kim and Eaton, 2015). However, of all the remaining 26 studies of the bifactor model of psychopathology which assessed both depression and anxiety (see section 3.1), only three allowed for this possibility – two showed better fit with internalizing than when anxiety and mood specific factors were separate (Lahey *et al.*, 2018b; Gluschkoff, Jokela and Rosenström, 2019) and the other tested exploratory factor models and these factors did not emerge separately (Murray, Eisner and Ribeaud, 2016). In these studies the internalizing domain was only indicated by 7-9 items, which may have limited the ability to observe distinct specific factors.). Of the prior models which extracted anxiety and mood separately, only Lahey et al externally validated these specific factors, but not using personality or environmental measures (other than childhood trauma (Lahey *et al.*, 2012)) as done in the present study. In both the present study and Lahey et al, differential validation patterns emerged for these separate internalizing specific factors, supporting the contention that these are indeed separate constructs. These validation patterns will be expanded on below for the present study.

In the social environment, both supportive friendships and positive parenting yielded protective effects on subsequent mood specific factor but were related to increased levels of the anxiety specific factor. This relationship with depressive symptoms has been well demonstrated (eg. (Yap *et al.*, 2014) systematic review and meta-analysis) and could represent social buffering of stressful events (Cohen and Wills, 1985). The increased anxiety related to supportive friendships and positive parenting could be explained by young people's internalization of the pressures to conform with normative standards and modelled behaviours (Biddle, Bank and Marlin, 1980) of individuals the young people care about. Failure to align with those standards may be anxiety-provoking, even more so if such standards are in conflict with the young person's identity explorations, which are important in both adolescence and emerging adulthood (Arnett, 2000). This contention is supported by the present findings that poor supervision and inconsistent discipline were related to decreased subsequent anxiety specific factor. Parents exhibiting these parenting styles are less likely to establish or enforce standards of behaviour, or provide consistent standards. As such, these approaches to parenting may be less anxiety-provoking for young people. However, parental supervision declines as children age (Frick, Christian and Wootton, 1999). Therefore, this subscale's negative prospective association with the anxiety specific factor could represent healthy independence and self-reliance in the young person. Indeed, over-involved "helicopter" parenting has been related to increased anxiety in young people (Yap *et al.*, 2014; Luebbe *et al.*, 2018). Future work should more closely assess these various hypotheses. Finally, stressful life events were not related with either mood or anxiety specific factors in the present study. While this differed

from longitudinal positive associations with internalizing symptoms (Amone-P'Olak *et al.*, 2009), such symptoms were not independent of *p* as in the present study. In the present study, life events only predicted *p.*

The differential association of mood and anxiety specific factors was also present with some of the personality measures. Prior work showed no concurrent and prospective association of prosocial personality with internalizing (Tackett *et al.*, 2013; Class *et al.*, 2019), although prosocial behaviour was positively associated with internalizing (Patalay *et al.*, 2015). In the present study, prosociality predicted increased worry specific factor two years later, but protected against the mood specific factor. Prosocial behaviours include being helpful and sensitive to other people's pain (Lahey *et al.*, 2008). Such responsivity to people's needs can result in anxiety for the well-being of others (Hay and Pawlby, 2003), which could explain prosociality's positive prospective association with the worry specific factor. Prosociality being associated with subsequent improved mood is consistent with findings that children low in prosociality exhibit increasing levels of depressive symptoms with age (Nantel-Vivier *et al.*, 2014). This is an important finding as such a disposition may seek to protect against subsequent mood problems. Impulsivity was negatively related with subsequent anxiety specific factor, but positively with the mood specific factor. This bifurcation explains the predominantly null findings reported in bifactor models when internalizing was modelled as one factor ((Olino *et al.*, 2014; Carragher *et al.*, 2016; Castellanos-Ryan *et al.*, 2016; Hankin *et al.*, 2017); see Table 3B.1 for alternative findings). Prior findings on the relationship of negative emotionality and daring disposition with an internalizing specific factor are more mixed (see Table 3B.1). The present findings support those studies reporting no association with the former (Tackett *et al.*, 2013; Olino *et al.*, 2014; Neumann *et al.*, 2016), and a negative association with the latter (Tackett *et al.*, 2013; Carragher *et al.*, 2016). However, given prior varied findings and that this is the first study to assess the relationship of personality with separate mood and anxiety specific factors, further study is warranted.

Associations of anxiety and mood specific factors with childhood trauma are generally consistent with the null findings reported previously in bifactor models with these factors (Lahey *et al.*, 2012) or a common internalizing factor (Caspi *et al.*, 2014). The only exception was the small association found with sexual abuse and subsequent worry and mood specific factors. While the positive association with worry could be anticipated based on similar findings with internalizing symptoms in a correlated-factors model (Schaefer *et al.*, 2018), the directionality of the mood findings are counter-intuitive and may be spurious due to a very low endorsement of sexual abuse. Work in further studies, particularly those which focus on traumatized youth, may help clarify these findings.

Finally, several protective factors emerged from the multivariate model (which did not include the trauma subsample): for the mood specific factor these were friendships, positive parenting, and prosociality, and for the anxiety specific factor this was principally a daring

disposition. Such a differential patterning of the key predictive measures underscores the importance of the mood and anxiety specific factors being distinct.

***Antisocial specific factor*:** Longitudinal associations with this factor were largely as anticipated (Table 3B.1). Many validation measures predicted the antisocial specific factor in the opposite direction to anxiety specific factor. This understandable, given that a worry specific factor has been negatively related to behaviour disorders three years later (Brodbeck *et al.*, 2014). Such a clear pattern was not seen in the only other study which externally validated a distinct anxiety specific factor alongside an antisocial specific factor (Lahey *et al.*, 2012). However, Lahey et al did not measure personality and parenting, where this pattern emerged in the present study (in addition to age and sex).

Of the social environment measures, friendships and stressful life events deviated from expectation in that they did not predict the antisocial specific factor. The later relationship was hypothesized based on contemporaneous data from a non-hierarchical model (Amone-P'Olak *et al.*, 2009), indicating that stressful life events are not predictive of antisocial behaviour which is independent of *p.* The friendships relationship was hypothesized based on findings from a bifactor model but with peer problems as the external validator (Patalay *et al.*, 2015). While related to friendships, peer problems are a somewhat different construct which may be more aligned to antisocial behaviour than friendships are. Indeed, in a sample of adolescents using the same peer and conduct problems subscales as Patalay et al ((Goodyer *et al.*, 2010); similar age to (Patalay *et al.*, 2015)), peer problems were related to conduct problems ($r=.24$, $p<.001$, $n=1,120$) but not to the present study's friendships measure ($r=-.05$, $p=.11$, $n=1,076$; unpublished data from (Goodyer *et al.*, 2010)). Thus, it is unsurprising no association was found with friendships and the antisocial specific factor in the present study. This finding suggests that friendships do not protect against antisocial behaviour which is independent of *p.* Such a null finding is consistent with observations regarding conduct disorder (for example), where peers may reject or acquiesce to antisocial behaviours, and affiliation with like peers consolidates behaviour (Burke, Loeber and Birmaher, 2002).

The present study's finding of physical punishment (the parenting subscale) predicting subsequent antisocial specific factor was supported by contemporaneous findings of physical abuse predicting an externalizing specific factor from a bifactor model (Lahey *et al.*, 2012). However, the present study's measure of physical abuse (trauma subscale) did not demonstrate this association. Certainly, the smaller sample size in the subsample who completed the trauma scales could contribute to this lack of association. This difference could also be due to a lower endorsement of the abuse questions due to their more extreme nature (eg. abuse item: "I got hit so hard by someone in my family that I had to see a doctor or go to the hospital" versus parenting item: "Your parents slap you when you have done something wrong"). This difference in association highlights the importance of assessing a broad range of behaviours to effectively measure a construct, particularly when a construct is of a sensitive nature.

Of the personality measures, the present findings are congruent with the literature's clarity that the externalizing specific factor is negatively associated with prosociality (Tackett *et al.*, 2013; Patalay *et al.*, 2015; Class *et al.*, 2019) and positively associated with a daring disposition (Tackett *et al.*, 2013; Carragher *et al.*, 2016; Class *et al.*, 2019) in young people both cross-sectionally and longitudinally. The literature is least clear with respect to this specific factor's association with negative emotionality (see Table 3B.1), yet the present lack of association concurs with the majority of studies (Caspi *et al.*, 2014; Olino *et al.*, 2014; Castellanos-Ryan *et al.*, 2016; Neumann *et al.*, 2016). One longitudinal study is in disagreement with our findings of impulsivity predicting the externalizing specific factor (Castellanos-Ryan *et al.*, 2016), whereas all cross-sectional findings concur with the present ones (Olino *et al.*, 2014; Carragher *et al.*, 2016; Neumann *et al.*, 2016; Hankin *et al.*, 2017).

Of the background, social environment, and personality variables included in the multivariate model, a daring disposition was most predictive of this specific factor, followed by prosociality's protection against subsequent antisocial specific factor. Interestingly, the univariate association with impulsivity did not survive in the multivariate model; this personality measure only predicted *p* (implications discussed in conclusions section).

3B.4.2 Predictive validity: Prospective associations of baseline factors on harmful behaviours and BMI two years later largely confirmed expectations (Table 3B.1). Of all the harmful behaviours, *p* most strongly predicted NSSI. This was consistent with Lahey et al's prospective finding in adults (Lahey *et al.*, 2012) using comparable harmful behaviours as the present study. NSSI was not related to the antisocial specific factor, consistent with a bifactor study of children and adolescents using a similar timeframe for NSSI reporting to the present study (past year for the present study, 6 months for (Deutz *et al.*, 2016)). Lahey et al found a positive association with this specific factor, but measured lifetime NSSI in adults (Lahey *et al.*, 2012); thus, reported rates of NSSI would be higher, which may have contributed to the presence of this association. The present study's bifurcation of the internalizing specific factor into anxiety and mood revealed that the latter predicted heightened levels of NSSI two years later to a small degree. A non-significant association of NSSI with the anxiety specific factor in the opposite direction helps explain the null effect with internalizing that other studies report (Lahey *et al.*, 2012; Deutz *et al.*, 2016). The present findings highlight that it is not only general distress (*p)* but also low mood independent of distress which predicts NSSI. This is consistent with the observation that adolescents engage in NSSI to regulate negative emotions (Nock and Prinstein, 2004), and helps support the interpretation of the distress-independent mood factor.

P also predicted use of cannabis and other illicit substances as anticipated based on prior longitudinal studies (Lahey *et al.*, 2012; Pettersson *et al.*, 2018; Laceulle *et al.*, 2019). *P* was not related to alcohol use, as previously found in adolescents and emerging adults (Laceulle *et al.*, 2019). This is in contrast to studies using the more severe alcohol abuse/dependency diagnoses,

which were predicted by *p* (Lahey *et al.*, 2012; Pettersson *et al.*, 2018). As excessive alcohol use is normative in adolescence and emerging adulthood (Arnett, 2000) - indeed, peaking in emerging adulthood (Brodbeck *et al.*, 2013) - this behaviour is more likely to be independent of *p* than the more severe alcohol abuse/dependency diagnoses. Nonetheless, it is of great public health importance to ascertain how various aspects of psychopathology predict alcohol use more broadly, given this behaviour's ubiquity and the problems with excessive use subthreshold to a diagnosis (Saunders and Lee, 2000). The more frequently used substances (alcohol, cannabis) were more strongly predicted by the antisocial specific factor than *p,* consistent with prior longitudinal studies of *p* in adults and adolescents (Lahey *et al.*, 2012; Laceulle *et al.*, 2019) and supporting the importance of this specific factor in predicting risk behaviour. Interestingly, when measured in childhood, *p* was most predictive of substance use/abuse diagnoses (Pettersson *et al.*, 2018). While this may point to a developmental difference or be related to parent- instead of self-reported symptoms being used to generate *p,* the outcomes from Pettersson et al were based on diagnoses from the few who made contact with specialist psychiatric services. In contrast, outcomes from the other three studies (the present one, (Lahey *et al.*, 2012; Laceulle *et al.*, 2019)) were based on self-reported behaviour or research interview diagnoses, which all participants undertook. It is therefore unsurprising that *p* was more related to the severer cases used in Pettersson et al. While the use of external data is commendable and both types of outcomes would be ideal to assess, I contend it is most relevant to use broader outcomes which pertain to the whole population to best understand the implications of a general population model of psychopathology. Finally, *p* predicted increased BMI two years later, anticipated based on prior longitudinal associations in bifactor models (Lahey *et al.*, 2012; Laceulle *et al.*, 2019), supported by the association of BMI with common mental disorders (McCrea, Berger and King, 2012) and the genetic correlation of BMI with multiple mental disorders and symptoms (Anttila *et al.*, 2018). The antisocial specific factor was not related to decreased BMI as Lahey et al reported in adults (Lahey *et al.*, 2012) but was in line with findings by Laceulle et al in adolescents and emerging adults (Laceulle *et al.*, 2019). Lahey's externalizing factor was heavily based on substance dependence disorders (4/5 of the indicators of externalizing), whereas in the present study and Laceulle et al, this factor was represented by aggressive and delinquent behaviour. The observed inverse relationship between BMI and substance use (Sansone and Sansone, 2013) could explain why this relationship was found in Lahey et al.

3B.4.3 Phenotypical Stability: Neither dynamic mutualism nor *p*-differentiation theories (Murray, Eisner and Ribeaud, 2016) were supported by the present findings of homotypic continuity in the general and specific factors. Moderate-to-large autocorrelations between all factors were stable from HQP1 to HQP2, and HQP2 to HQP3. Unsurprisingly, studies of bifactor models with eight or nine repeated assessments over childhood and adolescence yielded more fluctuant autocorrelations (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018). Consistent with studies

of children, adolescents, and adults repeatedly assessed after one to three years (Castellanos-Ryan *et al.*, 2016; Greene and Eaton, 2017; Snyder, Young and Hankin, 2017; McElroy *et al.*, 2018), *p* in the present study showed high levels of homotypic continuity. These reported associations (.64-.86) were understandably higher than those found after 10-12 years (.32) using different indicators and different informants (Class *et al.*, 2019). The magnitudes of the autocorrelations for each specific factor were also consistent with prior work which modelled symptoms over a comparable timescale to the present study: a moderate effect for the present self-confidence specific factor and a prosociality specific factor (Murray, Eisner and Ribeaud, 2016), moderately-high effects in the present antisocial specific factor and aggression and externalizing specific factors (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018), as well as the present mood and anxiety specific factors (combined) and internalizing specific factors (Murray, Eisner and Ribeaud, 2016; McElroy *et al.*, 2018). Such concordances with younger samples indicates that the stability of these specific factors is not greater for young adults, who are still undergoing brain development (Giedd *et al.*, 2015). However, much higher rates of stability were found for externalizing and internalizing/fear specific factors when item-level data was based on diagnostic criteria instead of symptoms (children (Olino *et al.*, 2018); adolescents (Castellanos-Ryan *et al.*, 2016); adults (Greene and Eaton, 2017)). Multiple symptoms loading onto specific factors captures greater nuance than diagnostic categories alone, and therefore allows greater possibility for change. As with *p*, the specific factors had much lower autocorrelations over approximately a decade (Class *et al.*, 2019). Finally, the present study is the first to present a stability coefficient for an aberrant thoughts specific factor. This factor was the most stable of all specific factors, and nearly as stable as the general factor.

The heterotypic continuity findings in the present study give limited support to both dynamic mutualism and *p*-differentiation theories. Principally, reciprocal associations of *p* with aberrant thoughts were found across all waves of data, implying that *p* strength may increase and yet become more specific at the same time. McElroy et al (McElroy *et al.*, 2018) suggested that these competing processes may cancel each other out, explaining the observed consistent variance explained by *p* and specific factors over time. The reciprocal associations of *p* with aberrant thoughts help validate this specific factor. Psychotic phenomena may be a marker of severity in common mental distress (Stochl *et al.*, 2015), and as such, this relationship with *p* is unsurprising. This notion is further supported by severity thresholds in the present model, which indicated that psychotic-like experience items loaded very highly on the distress continuum (St Clair *et al.*, 2017). It is important to note that this is only the second bifactor study to report any reciprocal associations between mental health factors, principally because most studies only generate a bifactor model at one or two waves of data. Heterotypic continuity may be observed with two waves of data, but three waves of data or more are required in order to observe reciprocal associations. Future studies should aim to further assess this phenomenon, as it gives bearing on how individuals move in and out of symptom profiles over the lifecourse.

Heterotypic continuity was also demonstrated by reciprocal inverse associations between self-confidence and mood specific factors. This was unsurprising due to the conceptual similarity of the original measures (mood (Costello and Angold, 1988) and well-being (Tennant *et al.*, 2007)). Despite being stripped of their commonalities by extraction of the general factor, these specific factors were still inversely related to a small degree over time. These reciprocal associations could also be partially due to methodology: eight cross-loading items were allowed if both items loaded >.20 (St Clair *et al.*, 2017). These cross-loading items represent 62% (8/13) of the items in the self-confidence specific factor and 27% (8/30) of the items in the mood specific factor. Being independent of *p*, the reciprocal associations evidenced between these factors does not argue for or against dynamic mutualism or *p*-differentiation, but instead suggests that self-confidence may be an important area for intervention for young people who exhibit mood problems. Finally, the absence of reciprocal associations between the mood and anxiety specific factors is notable, given that the majority of bifactor models construe these as one internalizing factor. Such a lack of association supports the independence of these constructs, once devoid of *p.*

Regardless, the heterotypic continuity exhibited in the present study is supported by findings from prior bifactor models in young people. Studies in adolescents and/or children (Castellanos-Ryan *et al.*, 2016; McElroy *et al.*, 2018; Olino *et al.*, 2018) also found small associations between factors, albeit with some differences to the present study (commonalities were found in the present study's mood and anxiety specific factors predicting the general factor and antisocial specific factor respectively from HQP1-2, just as the internalizing specific factor predicted both the general factor and externalizing specific factor across some of the waves in McElroy et al's study (McElroy *et al.*, 2018)). As each study consisted of different measures representing different facets of the mental landscape, it is more important to note the *presence* of heterotypic continuity across studies. This suggests that in these age ranges, the expression of mental illness changes to a small degree (or for a small percentage of individuals) over time. Individuals' symptoms may ebb and flow in an episodic nature, irrespective of diagnosis. However, studies in adolescents (Snyder, Young and Hankin, 2017) and adults (Greene and Eaton, 2017) did not support heterotypic continuity. While statistical power may have limited the ability to discern heterotypic continuity in the adolescent sample, the lack of heterotypic continuity in adults may suggest that mental states become more discrete with age. This notion is consistent with observations that brain development peaks in the third decade of life (Giedd *et al.*, 2015), but needs further testing in studies which model both young people and adults together. Finally, heterotypic continuity was not supported in the study which followed-up adolescents after approximately a decade (Class *et al.*, 2019), although as above, this effect may have been hampered by power (both studies had small between-factor associations at trend level: β=.2, p<.1, sample size approximately 500 (Snyder, Young and Hankin, 2017; Class *et al.*, 2019)).

3B.4.4 Developmental trends from ages 14-27: This is the first epidemiological study to reveal developmental trends in mean levels of *p* and the specific factors. The two studies modelling *p* over psychotherapy intervention and follow-up (Constantinou *et al.*, 2019; Aitken *et al.*, 2020) do not pertain to the naturalistic course of *p* and are thus not discussed. Accelerated growth models in the present study revealed that *p* increased over adolescence and emerging adulthood, from ages 14 to 27. The findings from the present representative UK sample are consistent with epidemiological findings showing an increase in mental disorders during this age range. Indeed, half of all lifetime mental disorders have been shown to emerge by age 14, and three-quarters by age 24, based on interview assessed mental disorders of a representative sample in the USA (Kessler, Berglund, *et al.*, 2005). Further, prevalence of any disorder increased from 18-29 year olds to 30-44 year olds (Kessler, Berglund, *et al.*, 2005), consistent with the continued upward trend in the present study. However, it is important to note that age of onset estimates rely largely on retrospective reports from cross-sectional community surveys (Kessler *et al.*, 2007). Thus, such estimates are based on prevalence by age population as opposed to the developmental trajectories able to be estimated with the present accelerated longitudinal design. However, this being the first epidemiological study to determine latent growth in a bifactor model of psychopathology, such comparisons are an appropriate way of contextualizing these findings.

Comparing mental disorders with elements of a bifactor model is also justified by prior associations of *p* (based on self-report data) with interview-based psychiatric diagnoses in adolescents. Based on a model using depression and anxiety symptoms, *p* was associated with concurrent behavioural, substance use, and eating disorders; *p* also had concurrent and 3-year prospective associations with affective and anxiety disorders (Brodbeck *et al.*, 2014). Validation of this model's two specific factors was also demonstrated with psychiatric diagnoses: a hopelessness specific factor including many of the items from the present mood specific factor (Brodbeck *et al.*, 2011; St Clair *et al.*, 2017) predicted affective diagnoses three years later, and a worry specific factor predicted anxiety diagnoses three years later (Brodbeck *et al.*, 2014). These findings support the appropriateness of comparing the present model's specific factors to mental disorders. If specific factors independent of *p* exhibit a similar developmental pattern to the associated mental disorder, this supports the interpretation that these residual factors still contain meaningful aspects of the comparable disorder.

The antisocial specific factor, based on antisocial behaviours (St Clair *et al.*, 2017), exhibited a quadratic decline over the 14-27 years of age studied. The median age of onset for conduct disorder is at ages 9-14 (Kessler *et al.*, 2007). Thus, most of these initial symptoms would have already been manifest by the youngest age in the present study, explaining no observed increase. A quadratic decline in the antisocial specific factor is consistent with the observed quadratic curve in antisocial and criminal activity which has been coined the age-crime curve (Loeber and Farrington,

2014). Such activity peaks in the late teens (ages 15-19) and subsequently declines over emerging adulthood, as is observed for the antisocial specific factor during this age range.

The worry specific factor, reflecting generalized worry ((St Clair *et al.*, 2017) appendix), showed a significant linear increase over the ages of 14-27 in the present study. The symptoms which load onto this specific factor are common across many anxiety disorders (American Psychiatric Association, 1994). While the median age of onset of any anxiety disorder is lower than conduct disorder, the age of onset is broader, and is diverse across types of anxiety disorder (Kessler, Berglund, *et al.*, 2005). After most phobias and separation anxiety onset in childhood or early adolescence, agoraphobia, PTSD and panic disorder all have a median onset in the early twenties, with 25% of cases emerging by the mid-teens. This is followed by generalized anxiety disorder's median onset at age 31, with 25% of these cases emerging by age 20 (Kessler, Berglund, *et al.*, 2005). Thus, the increase seen in the anxiety specific factor from the ages of 14-27 is understandable. Also, it is important to note that the present model is based on symptoms as opposed to diagnoses. Thus, non-pathological age-appropriate worry may be increasing for emerging adults - as they transition into greater independence they also enter into a period of heightened instability, particularly with respect to work, romantic relationships, and identity explorations (Arnett, 2000; Arnett, Žukauskiene and Sugimura, 2014). This can contribute to the heightened levels of anxiety reported by emerging adults, with 56% agreeing with the statement "I often feel anxious" (Arnett and Schwab, 2012). However, it is striking that this increase in the worry specific factor is still apparent even though this factor is independent of *p,* which also rises during this age range.

In contrast to the worry specific factor, the mood specific factor did not show a significant increase from ages 14-27. This stability matches 12-month prevalence rates found in the Dunedin longitudinal cohort study, which range from 16.5% to 17.3% measured at 18, 21, and 26 years (Moffitt *et al.*, 2010). Cross-sectional data from the separate adolescent and adult National Comorbidity Surveys in the USA also shows reasonably stable rates of mood disorder, at 10% for adolescents aged 13-17 (Kessler *et al.*, 2012), and 13% in emerging adults aged 18-29 (https://www.hcp.med.harvard.edu/ncs/ftpdir/NCS-R_12-month_Prevalence_Estimates.pdf). Growth modelling of longitudinal depressive symptoms from adolescence to emerging adulthood (also an American sample) revealed that these symptoms remained stable through late adolescence and declined into emerging adulthood, although there was considerable variability within individuals over time (ages 16-23 (Finan, Ohannessian and Gordon, 2018)). Taken together, these studies support the lack of change seen in the present mood specific factor from 14-27 years of age.

The self-confidence specific factor also did not significantly change over the age range studied, consistent with the finding that age was unrelated to well-being in 13-16-year-old adolescents, using the WEMWBS (Warwick-Edinburgh Mental Well-Being Scale), the measure from which the self-confidence items came (Clarke *et al.*, 2011). A declining non-significant trend was

observed in the present study, which was also found in the WEMWBS for young people and adults ages 16-34 (Tennant *et al.*, 2007).

The aberrant thoughts specific factor exhibited a significant linear decline from ages 14-27, a finding which was not anticipated. Based on obsessional/compulsive and psychotic-like experience items, this specific factor has elements of disordered thought which cut across the clinical typologies of obsessive-compulsive disorder, schizophrenia, all other psychoses, and paranoid states. However, age of onset data indicates that these disorders increase during adolescence and emerging adulthood. The median age of onset of obsessive-compulsive disorder has been shown to be 19 (interquartile range 14-30; (Kessler, Berglund, *et al.*, 2005)), non-affective psychosis emerges in the late teens to early 20s (Kessler *et al.*, 2007), and schizophrenia increases rapidly until the mid-20's (Häfner *et al.*, 1994). The majority of increases seen in these disorders could be due to *p*, which could explain why no such increase is apparent in the aberrant thoughts specific factor. Also of note is finding that 75–90% of developmental psychotic experiences appear to be transitory and disappear with time (Van Os *et al.*, 2009). Thus, the aberrant thoughts specific factor may be capturing symptoms such as delusions and hallucinations that are not associated with a disorder and so follow a typical declining course.

3B.4.5 Gender and growth curves: The relationship of gender with the age 14 intercepts of each of the factors was consistent with the association of gender reported in the overall baseline findings where the mean age was 19 years (St Clair *et al.*, 2017). In contrast to a lack of association with *p* and gender reported in adolescents ages 12 and 13 (Patalay *et al.*, 2015; Carragher *et al.*, 2016; Afzali *et al.*, 2018), the present study revealed heightened levels of *p* for females at age 14. However, a significant relationship of slope with gender indicated that *p* increased more for males than females over the twelve years the growth curve was estimated. This relatively greater increase for males could explain the lack of gender findings seen in adulthood (Caspi *et al.*, 2014).

Gender differentiated some of the specific factors at age 14, consistent with prior bifactor models in adolescents. Comparable to findings for the externalizing specific factor in adolescents (Patalay *et al.*, 2015; Carragher *et al.*, 2016; Afzali *et al.*, 2018), males in the present study exhibited higher levels of the antisocial specific factor at age 14. Gender was not related to the slope of this specific factor, demonstrating that the magnitude of these gender differences is unaltered over time. This is to be expected, as males also exhibit higher levels of the externalizing specific factor in adults (Caspi *et al.*, 2014).

The bifurcation of internalizing symptoms into mood and anxiety specific factors allowed new associations with gender to emerge which have been previously masked in bifactor models which extracted one internalizing specific factor. Gender differences reported in the internalizing specific factor in adolescents (Patalay *et al.*, 2015; Carragher *et al.*, 2016; Afzali *et al.*, 2018) were subsumed by the worry specific factor in the present study: at age 14, females were higher on this factor whilst no gender differences were apparent in the mood specific factor. Gender was related to

the slope of both these specific factors, but in the opposite direction: females exhibited greater levels of worry specific factor with time than males, but males exhibited greater levels of mood specific factor with time. Such a cross-over effect explains the continued observation of a female preponderance for the internalizing specific factor in adults (Caspi *et al.*, 2014) comparable to that observed in adolescents. These findings highlight the heterogeneity in the anxiety and mood specific factors which is masked by extracting a joint internalizing specific factor as many studies do.

There were no gender differences found in the remaining two specific factors. A lack of association of gender with aberrant thoughts specific factor at age 14 is consistent with findings from bifactor models in adolescents where a thought disorder specific factor was extracted ((Carragher *et al.*, 2016; Afzali *et al.*, 2018) - no adult studies have extracted such a factor). This finding is also supported by data from the psychosis spectrum: while psychosis in the context of schizophrenia is more common in men, psychosis in the context of mood disorders (ie: bipolar and schizoaffective diagnoses) is more common in women (Castro *et al.*, 2018). When symptoms are reported irrespective of diagnosis as in the present study, null gender effects are understandable. Gender was also unrelated to slope of aberrant thoughts specific factor. Similarly, in the self-confidence specific factor gender was unrelated to intercept or slope. While men have reportedly higher levels of well-being on the WEMWBS (Tennant *et al.*, 2007), the present self-confidence specific factor is independent of *p*. WEMWBS items loaded negatively onto *p*, which exhibited a strong female preponderance. Thus, the gender differences previously found in the WEMWBS are likely predominantly due to the presence of *p* in this scale. (It is also possible that the inclusion of the WEMWBS scale contributed to the present study's observed gender differences in *p*, which have not been apparent in other models of *p*).

Finally, it is important to note the limitations of the accelerated growth models. Whilst measurement invariance has been established from HQP1-3, due to complexity of the model and sample size of the cohorts, measurement invariance across cohorts was not able to be tested. Longitudinal data from overlapping but different age cohorts is used to estimate a common developmental trajectory across the cohorts (Duncan and Duncan, 2004); however, the estimation of the growth curves for each cohort is still limited by the number of repeated assessments. With only three repeated assessments in the present dataset, a cubic curve cannot be estimated, and a quadratic curve is only just identified with a fixed term. Future work with accelerated designs should aim to assess an additional timepoint to allow a more comprehensive modelling of possible trajectories.

3.2 Chapter Summary and Conclusions:

Herein I present a longitudinal bifactor model of mental illness and wellness, represented by self-reported thoughts, feelings, and behaviours across adolescence and emerging adulthood. All factors except the anxiety specific factor exhibited adequate to excellent construct replicability over time. Thus, these latent constructs were well-represented by their contributing items, supporting their inclusion in the model as distinct entities. In future work, greater representation of various anxiety symptoms could help bolster this specific factor. Support for gender invariance of the model was established, as well as longitudinal invariance over one and two years. The demonstrated consistency in strength of the factors over time and presence of homotypic continuity did not support the theories of dynamic mutualism or *p*-differentiation (Murray, Eisner and Ribeaud, 2016). As reciprocal associations of *p* with aberrant thoughts were found across all waves of data, this implies that *p* strength may increase and yet become more specific at the same time. This phenomenon has been observed in a younger sample, but between different specific factors (McElroy *et al.*, 2018). Future studies should aim to generate a bifactor model with three waves of data or more, to shed more light on reciprocal associations between these putative distinct symptom profiles.

The longitudinal association with external validation measures greatly strengthens support for the meaning of *p* and the specific factors in the present study. Key protective factors (supportive friendships, prosociality, positive parenting) as well as risk factors (negative emotionality, trauma symptoms, daring disposition, stressful life events, impulsivity) have been identified as particular targets for intervention with different patterning across *p* and the specific factors. Such patterning was made particularly clear in findings from the multivariate model, which consisted of background, social environment, and personality variables – the first study to have cross-domain measures predict a bifactor model of psychopathology. That a differential pattern also emerged when the bifactor model predicted subsequent risks further supports the importance of these distinct factors. On several occasions, cross-sectional associations present in the data and reported by others became non-significant when prospectively associated over two years. This underscores the importance of external validation using longitudinal data, to ascertain more sustained effects of any associations. Further work might include comparing the present associations with external variables in similar analyses with a correlated factors model (Watts, Poore and Waldman, 2019) or conversely testing how the interpretation of *p* with respect to these variables may be distorted in a unidimensional model (Reise, Moore and Haviland, 2010). These additional analyses could shed further light on the relevance of the specific factors in the present bifactor model.

A unique aspect of the present study is that it models developmental trends of *p* and the specific factors over 14 years of adolescence and emerging adulthood. This helps increase understanding of these phenomena during this crucial period of change. Consistency with age of onset data of the associated mental disorders (Kessler, Berglund, *et al.*, 2005) and the

developmental course of these behaviours and thoughts which have not extracted *p* (Tennant *et al.*, 2007; Loeber and Farrington, 2014; Finan, Ohannessian and Gordon, 2018) further supports the interpretation of these factors as meaningful constructs. Taking a similar approach in younger and older samples would help extend our understanding of change in these factors over the lifecourse.

If symptoms are exhibited in the absence of distress (*p)*, the present findings (particularly from the multivariate model) indicate different intervention targets than if distress was also present. While heightened worry and depressed mood seem more challenging from a clinical perspective to disentangle from distress, both antisocial behaviour and psychotic symptoms can occur in the absence of distress (Vracotas *et al.*, 2007; Glenn, Johnson and Raine, 2013). The present findings indicate that for this profile of antisocial behaviour the focus of intervention should be on modifying characteristics related to personality instead of the social environment. Daring disposition was a key risk factor of the antisocial specific factor, and those with a daring disposition are predisposed towards engaging in risky behaviour (Lahey *et al.*, 2008). Such interventions should therefore focus on teaching the young person to think through the implications of risky behaviour more so than teaching self-control, as impulsivity was only related to *p* in the multivariate model. If disordered thoughts are exhibited independent of distress, the present findings indicate that interventions should focus on mitigating the deleterious aspects of being an ethnic minority which are unrelated to social environment. As the aberrant thoughts specific factor also predicted subsequent *p*, clarifying these underlying mechanisms should be an urgent focus of subsequent research.

Taken together, the present findings as a whole support the utility of not only *p* but also the specific factors. In research, these validated orthogonal factors would be particularly revealing when studying different dimensions of mental illness and wellness concurrently, where correlated measures would be problematic. In a clinical setting, these factors could be used as treatment outcomes for comorbid disorders (such as depression and anxiety), if treatment was thought to address both general and specific aspects of symptoms. While treatments such as the Improving Access to Psychological Therapies (IAPT) programme in the UK include holistic outcomes, such as the ability to function in various social contexts (National Collaborating Centre for Mental Health, 2018), the present study's well-being specific factor independent of *p* could also be an important aspect to include when measuring recovery. These specific indicators could highlight relative areas of strength and weakness in an intervention which may hitherto have been masked. Currently this approach could be taken when statistically assessing outcomes at the population level, if a comprehensive suite of outcomes were assessed. However, the method to operationalize these statistically-derived factors on the individual level, particularly for the specific factors, is yet to be determined (Caspi and Moffitt, 2018). Nonetheless*,* the present study is a step in the right direction towards a greater understanding of a transdiagnostic approach to mental health.

CHAPTER FOUR

Young people's mental health treatment in the community: the relationship with transdiagnostic distress and social functioning over two years

4.1 Introduction

There is a growing acknowledgement of the need to support young people with mild to moderate mental health conditions (Department of Health and Department for Education, 2017). This includes those who have low-level needs not meeting diagnostic criteria but who would benefit from support, and those with a mental health diagnosis who may or may not meet the threshold for specialist care. Including subthreshold individuals in care is crucial from a prevention perspective, as such individuals are at risk of developing a full-fledged mental disorder (McGorry, Bates and Birchwood, 2013), with the concomitant lifelong economic and social sequelae (Goodman, Joyce and Smith, 2011). Allowing individuals to receive care irrespective of a mental health diagnosis circumvents the growing discomfort with the distinct categorization of psychiatric disorders (Caspi and Moffitt, 2018).

The UK offers services to adolescents with mild to moderate mental health conditions, from the National Health Service, schools, and local authorities, for example (Department of Health and Department for Education, 2017). As a result of low availability of therapy for adults with mild-moderate disorder, England's Individual Access to Psychological Therapies (IAPT) programme provides evidence-based therapy for adults whose symptoms would not qualify them for mainstream mental health services (Wolpert *et al.*, 2016; Clark, 2018). It is important to establish in community samples of young people how effectively this current standard-of-care reduces symptoms of mental ill-health irrespective of mental health diagnosis, and whether there is a discernable mechanism by which these treatments might act.

Research addressing the effectiveness of treatment-as-usual mental health services is limited and needs expanding. In Chapter 2 I utilized a community cohort to demonstrate that adolescent mental health service contact was related to a reduction in subsequent depression three years later (Neufeld *et al.*, 2017). Crucially, this study had a non-service using comparison group, and advanced the few comparable studies on the subject by adjusting for participants' initial likelihood to access services as well as subsequent factors related to treatment and outcome, utilizing multiple imputation to deal with missing data, and employing a clinically relevant cut-off. However, since that paper was published in 2017, there have been no further studies with a non-service using comparison group addressing the effectiveness of mental health treatment for adolescents or emerging adults in the community. Therefore, more work in different samples is needed to test the ubiquity of this treatment effect. The previous findings could be expanded in

several ways: employing an outcome which is transdiagnostic across mental disorders, and obtaining a more diverse sample across age, socio-economic status, and ethnicity. Regarding diversity, the cohort used in assessing the effectiveness of community mental health service contact in Chapter 2 (Neufeld *et al.*, 2017) was based on a more affluent and ethnically white sample than that of the rest of the UK (Goodyer *et al.*, 2010). Therefore, expanding such findings to a more socio-economically and ethnically diverse sample would demonstrate relevancy across a wider population.

Testing whether treatment-as-usual mental health services are effective or not in a measure that is relevant across multiple domains of disorder would be advantageous. A single dimension measuring the commonality among mental disorders is gaining traction as a way of more accurately representing the symptom overlap found in psychiatry (Caspi and Moffitt, 2018). This continuous common factor is also thought to quantify risk for mental disorder. The present study will thus employ a general distress factor as the primary outcome. This distress factor is the highest order overarching factor from a bifactor model of 118 items of thoughts, feelings, and behaviours which cover multiple domains of mental illness and wellness, developed cross-sectionally (St Clair *et al.*, 2017) and validated longitudinally (Chapter 3). This measure will be useful at testing the effects (positive and/or negative) of mental health treatment as usual on common symptoms of distress among these problems. The present study does not include other potential harms of treatment such as increased dependency or change in medication needs (Duggan *et al.*, 2014; Jonsson *et al.*, 2014) as these outcomes were not assessed. However, analyses were adjusted for subsequent mental health treatment and harmful behaviours, which could be indicative of a treatment-related adverse event (Jonsson *et al.*, 2014).

There is a strong argument for testing whether treatment-as-usual mental health services remain effective beyond adolescence from ages 18-25, termed emerging adulthood (Arnett, 2000). Many mental disorders emerge during this age range: half of all lifetime mental disorders have emerged by age 14 and three-quarters by age 24 (Kessler, Berglund, *et al.*, 2005). There are also many life transitions in emerging adulthood which make this period developmentally distinct from adolescence and the rest of adulthood, such as moving away from parents for the first time, undergoing training or education for a career, or finding a life partner (Arnett, 2000). The greater instability, heightened identity exploration and self-focus found in this developmental period can explain the high levels of anxiety and depression reported in emerging adults (Arnett, Žukauskiene and Sugimura, 2014). Mental health systems have not yet adapted to recognize the impact of these life transitions, and practitioners may be ill-equipped to address them (Arnett, Žukauskiene and Sugimura, 2014). The majority of Child and Adolescent Mental Health Services (CAMHS) in multiple developed countries around the world terminate at age 18 (McGorry, Bates and Birchwood, 2013; Arnett, Žukauskiene and Sugimura, 2014). This creates a disconnect in services for young people and their families, which can result in loss of treatment for needy young people (Singh *et al.*, 2010).

Often moving simultaneously away from living with parents to a new educational and social experience, emerging adults may lose not only access to the services they relied on in adolescence, but also their primary mental health advocates. It is therefore important to test whether mental health treatment-as-usual keeps sufficient pace with the changing needs of emerging adults.

While there is no direct evidence for mental health treatment-as-usual improving social functioning compared with no treatment (NICE, 2019b), there is reason to believe that this may be the case. Community mental health treatment comes from various sources, such as psychiatrists, psychologists, and school counsellors (Department of Health and Department for Education, 2017; Neufeld *et al.*, 2017). Across disorders, these practitioners utilize common techniques such as listening, encouraging, and empathising, and thus model beneficial social skills. Social skills may be an explicit focus of therapy, for example in interpersonal therapy and cognitive-behaviour therapy (CBT). A key intervention recommended in the NICE guidelines for many mental disorders in young people and adults (eg. depressive, anxiety, behaviour disorders, and psychosis (NICE, 2014b, 2014c, 2015, 2016, 2017, 2019a; Clark, 2018)), CBT is the most prevalent therapy type used in IAPT (Wolpert *et al.*, 2016; Baker, 2018; Clark, 2018). Thus, it is reasonable to posit that young people's interaction with, and perception of, their social environment may improve following treatment. Feeling understood in therapy also helps renew both the patient's capacity for social understanding and their ability to learn from social experience (Fonagy and Allison, 2014). Thus, this therapeutic relationship alone may help improve social functioning. Indeed, social functioning is a measured outcome in the IAPT programmes in England, which aim to increase the quality of mental health services for children, young people, and adults by implementing evidence-based psychological therapies (Wolpert *et al.*, 2016; Clark, 2018).

Evidence-based psychological therapies have an explicit focus on both families and peers; thus, mental health treatment is likely to improve both family functioning and friendship support. In both adolescent and adult services, family and peer support is particularly encouraged for conditions such as PTSD and psychosis (NICE, 2014c, 2016, 2018). However, for more common mental disorders, there is more emphasis on these social environments for adolescent care than adult care. Challenges with the family and peers are assessed as part of care for adolescents with anxiety, depression, and antisocial behaviour (NICE, 2013, 2014a, 2019a). However, families are more likely to be integrated into interventions for these conditions, and generally have a greater focus in treatment. In the UK, for young people with mild to moderate mental health problems in CAMHS (up to and including age 18), family issues must be addressed if they appear to have an adverse effect on the young person (NHS England, 2015b). In Children and Young People's (CYP) IAPT, the most common presenting problem was family relationship difficulties; all explicit family focused therapies combined were more prevalent than CBT (Wolpert *et al.*, 2016). In adult IAPT, family therapy is not mentioned (Baker, 2018; Clark, 2018). It is therefore more likely that treatment may be related to improved family functioning in young people who remain in CAMHS compared to

those who have transitioned to adult services. Similarly, treatment is more likely to be related to improved peer support in adolescents in CAMHS than young people in adult services. However, improvements in peer support are less likely to be seen in this age group than improvements in family functioning, given the greater focus on families than peers in CAMHS mentioned above.

Both families and friendships appear to play an important role in young people's mental health across mental disorders. While friendships and romantic relationships become an increasing focus during adolescence and emerging adulthood, relationships with families deepen (Larson and Richards, 1991; Larson *et al.*, 1996; Arnett, 2000). Adolescents and adults across a broad spectrum of mental disorders have poorer perceived family functioning than controls, irrespective of disorder (Friedmann *et al.*, 1997). Sixty percent of CYP ages 5-19 with any type of mental disorder lived in households with unhealthy family functioning (Sadler *et al.*, 2018). In young adolescents, peer problems have been positively correlated with a general factor of self-reported externalizing and internalizing symptoms (Patalay *et al.*, 2015). Poorer perceived family functioning at age 14 has been related to new cases of non-suicidal self-injury three years later (Cassels *et al.*, 2018). Post-secondary school, parental support and reciprocity has been prospectively related to emerging adults' improved emotional adjustment (Wintre and Yaffe, 2000; Levitt, Silver and Santos, 2007).

When compared together, the evidence is not clear as to whether family or friendship support is more strongly connected to young people's mental health. In adolescents aged 12-19, relationships with parents were more closely related to young people's wellbeing than relationships with friends (Greenberg, Siegel and Leitch, 1983). Similar findings have emerged in longitudinal studies of emerging adults. A positive relationship with parents in adolescence predicted reduced depressive symptoms in later adolescence and emerging adulthood, with no such relationship found for close friends (Finan, Ohannessian and Gordon, 2018). Similarly, first major depressive episodes by age 20 were predicted by poor family relationships but not poor peer relationships two years prior (Eberhart and Hammen, 2006). This could be largely because family relationships tend to be long-lasting (Bengtson, 2001) but peer relationships are more transitory during this age range (Poulin and Chan, 2010). However, an adolescent study has shown equivalent protective effects of supportive families and friendships on subsequent depressive symptoms (van Harmelen *et al.*, 2016). Chapter 3 indicates that in adolescents and emerging adults, both positive parenting and supportive friendships are independently related to *p* two years later, although friendships have a greater effect. In the same sample as Chapter 3, perceived social support from friends but not family was related to improved subsequent resilient functioning (van Harmelen *et al.*, 2017). It therefore appears the relative importance of family and friends may vary depending on the context and studied outcome.

A handful of adolescent studies have assessed directionality in the relationship between specific indicators of mental health and social support with mixed findings. Lack of parental support in adolescence has been shown to predict subsequent aggression, but the reciprocal relationship

was not supported (Hale *et al.*, 2008). Peer relationships and behavioural problems appear to have a reciprocal relationship: prior aggression has been shown to predict the choice of deviant friends in adolescence, but prior peer rejection also predicted adolescent rule breaking (Ettekal and Ladd, 2015). For adolescents exhibiting depressive behaviour, there appears to be a subsequent decline in peer support, but not the converse relationship (van Harmelen *et al.*, 2016; Ren *et al.*, 2018). While studies have shown adolescent depressive symptoms predict decreased parental and family support (but not the reciprocal (Hale *et al.*, 2008; van Harmelen *et al.*, 2016)), reciprocal evidence has also emerged for depressive and anxiety symptoms (Rueter *et al.*, 1999; Finan, Ohannessian and Gordon, 2018). Given such equivocal findings regarding directionality, any tests of mechanisms of treatment should assess whether social support mediates treatment effects on symptoms, or whether symptoms mediate treatment effects on social support.

In the present study, I utilize a longitudinal community sample of adolescents and emerging adults aged 14-24 (Kiddle *et al.*, 2018) to determine whether current mental health treatment at baseline is associated with reduced distress symptoms two years later. As previously mentioned, distress is a general factor representing the commonality among multiple measures of mental illness and wellness (St Clair *et al.*, 2017). The sample is broadly representative of England and Wales' youth population based on census data on gender, ethnicity, foreign birth, parental education, and deprivation (Kiddle *et al.*, 2018). I further test whether perceived family functioning or friendship support mediates the effect of treatment on distress. The equivocal findings outlined above regarding directionality of effects will be addressed by testing the converse model, that symptoms may mediate the effect of treatment on either measure of social functioning. Longitudinal structural equation modelling will be employed to test treatment effects and mediation of these effects, incorporating measured and unmeasured confounding and measurement error where possible (Goldsmith *et al.*, 2018a).

Given that treatment was not randomized, propensity weighting will be performed to balance covariates between treated and untreated groups, similar to a randomized control trial (Rosenbaum and Rubin, 1983). With community effectiveness studies, which by definition cannot be randomized, propensity weighting represents the best quality evidence. There are a large percentage of young people who require mental health treatment but do not receive it (34-56% from studies reviewed in Chapter 2 (Neufeld *et al.*, 2017)). Thus, it is most relevant to utilize propensity weighting to determine the average treatment effect in a population which may need such services, not just what the effect is in those who are actually treated (Austin, 2011). In Chapter 2 I have shown the importance of establishing treatment need when estimating treatment effects via propensity weighting (Neufeld *et al.*, 2017). In Chapter 2, the presence of a mental health diagnosis was used to determine treatment need. As the present study does not have this data, analyses will therefore comprise those above the population mean on distress. Such a cut-off sets a threshold for initial impairment which serves several purposes. In addition to establishing a population who potentially

require mental health treatment in order to estimate the population-level average treatment effect, a cut-off means that improvement is more likely to be discernible and also enables the treated and untreated groups to be more comparable on covariates. As there must be sufficient overlap between these groups to estimate causal effects (Stuart, 2010), an impairment cut-off facilitates balance of the treated and untreated groups using the propensity weight. Young people with higher levels of general distress are more likely to be mentally unwell and receive treatment (St Clair *et al.*, 2017); similarly, those with more symptoms of common mental disorder are more likely to receive treatment (Mcmanus *et al.*, 2016). Using a threshold such as those above the population mean on distress is therefore likely capture the large majority of participants who have received treatment, including those who meet criteria for a mental health diagnosis and also potentially individuals with subthreshold symptoms. Those receiving treatment below the population mean on distress may represent those with no history of mental illness, or those in remission from mental illness – in such individuals, it is difficult to demonstrate further improvement, for the improvement has already happened.

I also aim to test the role of age in the effects of treatment on distress and social support measures, with the above mediation models tested separately by age group if warranted. During their nineteenth year, most young people finish compulsory education and leave their parents' home (Arnett, 2000). In the UK, the transition from CAMHS to adult services may occur from age 16 to 25, depending on the specific mental health problem and location where services are received (Singh *et al.*, 2008, 2010; McGorry, Bates and Birchwood, 2013). However, the majority make the transition shortly after their eighteenth birthday (Singh *et al.*, 2010). This is supported by IAPT data from England. From 75 services of CYP IAPT, using records of care from those aged 15-25, only 0.5% represent emerging adults aged 20-25 (Wolpert *et al.*, 2016). In adult IAPT, only 2% of referrals and treatment are in those under 18 (Baker, 2018). I will therefore create subgroups by baseline age to perform separate analyses on the adolescents (aged 14-18) and emerging adults (ages 19-24) in the present sample in order to develop a propensity score specific to each subgroup (Green and Stuart, 2014), with sufficient numbers for each subgroup.

In a community sample of young people who report distress above the population mean, I anticipate that treatment effects will be related to decreased distress over two years following propensity weighting. These effects will be compared between the adolescents and emerging adults. I anticipate that treatment effects on perceived family functioning will only be apparent in the adolescents. In this age group, family functioning or distress may be a mediator of treatment effects. Although less likely than for family functioning, improvements in perceived friendship support may be related to treatment particularly in adolescents. Regardless, as treatments do target social skills in general, the possibility of improved social support across the whole age range is worth exploring.

4.2 Method

4.2.1 Sample: Young people aged 14-24 were recruited in Cambridgeshire and London as part of the Neuroscience and Psychiatry Network (NSPN). Invitations to participate were sent via general practitioners, schools, colleges, and purposive advertisements (Kiddle *et al.*, 2018). Baseline data were collected from November 2012 until April 2016, with 84% (2,026/2,403) collected in 2013. The recruited sample was age- and sex-stratified, with baseline (T1) data being collected at an average of age of 19.1 years (SD=3.00). Follow-up assessments were completed annually, with T2 occurring at a mean age of 20.1 years (SD=3.10), and T3 at 21.3 years (SD=3.15).

4.2.2 Measures

Predictor: Mental health treatment was based on response to the question "Are you currently being treated for any emotional, behavioural or mental health problem?". At T1, parents/carers responded if the participant was under age 16, otherwise the youth responded. Self-reported diagnoses were coded from a follow-up open-ended question. These questions were repeated at T2 and T3. Treatment could have consisted of a psychological and/or a pharmacological intervention. Length of treatment was not systematically recorded. While the source of treatment was unknown, no participants were institutionalized, so treatment was community based. This could therefore include mental health services from schools, GP practices, the voluntary sector, and specialist mental health services (Care Quality Commission, 2017). Treatment may also have been received from other sources such as private care.

Potential mediators: Perceived family functioning was measured by the self-report 12-item general functioning subscale from the McMaster Family Assessment Device (FAD-GF) (Epstein, Baldwin and Bishop, 1983). Using a balance of positively and negatively worded items, FAD-GF assesses problem solving, family communication, and affective responsiveness. Based on exploratory and confirmatory factor analyses, FAD-GF has been shown to provide the best summary score of the larger measure (Ridenour, Daley and Reich, 1999). In a community sample, FAD-GF has demonstrated reliability and discriminability between who are and are not receiving mental health services (Mansfield, Keitner and Sheeran, 2018). Across several clinical or help-seeking samples, FAD-GF has shown the expected divergent validity with a structured interview of family functioning and a measure of couple or marital satisfaction (Mansfield, Keitner and Sheeran, 2018). FAD-GF has demonstrated acceptable test-retest reliability and low correlations with social desirability (Miller *et al.*, 1985). The FAD-GF has been used to assess change in families following treatment. For example, in clinical trials of depression and bipolar disorder, FAD-GF exhibited significant improvements from baseline to post-treatment 3-4 months later (Mansfield, Keitner and Sheeran, 2018). Finally, FAD-GF exhibited excellent internal consistency at all three timepoints in the present study (α=.92-.93). Items were coded such that a higher score reflected worse family functioning.

Perceived friendship support was measured using the 8 item self-report Cambridge Friendship Scale (CFQ). CFQ assesses the number, availability, and quality of friendships, with three negatively worded questions inversely coded for the total score. Developed for use in an adolescent cohort study (Goodyer *et al.*, 2010), this measure has demonstrated invariance, reliability, and validity (van Harmelen *et al.*, 2016). In the present study, CFQ demonstrated acceptable internal consistency at all three timepoints (α=.71-.73).

The ***primary outcome*** was general distress, the general factor (*p)* from the bifactor model developed in Chapter 3, which explained 73%-75% of the common variance in depression, anxiety, antisocial behaviour, psychotic-like symptoms, obsessionality, self-esteem, and well-being self-report measures (St Clair *et al.*, 2017) (Chapter 3). Participants reported on their thoughts, feelings, and behaviours over the past two weeks. More details of this model are found in Section 3A.4. General distress (hereafter termed "distress") exhibited excellent construct replicability at all three timepoints (*H*=.99, Chapter 3).

Covariates were background variables (age, sex, centre, ethnicity, deprivation (Smith *et al.*, 2015), foreign birth), as well as self-reported adverse life events (Life Events Questionnaire (Goodyer, Kolvin and Gatzanis, 1985; Goodyer *et al.*, 2000)), which has been related to mental disorder (Goodyer, Wright and Altham, 1990), and harmful behaviours (cigarette use, alcohol use, drunkenness, cannabis use, and other illegal drug use, Non-Suicidal Self Harm [NSSI] (Goodyer *et al.*, 2010; Wilkinson *et al.*, 2018), which for the most part increase during adolescence and emerging adulthood (Brodbeck *et al.*, 2013; Plener *et al.*, 2015; Stormshak *et al.*, 2018). Life events and harmful behaviours were repeatedly assessed at each wave. Life events were reported over the past 18 months at baseline, and the past year at subsequent waves. Harmful behaviours were reported over the past month except for NSSI, which was reported over the past year. The presence of past mental health problems was probed at baseline following the treatment question above ("Have you had any similar or related problems in the past?").

4.2.3 Statistical analysis

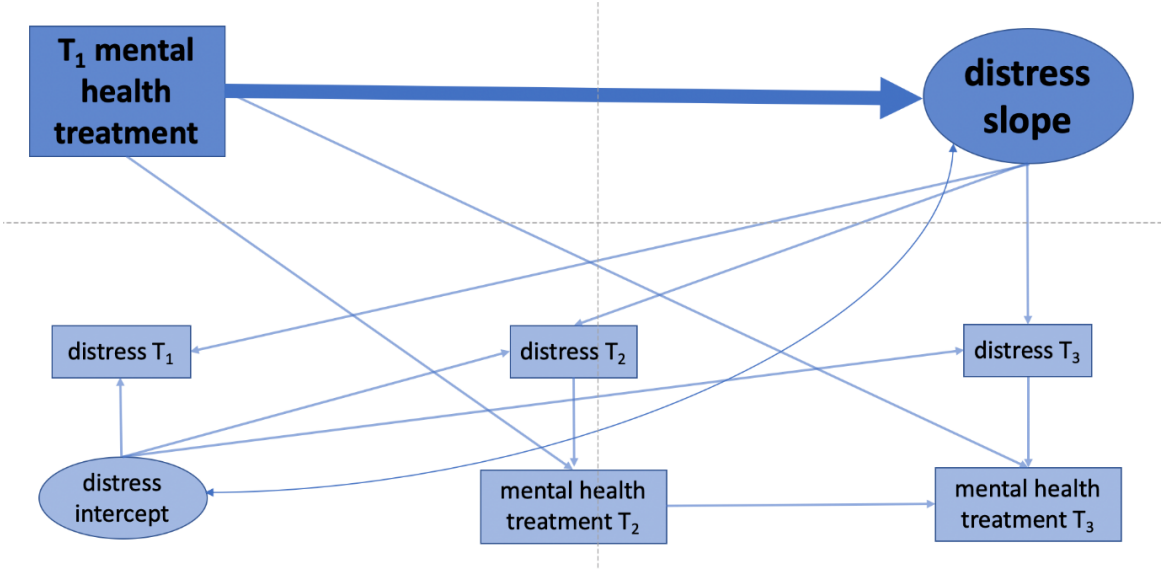
Propensity weighting: In the above average distress sample, a propensity score was generated to balance levels of baseline variables between the treated and untreated group, similar to a randomized control trial (Austin, 2011). In the present study, baseline variables assessed for balance included all covariates, mediator, and the outcome listed above and in Table 4.1. As any missing data in these variables would result in listwise deletion of participants in generating the propensity score, single imputation was performed prior to estimating the propensity score. For the purposes of generating a propensity score, Monte Carlo simulations have shown that single imputation performs as well as multiple imputation, provided the propensity score model includes missing data indicators (Leite, Stapleton and Bettini, 2019). This allows for balance in missing data patterns as well as the observed variables (Rubin and Rosenbaum, 1984). Imputation in the

distressed sample was performed using mi impute in Stata with the predictor, mediator, outcome, and all covariates at all timepoints.

The propensity score was then generated from a logistic regression predicting baseline mental health treatment. All baseline variables related to the outcome (distress slope, standardized estimate>.1) were included in the model, regardless of their association with treatment, as this decreases bias and variance (Brookhart *et al.*, 2006; Pirracchio, Resche-Rigon and Chevret, 2012). Indicators of missingness in these variables were also included. The propensity method used was inverse probability of treatment weighting (IPTW). This method has been shown to be less biased in studies with a low prevalence of treatment (Pirracchio, Resche-Rigon and Chevret, 2012), as in the present study. Further, IPTW allows for estimating the average treatment effect in a whole population, unlike some other propensity methods which estimate the average treatment effect in the treated only (Austin, 2011). Correct model specification was determined by assessing balance of baseline variables after weighting. Balance strove to minimize the number of variables with large standardized differences (>.25) between treated and untreated groups (Stuart, 2010). If significant imbalance was detected, the propensity score model was re-specified by iteratively including the unbalanced variables (Rubin and Rosenbaum, 1984). The IPTW was then used to weight the relevant Mplus models, requiring a robust estimator. As IPTW can result in extreme weights, weighted analyses were then limited to the region of common support – the range of propensity scores which are present in the treated and untreated groups (Heckman *et al.*, 1996).

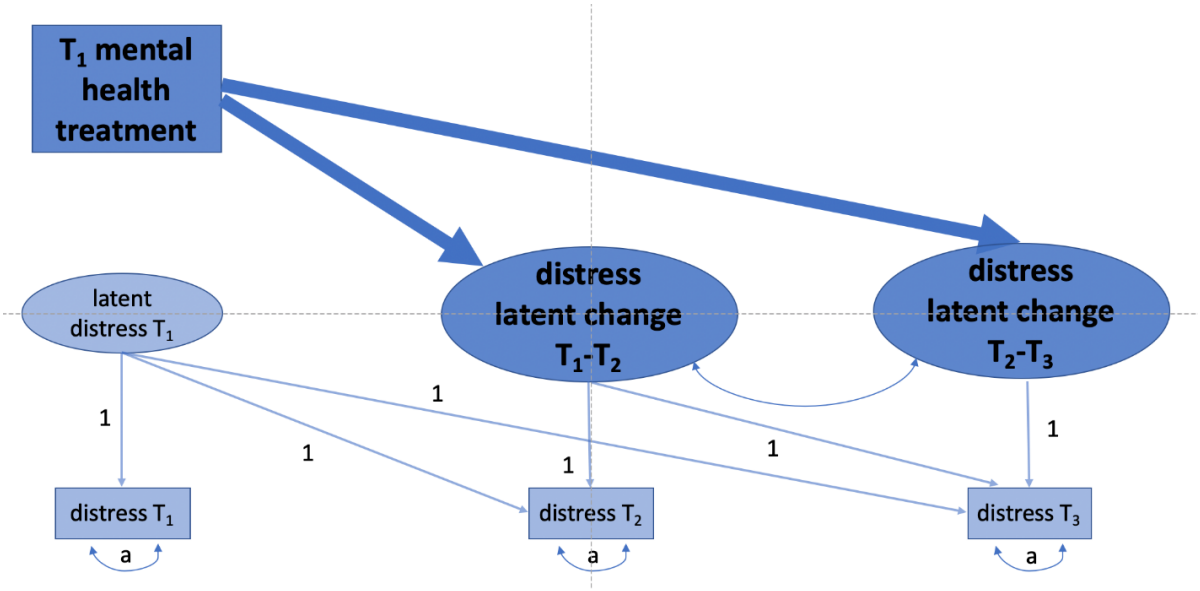
Treatment on latent slopes: As intraindividual change over time related to mental health treatment was considered to be of primary importance, the potential mediators and outcome were modelled as latent slopes. Slope models are presented with increasing robustness: unadjusted, propensity weighted, and propensity weighted adjusted by subsequent treatment (as possible). As propensity weighting results in treatment being unrelated to baseline variables, the relationship with each model's intercept was not required (Goldsmith *et al.*, 2018a). These models were first specified separately for outcome and mediators and tested together in a mediation model if effects warranted this. Latent growth models were specified from T1-3, testing the appropriateness of linear and quadratic change. Starting values for all growth models were set at the mean time of each assessment since baseline. Figure 4.1A depicts the latent growth model, specified as recommended (Goldsmith *et al.*, 2018a) but also including treatment2 and treatment3 specified as time-varying covariates (Muthén and Muthén, 2017). For sensitivity analyses, a latent change model allowed for change to be different over T1-T2 compared with T2-T3 (Goldsmith *et al.*, 2018a). Figure 4.1B depicts the latent change model, specified as recommended (Steyer, Eid and Schwenkmezger, 1997; Goldsmith *et al.*, 2018a). This model had the added benefit of allowing the mediator to precede the outcome in time. For both models in Figure 4.1, T1 mental health treatment is propensity weighted and thus uncorrelated with T1 distress.

Figure 4.1: Latent models of baseline (T1) mental health treatment on distress slope
A: Latent growth model, adjusted by mental health treatment at T2 and T3 as time-varying covariates[a].



[a] Treatment2 and 3 were predicted by prior treatment variables in order to appropriately estimate these categorical variables in the presence of missing data.

B: Latent change model[a]

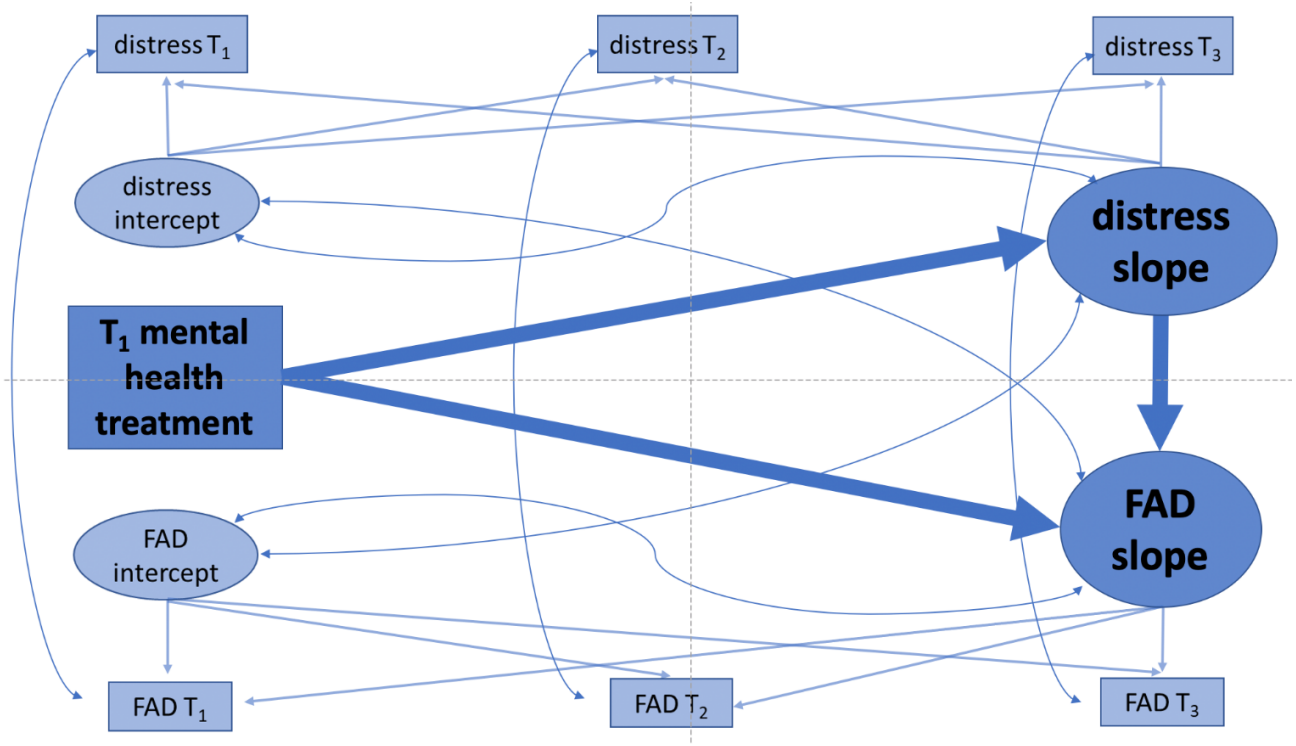


[a] All factor loadings were set at 1. Residual correlations were constrained to be equal (a) to enable identification.

Mediation: The latent growth and latent change models offer different advantages for mediation. Mediators modelled as T1-T2 latent change precede the outcome in time, since the outcome is modelled as T2-T3 latent change. A T1-T3 latent growth model allows a longer time period for treatment to show an effect on the mediator but removes the temporal nature of the mediator and outcome. Such comparison of various models has been encouraged as a robust approach to the parameterization of mediation effects (Goldsmith *et al.*, 2018a). Both latent variable

approaches to mediation are advantageous in helping account for bias. The effect of the mediator on the outcome is independent of measurement error (Goldsmith *et al.*, 2018b). Further, the inclusion of error covariances can address unmeasured confounding, representing factors which influence both meditator and outcome on the same day (contemporaneous) or over time (lagged) (Goldsmith *et al.*, 2018b). Therefore, models which included lagged measurement error covariances were compared against those with contemporaneous measurement error covariances. The effect of constraining these covariances was also assessed, with the fit statistic Sample Size adjusted Bayesian Information Criterion (SS BIC) being used to compare models (ΔSS BIC<2 denotes equivalent models (Kass and Raftery, 1995)). A depiction of a latent growth mediation model with distress as the mediator and contemporaneous error covariances is shown in Figure 4.2.

Figure 4.2: Latent growth mediation model with distress as mediator and contemporaneous error covariances[a]



[a] T1 Mental health treatment is propensity weighted and thus uncorrelated with T1 FAD or distress. T=timepoint; FAD=Family Assistive Device general perceived family functioning.

To address confounding of treatment on putative mediators and outcome, all mediation models were propensity weighted and adjusted for subsequent treatment. (As propensity weighted models require a robust estimator, bootstrapping was not available for indirect effects (Muthén and Muthén, 2017)). To address measured confounding of the mediator-outcome path (VanderWeele, 2016), the relationship of each time-invariant covariate on these individual slopes was assessed

alongside treatment 1-3 (propensity weighted models). Covariates with estimates>.1 and p<.1 on both mediator and outcome were included in the adjusted mediation model. Time-varying covariates related to mediator and outcome at baseline (r/rho>.1) were similarly assessed. In the presence of multicollinearity, covariates most related to mediator and outcome were retained. Autocorrelations and cross-correlations of time-varying covariates were included in adjusted mediation models if p<.1.

Missingness: Full-information maximum likelihood (FIML) was used to deal with missing data in the longitudinal SEMs, an appropriate technique for SEMs such as growth modelling (Allison, 2003). FIML uses all available information from the observed data and Mplus implements procedures which are robust to non-normality (Allison, 2003; Muthén and Muthén, 2017). In longitudinal data, FIML estimates are similar to those obtained through multiple imputation (Ferro, 2014). FIML has been used for growth modelling in longitudinal data sets with retention rates as low as 25% (Finan, Ohannessian and Gordon, 2018). In the present study, missing data patterns were assessed to ascertain levels of bias due to missingness (Little, 1988). Monte Carlo integration with a robust estimator was used to address missingness in post-baseline categorical covariates (Muthén and Muthén, 2017). Due to this missingness, only raw estimates were available when post-baseline categorical variables were in the model, and the goodness-of-fit statistics below were also unavailable. Therefore, for comparative purposes, raw estimates are presented in tables. For treatment effect sizes equivalent to Cohen's *d*, model estimates are reported where possible standardized with respect to the outcome (small, medium, and large effect sizes correspond to *d* of .2, .5, and .8 respectively, (Cohen, 1988)).

Baseline analyses and determination of auxiliary variables were performed in STATA 14; all other analyses were performed in Mplus 8 (Muthén and Muthén, 2017). Continuous variables were centred or standardized according to which afforded the best model fit, and to aid interpretation of estimates. For unweighted models, a robust estimator was used if variables exhibited non-normality. Models were specified firstly to obtain acceptable fit (RSMEA<.05, and CFI and TLI>.950; van de Schoot et al. 2012; Hu & Bentler 1999), although slightly more lenient values were accepted if necessary (CFI > 0.90 (Tucker and Lewis, 1973; Bentler, 1990, 1992) and RSMEA < 0.08; (MacCallum, Browne and Sugawara, 1996)). Thereafter, models with a lower SS BIC were favoured; a difference of less than two indicated models of equivalent fit (Kass and Raftery, 1995).

4.3 Results

4.3.1 Descriptive statistics: In the full sample of 2,403, 6.6% (158/2387) reported current mental health treatment at baseline. Rates were 4.7% (60/1,269) for adolescents aged 14-18 (mean=16.66, SD=1.43), and 8.8% (98/1,020) for emerging adults aged 19-24 (mean=21.82, SD=1.67). A fifth (19.5%, n=466) of all responses to the treatment question were obtained from primary caregivers (and 37% [466/1.269] of the adolescent responses), as these participants were

under age 16. Those with mental health treatment were more likely to be female, older, of white ethnicity, and to have had past mental health problems (Table 4.1). They experienced more distress and adverse life events, worse family functioning and friendship support. They engaged in more NSSI but less cigarette use. Mental health diagnoses were only self-reported for those who endorsed current mental health treatment, most common being depression (42%, 66/158) and anxiety (30%, 47/158), with behavioural (8%, 13/158), and psychosis/schizophrenia (4%, 6/158) less common. Sixteen percent endorsing treatment (26/158) did not report a diagnosis. Among all who reported no current mental health treatment or reported a diagnosis and treatment (n=2,361), females were more likely to report depression (OR=2.37 [95% CI: 1.37, 4.10], p=.002) and anxiety diagnoses (OR=1.88 [1.01, 3.49], p=.046), whereas males were more likely to report behavioural diagnoses (OR=6.38 [1.41, 28.85], p=.016).

Table 4.1: Relationship of baseline variables with baseline current mental health treatment, full sample[a]

Baseline variables	**Mean (SD) or % (n)**		**Covariate with treatment (n=2,235-2,373)**	
	treated (n=147-157)	**untreated (n=2,085-2,216)**	**r/rho**	***p***
Background				
Age	20.15 (2.80)	19.00 (3.01)	0.10	<0.001
Sex (1=female)	63% (99)	53% (1,173)	0.12	0.016
Centre (1=Cambridge)	65% (102)	58% (1,278)	0.09	0.079
Ethnicity (1=non-white)	12% (19)	22% (486)	-0.18	0.003
Deprivation	15.14 (10.90)	15.13 (11.89)	0.00	0.99
Foreign birth	8% (13)	14% (301)	-0.14	0.052
Past mental health problems	48% (72)	9% (201)	0.59	<0.001
Environment				
Poor family functioning	27.89 (8.14)	23.84 (6.91)	0.14	<0.001
Supportive friendships	20.03 (5.44)	22.75 (4.01)	-0.16	<0.001
Adverse life events	2.10 (1.88)	1.09 (1.42)	0.17	<0.001
Harmful behaviour				
Non-suicidal self-injury	59% (88)	16% (327)	0.56	<0.001
Cigarette use			0.12	<0.001
None	59% (93)	79% (1,748)		
Occasionally	16% (25)	8% (186)		
Often	25% (39)	12% (274)		
Alcohol use			0.03	0.16
None	29% (46)	36% (784)		
Occasionally	59% (92)	53% (1,177)		
Often	12% (19)	11% (243)		
Drunkenness			0.01	0.69
None	46% (72)	46% (1,015)		
Occasionally	29% (45)	30% (672)		
Often	26% (40)	23% (509)		
Cannabis use			0.00	0.90
None	87% (135)	87% (1,930)		
Occasionally	10% (16)	9% (199)		
Often	3% (5)	4% (79)		
Other illegal drug use	5% (8)	4% (96)	0.04	0.55
Outcome (distress)	0.96 (0.96)	-0.05 (0.92)	0.26	<0.001

[a] Those with baseline data on mental health treatment and at least one timepoint of distress, n=2373. Data is at least 94% present for each baseline variable.

Hereafter, all statistics pertain to the above average distress sample unless specified. Of the 1,157 participants above the population mean on distress, 12% (n=133) reported mental health treatment, representing 85% (133/157) of those who reported treatment in the whole sample (one reporting treatment did not have enough data to generate the baseline distress factor). Compared with the full sample, participants in this subsample exhibited similar relationships of baseline variables with treatment to that of the full sample (Δr/rho from 0 to 0.08). However, participants above the population mean on distress had half the number of extreme standardized differences (>.5) between treated and untreated groups, making the balance of baseline variables between groups possible. Compared with those at or below the population mean on distress, participants with higher distress scores were no different in age but more likely to be female, non-white, foreign born, and from London (Table 4.2). They were more deprived, disadvantaged on all environmental variables, and more likely to engage in all harmful behaviours, with the exception of alcohol use. Further, they were more likely to report previous mental health problems. Of the full sample who reported a current mental health diagnosis (n=132), those above the population mean in distress were more likely to report depression (54% versus 25% below distress mean, $\chi^2(1, n=131)=4.70$, p=.030), but somewhat less likely to report anxiety (32% versus 56% $\chi^2(1, n=131)=3.57$, p=.059). All reported psychosis/schizophrenia diagnoses and 77% (10/13) of behavioural disorders were in the higher distress sample.

Table 4.2: Relationship of baseline variables with distress dichotomized at mean[a]

Baseline variables	Mean (SD) or % (n)		Variable with distress dichotomized at mean (n=2,221-2,356)	
	Average and below (n=1,131-1,199)	Above average (n=1,090- 1,157)	r/rho	*p*
Background				
Age	19.08 (3.07)	19.08 (2.95)	0.00	0.97
Sex (1=female)	49% (582)	59% (682)	0.16	<0.001
Centre (1=Cambridge)	61% (730)	56% (644)	-0.08	0.011
Ethnicity (1=non-white)	18% (217)	25% (283)	0.14	<0.001
Deprivation	14.40 (11.60)	15.88 (12.02)	0.06	0.003
Foreign birth	11% (134)	16% (177)	0.12	0.003
Past mental health problems	6% (76)	17% (195)	0.34	<0.001
Environment				
Poor family functioning	21.48 (5.91)	26.82 (7.15)	0.38	<0.001
Supportive friendships	24.18 (3.13)	20.91 (4.46)	-0.39	<0.001
Adverse life events	.74 (1.11)	1.59 (1.66)	0.29	<0.001
Harmful behaviour				
Non-suicidal self-injury	6% (67)	32% (346)	0.60	<0.001
Cigarette use			0.14	<0.001
None	84% (1,006)	71% (824)		
Occasionally	6% (68)	12% (139)		
Often	10% (118)	17% (193)		
Alcohol use			0.01	0.71
None	35% (413)	36% (409)		
Occasionally	55% (660)	52% (602)		
Often	10% (121)	12% (139)		
Drunkenness			0.08	<0.001
None	49% (584)	43% (493)		
Occasionally	30% (357)	31% (355)		
Often	21% (245)	26% (302)		
Cannabis use			0.09	<0.001
None	90% (1,079)	84% (972)		
Occasionally	7% (80)	11% (132)		
Often	3% (34)	4% (50)		
Other illegal drug use	4% (42)	5% (61)	0.12	0.043
***Outcome* (distress)**	-0.72 (0.56)	0.79 (0.61)	0.79	<0.001
Predictor				
Mental health treatment	2% (24)	12% (133)	0.48	<0.001
Diagnosis related to treatment				
Depression	0.3% (4)	5% (62)	0.58	<0.001
Anxiety	0.8% (9)	3% (37)	0.35	<0.001
Behavioural	0.3% (3)	0.9% (10)	0.26	0.052
Psychosis / schizophrenia	n=0	0.5% (6)	1.00	0.014

[a] all cases have baseline distress and treatment data. Data is at least 94% present for each baseline variable.

4.3.2 Missing data: All 1,157 participants above the population mean on baseline distress had data on baseline mental health treatment. 67% (n=777) and 44% (n=507) had distress data at T2 and T3 respectively, with similar rates for FAD and CFQ. The retention rate at T3 is similar to other studies of adolescents and emerging adults two years after baseline (Levitt, Silver and Santos, 2007). Little's MCAR (Little, 1988) was non-significant in separate models of distress and FAD at all three timepoints, indicating these data are missing completely at random (MCAR: distress $\chi^2(5, n=1,157)=3.62$, p=.61; FAD $\chi^2(9, N=1138)=9.23$, p=.42). While CFQ data did not

appear to be MCAR (χ^2(8, n=1,153)=21.50, p=.006), it did demonstrate covariate-dependent missingness (CDM) based on mental health treatment and propensity score variables (χ^2(104, n=988)=92.46, p=.78; Table 4.3 for variables). CDM is a type of missing at random data. When data are MCAR or exhibit CDM, analysis of complete cases is not biased, however methods which use all the data (eg. FIML) are more efficient (Little, 1995). The lack of bias in these patterns of missingness means there are no systematic differences between the missing values and the observed values. Therefore, employing FIML to analyse the present data means the findings will not be subject to bias due to missingness.

Table 4.3: Standardized estimates of baseline variables on distress slope[a]

Baseline variables	**All above average distress (n=1157)[b]**		**Adolescents (ages 14-18, n=625)[c]**		**Emerging adults (ages 19-24, n=532)[d]**	
	estimate	***p***	**estimate**	***p***	**estimate**	***p***
Background						
Age	**0.134**	0.007	**-0.130**	0.056	0.034	0.67
Sex (1=female)	0.018	0.87	**0.198**	0.18	**-0.233**	0.17
Centre (1=Cambridge)	**-0.156**	0.14	**-0.163**	0.23	**-0.135**	0.42
Ethnicity (1=non-white)	**0.104**	0.40	-0.026	0.87	**-0.171**	0.40
Deprivation	**-0.132**	0.015	**-0.140**	0.037	**-0.121**	0.17
Foreign birth	**0.098**	0.056	0.084	0.16	0.076	0.37
Past mental health problems	**-0.106**	0.049	-0.048	0.52	-0.074	0.35
Environment						
Poor family functioning	-0.040	0.46	-0.054	0.44	0.003	0.98
Supportive friendships	0.086	0.16	**0.106**	0.12	-0.007	0.93
Adverse life events	0.014	0.80	-0.032	0.66	0.006	0.94
Harmful behaviour						
Non-suicidal self-injury	**-0.149**	0.005	**-0.140**	0.037	**-0.147**	0.081
Cigarette use	0.045	0.43	0.021	0.78	0.015	0.84
Alcohol use	0.053	0.33	0.021	0.75	-0.062	0.58
Drunkenness	0.013	0.81	-0.043	0.54	-0.030	0.73
Cannabis use	**0.230**	0.019	**0.137**	0.074	**0.128**	0.31
Other illegal drug use	0.023	0.71	0.045	0.57	-0.012	0.90
***Outcome* (distress)**	0.008	0.95	**0.106**	0.62	**-0.160**	0.29

[a] Latent growth model, T1-T3. Estimates standardized with respect to x and y for continuous predictors and with respect to y for categorical predictors. Baseline variances were included if missing data was in the baseline variable, allowing this missingness (no more than 6%) to be estimated using FIML.
[b] All RMSEA<=.026, CFI>=.995, TLI>=.987. Once all variables with standardized estimates (STD est)>=.1 were included in the propensity model, distress, cigarette use, and life events were iteratively added to further balance the model.
[c] All RMSEA<=.043, CFI>=.990, TLI>=.971. Once all variables with STD est>=.1 were included in the propensity model, past mental health problems, life events, alcohol use, drunkenness, and family functioning were iteratively added to further balance the model.
[d] All RMSEA<=.049, CFI>=.988, TLI>=.965. Once all variables with STD est>=.1 were included in the propensity model, past mental health problems, foreign birth, friendships, and cigarette use were iteratively added to increase balance.

4.3.3 Treatment predicting distress: Due to a non-significant quadratic term in the growth model of distress (Table 4.4), a linear model was pursued. Unadjusted, mental health treatment predicted decreased distress over the course of two years to a moderate effect size (*d*=-0.41, Table 4.4). The latent change model revealed this effect to be present one to two years following baseline treatment as opposed to immediately following treatment.

Table 4.4: Baseline mental health treatment on distress slope[a]

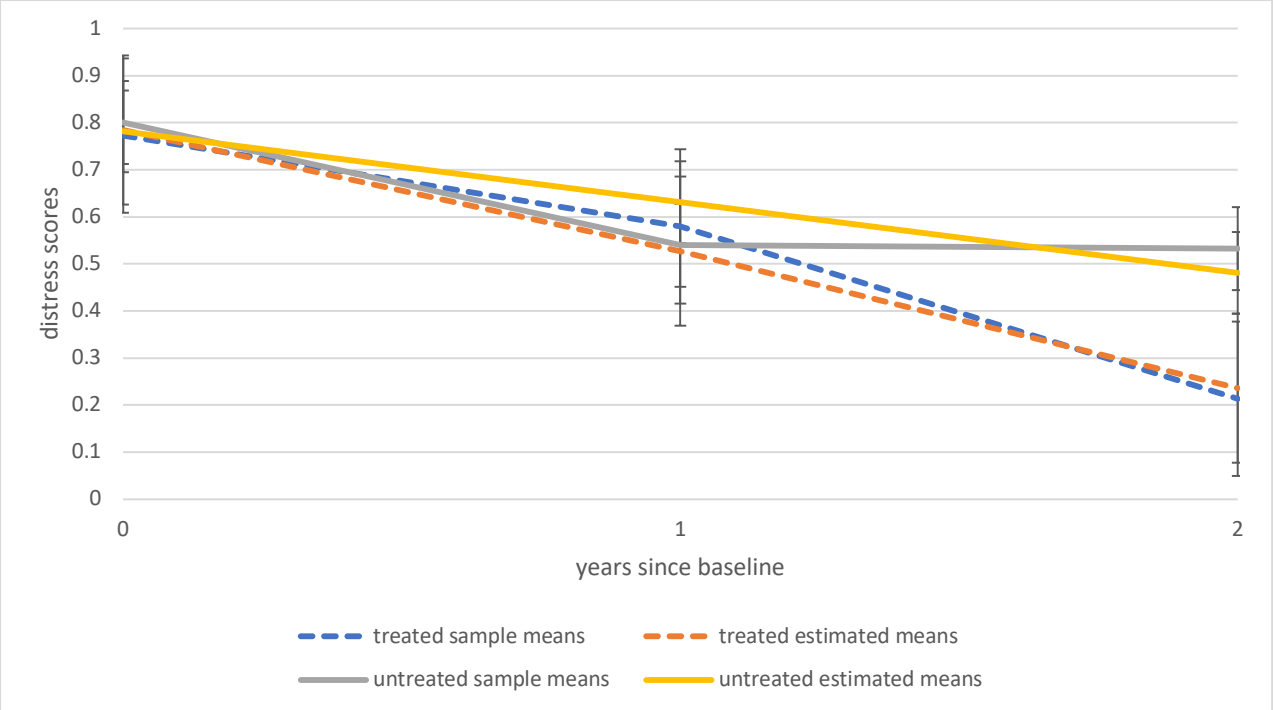
	Raw estimates (S.E.)	p-value	SS BIC	Free parameters	RMSEA	CFI	TLI
A) Distress latent growth model			4919	8	0.000	1.000	1.007
Intercept mean	0.000 (0.018)	0.98					
Slope mean (linear)	0.010 (0.016)	0.54					
Intercept with slope	-0.001 (0.018)	0.94					
Quadratic slope mean (separate model[b])	-0.006 (0.017)	0.74	4930	12			
B) Treatment on distress slope – latent growth model							
Unadjusted model	-0.119 (0.045)	0.008	4838	10	0.000	1.000	1.007
Propensity weighted (IPTW)[c]	-0.137 (0.069)	0.046	5072	9	0.021	0.991	0.982
IPTW with T2 & T3 treatment[d]	-0.209 (0.064)	0.001	5809	16			
Common support sample[e]	-0.135 (0.069)	0.049	4938	9	0.023	0.990	0.980
Common support sample with T2 & T3 treatment[d]	-0.208 (0.064)	0.001	5655	16			
C) Treatment on distress slope – latent change							
Unadjusted			5488	13	0.000	1.000	1.009
T1-T2	-0.058 (0.081)	0.48					
T2-T3	-0.209 (0.090)	0.020					
Propensity weighted[c]			5071	10	0.008	0.999	0.998
T1-T2	0.049 (0.119)	0.68					
T2-T3	-0.392 (0.159)	0.013					
Common support sample IPTW (n=1123)			4937	10	0.009	0.999	0.997
T1-T2	0.049 (0.120)	0.68					
T2-T3	-0.389 (0.159)	0.014					

[a] n=1157 unless common support sample (see [e]). Raw estimates are reported for comparison with models where standardized estimates are not available.
[b] Model was not identified with starting values for growth models set at mean time of each assessment since baseline, so exact time of assessment per individual was used. For such a random effects model, only raw estimates are available and no goodness-of-fit statistics (GoF: RMSEA, CFI or TLI)
[c] variables in propensity model: age, centre, deprivation, foreign born, past treatment, distress, life events, NSSI, cannabis use, cigarette use, and indicators of missingness in these variables. Raw estimate=-0.141 (SE=0.071) for comparison with models using categorical post-baseline variables.
[d] Only raw estimates were available for this model due to missingness in categorical post-baseline variables. GoF statistics were also unavailable for this reason.
[e] Sample trimmed to include only those propensity scores which existed in treated and untreated groups. n=1123: 99% (131/133) of the treated sample and 97% (992/1,024) of the untreated sample

***Propensity weighting**:* The final propensity score model (Table 4.3) balanced all baseline variables (standardized differences from .001 to .216, mean of .085), indicating appropriate specification of the model. Following propensity weighting, mental health treatment effects on distress became stronger in both the latent growth (*d*=-.50) and latent change models (*d*=-.80; Table 4.4; Figure 4.3), with both models yielding equivalent fit. Very comparable effects were seen in the common support models, but model fit was substantially better due to extreme propensity weights being trimmed from this subsample. In the growth model, subsequent treatment could be modelled

as a time-varying covariate (ie: treatment2 was regressed onto distress2, and treatment3 onto distress3 (Muthén and Muthén, 2017)). Both precision and effect sizes of baseline treatment on distress slope increased with this addition (Table 4.4). (Subsequent treatment was not added to the latent change model, as treatment2 would become a predictor of T2-T3 distress slope. An additional propensity score for treatment2 would then need to be estimated to properly quantify these effects (Linden and Adams, 2010). However, these effects were anticipated to be non-significant, similar to the proximal effects of treatment1.)

Figure 4.3: Baseline mental health treatment on distress slope[a]



[a] for illustrative purposes of the estimated means, a propensity weighted latent growth model was run separately for the treated and untreated group. Sample means are also propensity weighted.

***Perceived family functioning and friendship support*:** In the latent change models, treatment was not significantly related to improved family functioning or friendship support from T1-T2 in unadjusted and propensity weighted models (Table 4.5. The latent change models including T1-T2 and T2-T3 change (Goldsmith *et al.*, 2018a) had latent variable covariance matrices which were not positive definite, so T1-T2 only models were specified (Selig and Preacher, 2009)). The latent growth models of FAD and CFQ each had a non-significant quadratic term and thus were modelled linearly. Treatment was not significantly related to FAD or CFQ slope over two years in unadjusted and propensity weighted models, with and without subsequent treatment.

Table 4.5: Baseline mental health treatment on FAD[a] and CFQ[b] slopes

	Raw Estimate (S.E.)	p-value	SS BIC	Free parameters	RMSEA	CFI	TLI
A) Treatment on FAD – T1-T2 latent change model[c]							
Unadjusted	-0.179 (0.096)	0.063	5580	9	0.000	1.000	1.000
Propensity weighted (IPTW)	-0.030 (0.130)	0.82	4995	6	0.000	1.000	1.012
Common support sample (n=1104)	-0.037 (0.130)	0.78	4847	6	0.000	1.000	1.013
B) FAD latent growth model			5962	8	0.000	1.000	1.003
Intercept mean	-0.007 (0.029)	0.80					
Slope mean (linear)	0.017 (0.017)	0.31					
Intercept with slope	0.015 (0.031)	0.63					
Quadratic slope mean (separate model)[d]	0.007 (0.019)	0.71	5968	12			
C) Treatment on FAD slope – latent growth model							
Unadjusted	-0.084 (0.047)	0.07	5960	10	0.040	0.994	0.983
Propensity weighted	-0.060 (0.069)	0.39	6041	9	0.000	1.000	1.015
IPTW with T2 & T3 treatment[e]	-0.086 (0.070)	0.22	6848	16			
Common support sample (n=1105)	-0.057 (0.070)	0.41	5865	9	0.000	1.000	1.017
Common support sample with T2 & T3 treatment[e]	-0.085 (0.071)	0.23	6648	16			
A) Treatment on CFQ slope – T1-T2 latent change model							
Unadjusted	-0.080 (0.090)	0.37	5840	9	0.000	1.000	1.000
Propensity weighted (IPTW)	-0.126 (0.122)	0.30	5233	6	0.013	0.997	0.994
Common support sample (n=1119)	-0.124 (0.122)	0.31	5072	6	0.018	0.994	0.987
B) CFQ latent growth model			6431	8	0.000	1.000	1.009
Intercept mean	-0.002 (0.029)	0.93					
Slope mean (linear)	-0.004 (0.020)	0.83					
Intercept with slope	-0.059 (0.045)	0.19					
Quadratic slope mean (separate model)[d]	-0.009 (0.020)	0.67	6434	12			
C) Treatment on CFQ slope – latent growth model							
Unadjusted	-0.012 (0.059)	0.84	6423	10	0.000	1.000	1.007
Propensity weighted	-0.001 (0.081)	1.00	6498	9	0.000	1.000	1.003
IPTW with T2 & T3 treatment[e]	-0.032 (0.116)	0.78	7314	16			
Common support sample (n=1119)	-0.006 (0.081)	0.94	6285	9	0.003	1.000	1.000
Common support sample with T2 & T3 treatment[e]	-0.046 (0.116)	0.69	7079	16			

[a] n=1138, as 19 did not have FAD data at any timepoint.
[b] n=1153, as 4 did not have CFQ data at any timepoint
[c] n=1137: one person had FAD data only at T3, so was not able to be used for T1-T2 only models.
[d] Model was not identified with starting values for growth models set at mean time of each assessment since baseline, so exact time of assessment per individual was used. For such a random effects model, only raw estimates are available and no goodness-of-fit statistics (GoF: RMSEA, CFI or TLI). 1 case missing time score.
[e] Only raw estimates were available for due to missingness in categorical post-baseline variables. GoF statistics were also unavailable for this reason.

Adolescents *(ages 14-18 [mean=16.75, SD=1.40], n=625)* ***and emerging adults*** *(ages 19-24 [mean=21.81, SD=1.68], n=532)***:** In the above average distress sample, emerging adults had a twofold odds of reporting treatment compared with adolescents (OR=2.05 [1.42, 2.97], p<.001; 15% [82/532] versus 8% [51/625]). However, for each group, 85% of participants who reported treatment were in the above average distress subsample. Gender was evenly distributed between adolescents and emerging adults (p=.25; 61% [n=378] and 57% [n=304] female, respectively). Similarly, there were no differences in reported diagnoses between adolescents and emerging adults (depression p=.49, 50% versus 57% respectively; anxiety p=.37, 37% versus 29% respectively; behavioural 9% each; psychosis/schizophrenia 4% versus 6% respectively). Baseline variables were balanced for adolescents and emerging adults by devising separate propensity models based on each groups' predictors of distress (Table 4.3; standardized differences following IPTW: adolescent range .001-.424, mean=.137; emerging adult range .002-.305, mean=.090).

As in the full analysis, latent growth models in the adolescent and emerging adult samples yielded non-significant quadratic terms (p>.44), so linear models were specified. All latent change scores from T1-T2 were non-significant with treatment in adolescent (p>.14) and emerging adult (p>.32) groups.

In both adolescents and emerging adults, treatment was significantly related to decreased distress when the model was propensity weighted and subsequent treatment was included (Tables 4.6-4.7). These effects were comparable in the common support sample with subsequent treatment. In both models, effects were smaller for the emerging adults, but standard errors yielded overlapping 95% confidence intervals, indicating that effects were not significantly smaller. In all propensity weighted models, treatment was related to improvements in perceived family functioning in the adolescents only (Tables 4.6-4.7). Treatment was not related to change in perceived friendship support in either adolescents or emerging adults.

Table 4.6: Adolescents (ages 14-18): baseline mental health treatment on distress, poor family functioning, and friendship support slopes[a]

A) Treatment on FAD slope (n=616, missing 9)	**Raw Estimate (S.E.)**	**p-value**	**SS BIC**	**Free parameters**	**RMSEA**	**CFI**	**TLI**
Unadjusted	-0.113 (0.081)	0.16	3316	10	0.017	0.999	0.997
Propensity weighted (IPTW)[b]	-0.184 (0.059)	0.002	3328	8	0.017	0.996	0.994
IPTW with T2 & T3 treatment[b,c]	-0.208 (0.051)	<0.001	3749	15			
Common support sample IPTW (n=504, missing 7)[b]	-0.184 (0.058)	0.002	2708	8	0.021	0.994	0.991
Common support sample with T2 & T3 treatment[b,c]	-0.207 (0.050)	<0.001	3070	15			
B) Treatment on CFQ slope (n=622, missing 3)							
Unadjusted	0.085 (0.099)	0.39	3595	10	0.000	1.000	1.005
Propensity weighted	0.107 (0.131)	0.41	3573	9	0.000	1.000	1.034
IPTW with T2 & T3 treatment[c]	0.146 (0.114)	0.20	3984	16			
Common support sample IPTW (n=509, missing 2)	0.134 (0.128)	0.29	2915	9	0.000	1.000	1.052
Common support sample with T2 & T3 treatment[c]	0.169 (0.113)	0.14	3275	16			
C) Treatment on distress slope (n=625)							
Unadjusted	-0.124 (0.076)	0.10	2747	10	0.000	1.000	1.031
Propensity weighted	-0.208 (0.123)	0.091	2838	9	0.000	1.000	1.071
IPTW with T2 & T3 treatment[c,d]	-0.265 (0.112)	0.018	3210	15			
Common support sample IPTW (n=511)[b]	-0.206 (0.134)	0.12	2364	8	0.023	0.983	0.974
Common support sample with T2 & T3 treatment[b,c]	-0.262 (0.110)	.017	2686	15			

[a] All latent change scores from T1-T2 are non-significant with treatment (p>.14) and therefore not presented for simplicity. FAD=poor family functioning; CFQ=Cambridge Friendship Questionnaire

[b] Not positive definite latent variable covariance matrix was best resolved by respecifying models, taking off the correlation of slope with intercept. This was a non-significant term in all models (p>=.10, estimates<.13). Making this change resulted in models with only very slightly worse model fit (SS BIC increased by values from 2.4 to 5.6) and very minimal change to estimates (.001 to .01). Significant models in table either had an identical p-value to the original or the original was better by up to .007.

[c] Only raw estimates were available for due to missingness in categorical post-baseline variables. Goodness-of-fit statistics were also unavailable for this reason.

[d] as [b], but model fit was equivalent (ΔSS BIC=1.3)

Table 4.7: Emerging adults (ages 19-24): baseline mental health treatment on FAD, CFQ and distress slopes[a]

A) Treatment on FAD slope (n=522, missing 10)	**Raw Estimate (S.E.)**	**p-value**	**SS BIC**	**Free parameters**	**RMSEA**	**CFI**	**TLI**
Unadjusted	-0.106 (0.054)	0.051	2643	10	0.027	0.998	0.994
Propensity weighted	-0.055 (0.077)	0.47	2703	9	0.000	1.000	1.021
IPTW with T2 & T3 treatment[b]	-0.068 (0.074)	0.36	3091	16			
Common support sample IPTW (n=437, missing 10)	-0.057 (0.076)	0.46	2301	9	0.000	1.000	1.021
Common support sample with T2 & T3 treatment[b]	-0.073 (0.074)	0.33	2657	16			
B) Treatment on CFQ slope (n=532, missing 1)							
Unadjusted	-0.015 (0.064)	0.81	2817	10	0.000	1.000	1.021
Propensity weighted	-0.066 (0.076)	0.39	2872	9	0.000	1.000	1.021
IPTW with T2 & T3 treatment[a]	-0.010 (0.080)	0.90	3252	16			
Common support sample IPTW (n=446, missing 1)	-0.066 (0.079)	0.45	2440	9	0.000	1.000	1.026
Common support sample with T2 & T3 treatment[b]	-0.003 (0.083)	0.97	2787	16			
C) Treatment on distress slope (n=532)							
Unadjusted	-0.132 (0.047)	0.005	2090	10	0.008	1.000	0.999
Propensity weighted	-0.089 (0.057)	0.12	2211	9	0.028	0.991	0.982
IPTW with T2 & T3 treatment[b]	-0.156 (0.063)	0.014	2580	16			
Common support sample IPTW (n=447)	-0.104 (0.059)	0.077	1869	9	0.035	0.982	0.963
Common support sample with T2 & T3 treatment[b]	-0.170 (0.064)	0.008	2208	16			

[a] All latent change scores from T1-T2 are non-significant with treatment (p>.32) and therefore not presented for simplicity
[b] Only raw estimates were available for due to missingness in categorical post-baseline variables. GoF statistics were also unavailable for this reason

4.3.4 Mediation in adolescents: Based on the above findings, the relative appropriateness of either FAD or distress as a mediator was assessed in the adolescents only. Firstly, models were specified with all correlations between intercepts and slopes (Goldsmith *et al.*, 2018a), and respecified if required to resolve any problems with estimation. In order for model comparison to be possible, respecification was done in the same way across all models with the same mediator, using the method which resolved the maximal number of estimation problems. In the case of two equally plausible models, estimates from both are presented, as this reflects the level of certainty in the findings (Goldsmith *et al.*, 2018a).

Mediator-outcome confounds were determined in latent growth models (propensity weighted with baseline treatment in the model and treatment2-3 as time varying covariates) separate for mediator and outcome. Of the time-invariant covariates, only foreign birth was related to both slopes (estimates>=.1, p<.05). All other time-invariant covariates were not confounders (estimates<=.05, p=>.27). Of the time-varying covariates, CFQ, LEQ, NSSI, cigarette use, and cannabis use were related to both distress and FAD at baseline (r/rho>.1; all others had r/rho<.07 with distress).

Cigarette and cannabis use were highly correlated (rho=.76), but cigarette use was more related to both baseline distress and FAD, so cannabis use was not retained. All other intercorrelations were acceptable (<.41). In separate latent growth models, CFQ, NSSI, and cigarette use were related to both FAD and distress at each wave (all estimates>=.14 and p<.05 except NSSI2 on FAD2: p=.084 but estimate=.25) and thus were considered for inclusion in the adjusted model along with foreign birth. LEQ, however, was not (estimates<.04, p>.14). Finally, to make models as parsimonious as possible while still addressing confounding, covariates meeting the cut-off were included into a mediation model to test if any of these effects became negligible and thus could be dropped. Foreign birth had negligible effect on FAD slope (estimate=-.003, p=.96) but all other covariates remained substantial (all estimates>=.12 and p<.05 except NSSI2 on FAD2: p=.065 but estimate=.25). Covariate relationships in common support models were comparable.

With FAD as mediator, most adjusted mediation models failed to converge when all correlations between intercepts and slopes were modelled. However, in separate treatment models of FAD and distress (adjusted as for mediation), neither intercept was significantly correlated with its slope (estimates<.05, p>=.37), so these correlations were both removed from all the FAD mediator models. This change resulted in all but one of these models being estimated normally (Table 4.8). For all models with distress as mediator, there was a small negative residual variance for FAD slope (.02 to .06, p<.65 and >=.18). These models were estimated normally following fixing this residual at zero, which resulted in equivalent or better model fit than the original model (the difference in SS BIC between the original model versus those which fixed the residual at 0 ranged from 1.2 to 7.4).

In all cases, contemporaneous constrained and unconstrained models yielded equivalent fit. Inspection of the error covariances in the four unconstrained contemporaneous models revealed a different pattern over time in the significance of the unconstrained covariances (T1 and T2 p< 0.05, T3 p>=0.25), and therefore unconstrained was favoured to capture this theoretically plausible difference (Goldsmith *et al.*, 2018b). The only exception was the FAD mediator common support model, where the unconstrained model resulted in a negative residual, and thus the constrained model was favoured. In full propensity weighted and common support samples, FAD did not significantly mediate effects of treatment on distress (Table 4.9). Conversely, distress significantly mediated effects of treatment on FAD, 96% in the full propensity weighted sample and from 77-81% in equivalent fitting common support models (Table 4.9). In all cases, the common support sample yielded substantially better-fitting models, and is therefore considered the best estimate of indirect effects.

Table 4.8: Mediation models[a] for adolescents: model comparison for various error covariance specifications

	FAD mediator, distress outcome			Distress mediator, FAD outcome[c]	
	Original model issues	**SS BIC[b]**	**Free parameters**	**SS BIC[c]**	**Free parameters**
(A) Error covariance specification					
Full propensity weighted sample (n=625)					
No error covariances	no convergence	12994	84	12995	83
Contemporaneous unconstrained	negative residual - distress slope	**12985[d]**	87	**12982[d]**	86
Contemporaneous constrained	no convergence	12986	85	12984	84
Lagged unconstrained	no convergence	13001	86	12985	85
Lagged constrained	no convergence	12998	85	12987	84
Common support sample (n=511)					
No error covariances	no convergence	10768	84	10765	83
Contemporaneous unconstrained	negative residual - FAD slope	10761 negative residual - distress slope	87	**10757[d]**	86
Contemporaneous constrained	negative residual - FAD slope	**10761**	85	10758	84
Lagged unconstrained	no convergence	10774	86	**10757**	85
Lagged constrained	no convergence	10771	85	10760	84

[a] All propensity weighted and adjusted by T2 and T3 treatment, and time-varying covariates related to FAD and distress (friendship support, NSSI, cigarette use).
[b] Neither intercept was correlated with its slope, to resolve convergence issues.
[c] All models had a small negative residual variance for FAD slope (.02 to .06, p<.65 and >=.18) which was fixed to zero. This change resulted in model fit being comparable (SS BIC declined by 1.2 in full propensity model with no error covariances) or improving slightly (all other models SS BIC declined by 2.4 to 7.4).
[d] Favoured over constrained model as error covariances were different over time: p<.05 at T1 and T2, but p>=.25 at T3.

Table 4.9: Best-fitting mediation model[a] estimates for adolescents

	FAD mediator, distress outcome		Distress mediator, FAD outcome	
	Estimate (S.E.)	**p-value**	**Estimate (S.E.)**	**p-value**
Full propensity weighted sample (n=625)				
Treatment on mediator	-0.138 (0.042)	0.001	-0.220 (0.073)	0.003
Mediator on outcome	1.088 (0.586)	0.063	0.585 (0.159)	<0.001
Treatment on outcome (direct effect)	-0.070 (0.111)	0.53	-0.006 (0.064)	0.93
Indirect effect	-0.150 (0.090)	0.096	-0.129 (0.059)	0.028
Total effect	-0.220 (0.074)	0.003	-0.135 (0.042)	0.001
% of total effect mediated	Not applicable		**95.6%**	
Common support sample (n=511)				
Treatment on mediator	-0.147 (0.044)	0.001	-0.205 (0.070)	0.003
Mediator on outcome	2.382 (3.923)	0.54	0.572 (0.154)	<0.001
Treatment on outcome (direct effect)	0.148 (0.573)	0.80	-0.027 (0.061)	0.65
Indirect effect	-0.350 (.562)	0.53	-0.117 (0.053)	0.028
Total effect	-0.203 (.074)	0.006	-0.145 (0.042)	0.001
% of total effect mediated	Not applicable		**80.7%**	
Equivalent fitting model in common support sample	No equivalent fitting model			
Treatment on mediator			-0.206 (0.075)	0.006
Mediator on outcome			0.567 (0.150)	<0.001
Treatment on outcome (direct effect)			-0.035 (0.068)	0.61
Indirect effect			-0.117 (0.057)	0.038
Total effect			-0.152 (0.046)	0.001
% of total effect mediated			**77.0%**	

[a] All propensity weighted and adjusted by T2 and T3 treatment, and time-varying covariates related to FAD and distress (friendship support, NSSI, cigarette use). Best-fitting models from Table 4.8.

4.4 Discussion

The present study demonstrates that young people who receive community mental health treatment have reduced general distress over two years compared to those who have not received this treatment. This expands on findings in Chapter 2 (Neufeld *et al.*, 2017) by using a measure of symptoms which is transdiagnostic across mental disorders, demonstrating equivalent treatment effectiveness in emerging adults as well as adolescents, and utilizing a sample more socio-economically and ethnically representative of the general UK population (Kiddle *et al.*, 2018). I utilized propensity weighting to balance treated and untreated groups based on background, environmental, and harmful behaviour covariates, as well as baseline levels of the mediator and outcome, similar to a randomized control trial (Austin, 2011), and controlled for time-varying covariates. Further, I used latent slopes to model intraindividual change over time related to mental health treatment, FIML to deal with any bias due to missing data, and compared competing models to more clearly parameterize effects (Goldsmith *et al.*, 2018a).

4.4.1 Mental health treatment rates

The validity of the present study's self-reported rates of mental health treatment is supported by commensurate rates from NHS data. In the present study, 4.7% of adolescents aged 14-18, and 8.8% of emerging adults aged 19-24 in Greater London and Cambridgeshire reported current mental health treatment at baseline from 2012-2016 (84% reported in 2013). While regional variation complicates comparisons, NHS data from England in 2012/2013 reveals similar rates: 3.3% of 15-19-year-olds and 6.5% of 20-24-year-olds used adult IAPT and specialist mental health services (Health and Social Care Information Centre, 2014). Young people also may have received community care from CYP's mental health services. For example, 1.7% of 0-17-year-olds in London and 1.4% in East of England received such care in 2015 (NHS England, 2016). This community care could consist of services from schools, GP practices, or the voluntary sector, as well as specialist mental health services (Care Quality Commission, 2017). Treatment rates reported in the present study may also include private care or informal care not considered in the above documents. The gender difference in mental health treatment rates found in the present study (62.7% women) is also supported by NHS data. For young people using adult IAPT and specialist mental health services in England in 2012/2013, the female preponderance for those aged 15-24 was very similar at 61.3% (Health and Social Care Information Centre, 2014). The present study also showed that young women more commonly reported depression and anxiety and young men more commonly reported behavioural diagnosis, as found the most recent surveys of mental health in England (Mcmanus *et al.*, 2016; Sadler *et al.*, 2018). These concordances help bolster the accuracy of the present study's self-report data as well as the generalizability of the findings.

Self-reported treatment rates in the present study are much smaller than national rates of young people with a mental disorder, indicating barriers to care for those with mental health problems. The most recent surveys in England report that 14.4% of 11-16-year-olds have a mental disorder, and 17% of both 17-19-year-olds (Sadler *et al.*, 2018) and 16-24-year-olds have a mental disorder (Mcmanus *et al.*, 2016). Thus, treatment rates in the present sample (reported in previous paragraph) represent at most a quarter to a third of adolescents with a mental disorder and half of emerging adults with a mental disorder, based on the above national rates of mental disorder. However, the percentage of those with a mental disorder receiving treatment is likely to be even lower than these figures. While we did not ascertain mental disorder in the present study, 15% of those reporting current treatment were less distressed than the sample average, and are therefore unlikely to have a mental disorder. Regardless, these findings indicate a substantial amount of unmet need for both adolescents and emerging adults with a mental disorder.

It is notable that nearly double the rate of mental health treatment was reported by emerging adults than adolescents. Given that rates of mental disorder for both adolescents and emerging adults in England appear to be roughly comparable for the age ranges in the present study (see previous paragraph), this difference is unlikely to be based on increased need alone. Higher

treatment rates in emerging adults may indicate that this age group has better access to mental health services than adolescents. Analyses of transitions to care from CAMHS to adult services from 2005 data have indicated that transitions were poor for the vast majority of service users (Singh *et al.*, 2010). While this implies that access to services is likely to be poor in emerging adults, Singh et al's 2010 study took place prior to the launching of adult IAPT in 2008. A primary goal of adult IAPT was to initiate new services in order to increase access to care (London & South East CYP IAPT Learning Collaborative, 2016), and indeed the number of patients treated in adult IAPT has increased over time (Clark, 2018). In 2013/2014, emerging adults aged 20-24 received nearly the highest rate of IAPT services of any age group, second only to those aged 25-29 (Health and Social Care Information Centre, 2014). CYP IAPT has been focused on improving the quality of existing services as opposed to providing more services (London & South East CYP IAPT Learning Collaborative, 2016). However, access to CYP's mental health services has been found to be limited, based on a fragmented system of care drawn from multiple sources which may struggle for funding (Care Quality Commission, 2017). Thus, the present study's lower rates of reported mental health treatment in adolescents may reflect relatively increased barriers to care for adolescents compared with emerging adults. These findings therefore add to other exhortations highlighting the need for improved access to mental healthcare in adolescents (Department of Health and Department for Education, 2017).

4.4.2 Mental health treatment effectiveness

The present study demonstrates how utilizing those above the population mean on distress can be an effective way of parameterizing treatment need in order to determine treatment effectiveness. This cut-off allowed those with and without current mental health treatment to be more comparable than the full sample on baseline variables, enabling balancing of these variables using the propensity weight. These participants were more deprived, socially disadvantaged, and more likely to engage in harmful behaviours than those with lower distress scores. Further, they were more likely to report previous mental health problems, and treatment rates in this subsample were doubled (8% and 15% for adolescents and emerging adults) and thus much closer to national rates of mental disorders. Half of the sample was included in the propensity-weighted treatment analysis in the present study, whereas only 11% of the measured population was included in a comparable analysis when treatment need was defined by presence of a mental disorder (Neufeld *et al.*, 2017). While it is possible that the present study included some individuals subthreshold for a mental disorder in the treatment estimate, this is unknown since we did not also diagnose presence or absence of mental disorder. Future studies should look to estimate a treatment effect in prediagnosable individuals, as such a preventative approach "…represents the frontier of research and service reform in mental healthcare" (McGorry, 2013).

Treatment effects on measures of perceived social functioning operated as anticipated. Based on differences in treatment focus in CAMHS and adult services (Wolpert *et al.*, 2016; Baker, 2018; Clark, 2018) as well as the changing nature of family and friendships over adolescence and emerging adulthood (Larson and Richards, 1991; Larson *et al.*, 1996; Arnett, 2000), I assessed effects in adolescents and emerging adults separately. Importantly, I implemented a propensity weight tailored to each subgroup, which adjusted for baseline levels of family functioning (and other covariates on Table 4.3). Following propensity weighting, the standardized differences in baseline family functioning between the treated and untreated groups was .03 or less. This therefore rules out the notion that more functional families were more likely to take up treatment. Following propensity weighting in each age group, treatment was related to improved family functioning over two years, but only in the adolescent sample. As family needs of young people are often not met in CAMHS (McGorry, Bates and Birchwood, 2013), it is heartening that treatment effects on family functioning were seen in those who were most likely to be accessing CAMHS. Services therefore appear to be meeting more holistic needs of adolescent patients. However, social functioning is also a measured outcome in adult mental healthcare (Clark, 2018). Lack of treatment effects on either measure of social functioning in the emerging adults subsample indicates that treatment may need to more specifically focus on this aspect of recovery. Prior work has identified that in the third decade of life, young emerging adults (age 21) report the lowest levels of perceived social support, which increases until the age of 30 (Pettit *et al.*, 2011). Emerging adults often experience a changing social environment, and romantic relationships become relatively more important, yet are by no means stabilized (Arnett, 2000). Therefore, families may still be the most important source of support during these transitory years. More focused study should be performed to better understand how treatment can improve social functioning for emerging adults with mental health needs.

4.4.3 Mediation

Mediation models in the present study indicate that in adolescents, distress mediates treatment effects on perceived family functioning. The common support sample, trimmed of those with extreme propensity weights (Heckman *et al.*, 1996), yielded substantially better fit than the full propensity models. Two common support models (with lagged or contemporaneous unconstrained error covariances) yielded equivalent fit (Table 4.8). These models revealed that improvements in distress mediated 77-81% of the treatment effects on family functioning (Table 4.9). As treatment was related to both mediator and outcome in two-year growth curves only, these effects were not temporally offset. However, directionality of effects is supported by the improved model fit observed when distress (as opposed to family functioning) was modelled as the mediator, in full propensity and common support models. This follows from Goldsmith et al's exhortation that competing mediation models should be compared and the best-fitting model favoured (Goldsmith *et al.*, 2018a). Directionality of effects would be more strongly demonstrated if data existed for an

additional timepoint, so that growth curves could be offset in time. Nonetheless, findings indicate that treatment effects on distress drive improvements in family functioning. This directionality is supported by a review of longitudinal models of psychosocial development in adolescence (Meeus, 2016). Studies reviewed consistently observed that adolescent psychopathology led to the erosion of relationships with parents, but findings in the opposite direction were less systematic (Meeus, 2016). Further, family functioning is influenced by each individual in the family. Such complexity no doubt contributes to the challenge of treatment altering family functioning in a way that improves young people's distress. Supporting this notion is the evidence that family therapy is not related to improved depression (NICE, 2019b).

Confidence in the findings from the mediation model is bolstered by the methodological rigour used in the present study. Confounding was robustly addressed for all pathways in mediation models as recommended (VanderWeele, 2016). Propensity weighting equalized the treated and untreated groups by baseline variables (Austin, 2011), thereby addressing confounding of the treatment-mediator and treatment-outcome paths. Time-varying covariates related to the mediator and outcome adjusted for potential confounding of the mediator-outcome path. Finally, the most optimal way to model unmeasured confounding was ascertained by assessing the relative fit of models when contemporaneous and lagged constrained and unconstrained error covariances were included (Goldsmith *et al.*, 2018b).

The large amount of total effects explained by the mediator could be due to both distress and family functioning being self-reported. The inclusion of the error covariances between these measures helps address this shared method variance (Goldsmith *et al.*, 2018b). Nonetheless, it is possible that mediation effects would be reduced if family functioning was assessed by another person than the participant, for example using the FAD-GF's comparable clinical rating scale or structured interview (Miller *et al.*, 2000). However, these are more time-intensive approaches and therefore unlikely to be used in the large studies required for analyses such as those performed in the present study. A parent report of the FAD-GF could instead be used; however, parents rate family functioning as being significantly better than adolescents using this scale (Bagley *et al.*, 2001). To my awareness, the comparative validity of the parent versus child report of FAD-GF has not been ascertained at any age of child, and thus a parent report may not provide a more accurate measurement of family functioning. While triangulated evidence from both sources may be illuminating, most parents were not involved in the present study, so their inclusion would still represent a practical challenge. Nonetheless, inclusion of an additional or alternative rater of family functioning could be an avenue for future research.

4.4.4 Treatment effects over time

The present study's findings demonstrate the importance of studying effects of treatment-as-usual over a longer period. Latent growth models revealed that treatment was effective over the

course of two years but latent change score models indicated that any treatment-related improvement was not significant during the first year. These differential findings also demonstrate the importance of utilizing various approaches to measuring change over time. The long-term effectiveness of mental health treatments is rarely studied, but the importance of collecting this data is now acknowledged (Goodyer *et al.*, 2017). Similar to the present study, Chapter 2 showed how treatment effects in adolescents were most apparent one-to-two years following service contact (Neufeld *et al.*, 2017).

These findings could be interpreted in several ways. Therapy may be teaching young people skills which they continue to apply and hone following the cessation of treatment, resulting in improved outcomes later in time (Goldsmith *et al.*, 2018a). This notion is supported by recent findings that following 6-7 months of therapy, depressive symptoms continued to decline over a year after start of treatment for the majority of adolescent patients (Davies *et al.*, 2019). Alternatively, findings could support the necessity of a longer duration of treatment or multiple courses of treatment for effectiveness to be demonstrated. Number of treatment sessions has indeed been shown to be positively related to treatment response (Clark *et al.*, 2018). While the present study did not measure length of treatment participants had engaged in at baseline or over the two-year follow-up, effects increased when subsequent treatment was included in models, indicating that continued treatment played a role in the findings. Finally, a delayed start of treatment may have contributed to the observed apparent delay in treatment effects. While it is most likely that those who endorsed current treatment were actively in therapy or taking psychotrophic medication, this self-report data may reflect participants being accepted into treatment but waiting for it to commence. Time from referral to start of treatment is often substantially delayed (Department of Health and Department for Education, 2017), so this in itself could contribute to effects being apparent well after current treatment was first reported. These possibilities speak to the importance of acquiring sufficient detail on mental health treatment in community-based studies (eg. referral date, commencement of treatment, and length of treatment), in order to refine conclusions.

4.4.5 Limitations and Conclusions

There are some final caveats to note about the present study. Firstly, despite a marked female preponderance of common mental disorders in both adolescents and emerging adults (Mcmanus *et al.*, 2016; Sadler *et al.*, 2018), as well as gender differences between social support factors and mental health (Pettit *et al.*, 2011; Finan, Ohannessian and Gordon, 2018), the present study did not assess treatment effects by gender in addition to age effects. This is primarily because sample size limited further subdivision of the adolescent and emerging adult subsamples; however, I am unaware of data to support differential treatment effects by gender (eg. NICE reviews). Secondly, propensity weighting balances the treated and untreated groups based on measured variables, but there are other covariates which were not measured that may also differentiate these

groups. For example, presence of parental psychiatric disorder is a known correlate of psychopathology in youth (Nomura *et al.*, 2002) which was not measured in the NSPN study. Nonetheless, many important factors relevant to mental health were measured, allowing for propensity weighting to cover an array of variables from many domains. Finally, the primary outcome measure of distress is advantageous in measuring change across multiple disorders, being transdiagnostic across depressive, anxiety, behavioural, obsessive, and psychotic symptoms and inclusive of well-being and self-esteem. However, such a measure has no benchmark for reliable improvement or reliable recovery, key outcome metrics for IAPT (Wolpert *et al.*, 2016; Clark, 2018). Similarly, no such benchmarks exist for the social functioning measures. The present findings are therefore limited to showing statistically significant improvement.

In sum, the present study adds to a dearth of studies on this effectiveness of mental health services in the community. Robustly addressing confounding, it relates self-reported mental health treatment across multiple disorders to a two-year decline in distress common across symptoms of these disorders, in both adolescents and emerging adults. Findings reveal that this decreased distress is a mechanism by which treatment acts to improve perceived family functioning in adolescents. Thus, for adolescents who access services, the positive benefit of mental health treatment appears to be more holistic than improved symptoms alone. While emerging adults may be particularly vulnerable to loss of care when transitioning into adult services (McGorry, Bates and Birchwood, 2013), the present findings reveal that adolescents report less access to care than emerging adults. Self-reported treatment rates in the present study are much lower than national rates of mental disorders, highlighting the ongoing need of increased access to mental health services.

CHAPTER FIVE

Conclusion and future directions

The present body of work is important principally in supporting the notion that mental health services do appear to effectively reduce subsequent symptoms of mental illness in young people to a greater extent than natural improvement without intervention. This was able to be demonstrated in two cohorts (ROOTS: (Goodyer *et al.*, 2010); NSPN: (Kiddle *et al.*, 2018)), in adolescents and emerging adults, using both specific (ie: depression) and transdiagnostic (ie: *p*) outcomes. These findings, in conjunction with the treatment-related gains in family functioning reported in Chapter 4, highlight the importance of young people's increased access to mental health services. Findings of both studies are bolstered by the use of advanced statistical methods to account for attrition and non-randomization of service usage in these community studies. Some themes emerge across this body of work which bear further discussion.

5.1 Measuring mental health service contact

I would first like to make a point about the measurement of mental health services in epidemiological samples. Both studies in Chapters 2 and 4 utilized self-report data to measure service contact. While this is a limitation, to date, there have been few studies comparing those with and without mental health service contact which have validated reported service contacts against clinical records. The Child and Adolescent Service Use Schedule (CA-SUS), originally developed for a randomized controlled trial (RCT; (Byford *et al.*, 1999)), has been used extensively as part of cost-analyses in subsequent RCTs of children and young people with a variety of mental health conditions (Barrett *et al.*, 2006; Byford *et al.*, 2007, 2015; Green *et al.*, 2011; Dalgleish *et al.*, 2015; Goodyer *et al.*, 2017; Kuyken *et al.*, 2017; Ford *et al.*, 2019). The CA-SUS obtains participant responses regarding number and duration of contacts with various services and professionals. However, none of these trials (or other studies to my knowledge) have assessed the CA-SUS for concordance with clinical records. Thus, self-report data used to compare the economic consequences of various interventions has not been externally validated. Further, the UK government's most recent child and adolescent mental health survey used a self-report measure of mental health service contact which has not been externally validated (Sadler *et al.*, 2018). The Services Assessment for Children and Adolescents (SACA (Hoagwood *et al.*, 2000)), which has been used in several of the studies reviewed in Chapter 2 (supplementary Table 1, Appendix A (Neufeld *et al.*, 2017)), has been assessed for external validity. A global "any use" variable from the SACA reported by primary caregivers across inpatient, outpatient, and school mental health services had excellent agreement with records (Hoagwood *et al.*, 2000). This more general variable,

similar to the ones used for analyses in Chapters 2 and 4, had a higher concordance than more specific reports of service use. The use of such variables is further bolstered by the high concordance between adolescent and caregiver reports of mental health service contact revealed in Chapter 2 (Neufeld *et al.*, 2017).

It is noteworthy to comment on the simplicity of the service use measures used, for consideration to be included in future assessments of mental health, whether for research or clinical purposes. In both chapters, mental health service contact was related to improved outcomes irrespective of disorder type, sector in which contact was made, or treatment length. In Chapter 4, mental health service contact was simply based on a question which asked whether the participant was "…currently being treated for any emotional, behavioural or mental health problem". Yet rates determined from this question were comparable to service usage reported by NHS England during a similar timeframe for comparable ages to the ones studied. While the lack of detail in such a measure could rightly be seen as a limitation, this simplicity could enable the assessment of mental health treatment use in studies which may not be able to ask more detailed questions due to participant burden. If such a measure were included alongside future mental health assessments (self-report or otherwise), more datasets could potentially be used to test effectiveness of mental health treatment-as-usual. This could help home in on regions of the country where services may not be as effective, for example. It is noteworthy that this simple measure yielded more comparable rates to NHS data than the measure used in the most recent governmental survey on mental health in children and young people (Sadler *et al.*, 2018). As in previous British surveys, this measure more generically assessed contact with professional services for a mental health reason. This included contact with teachers, and thus reported service contacts are much higher than those determined using more specific definitions (discussed in Chapter 2, (Neufeld *et al.*, 2017)). The measure in chapter 4 implies specificity to health professionals by being limited to treatment as opposed to more generic help-seeking. This balance of simplicity with specificity is an important distinction to make.

However, accurate measurement of mental health service contact in young people is still generally stymied by the pediatric-adult divide in mental health services. Chapter 4 importantly focuses on the age range from 14-24, encompassing the developmentally sensitive period of adolescence and emerging adulthood (Kessler *et al.*, 2007; Arnett, Žukauskiene and Sugimura, 2014; Giedd *et al.*, 2015). However, even comparing rates of mental health service usage reported in this chapter to national data was hampered by the artificial split between child and adult services. NHS publications report treatment for 15-17-year-olds in both adult IAPT (Health and Social Care Information Centre, 2014) and CAMHS, but CAMHS acknowledges that treatment may be transitional and/or up until the age of 25 (NHS England, 2016). National mental health survey estimates for adults include those aged 16 and above (Mcmanus *et al.*, 2016), whereas the comparable survey for children and young people only extends to age 19 (Sadler *et al.*, 2018).

Other countries' statistics are similarly hampered by a separate child and adult service model (eg. USA: (Kessler, Chiu, *et al.*, 2005; Merikangas *et al.*, 2010); Australia: (Korten and Henderson, 2000; Lawrence *et al.*, 2015)). If we are intent on making serious gains in the mental health of young people – considering both adolescents and emerging adults to be populations of great risk but high opportunity for intervention – it is vital that the same statistics be obtained continuously, for the whole age range from age 14 to at least 25. Only then can we gain an accurate picture of the scope of young people's mental health needs and service utilization during this sensitive period.

Considering the worse rates of mental health service contact for CAMHS compared with adult services discussed in Chapter 1, it is unsurprising that Chapter 4 data mirrors these findings. Treatment rates for adolescents (5%) and emerging adults (9%) were estimated to represent at most a quarter to a third of adolescents with a mental disorder and half of emerging adults with a mental disorder, based on national rates. This underscores the need for serious reform of both child and adult systems, but in particular, adolescent mental health services. Such reform is especially urgent in light of the lack of parity in economic outlay for CAMHS compared with adult mental health services (Children's Commissioner, 2018).

5.2 The structure of young people's mental health

How do the findings in Chapter 3 expand our understanding of the structure of mental health during this developmentally sensitive period? What is remarkable about these findings is the consistency they have with similar models from other studies which may or may not have covered at least part of the age range studied of 14-24 (for review see (Caspi and Moffitt, 2018)). This observation supports the validity of such an approach to understanding mental health across the lifespan. However, compared to previous studies, Chapter 3 makes a tighter argument about the validity and reliability of such a model over time, using in-depth statistical methodologies (measurement invariance testing, ascertaining construct replicability) as well as a bevy of external validation measures, and accelerated growth modelling never previously employed on such a model. This gives more support to the perhaps bold assertion that not only is the general transdiagnostic factor (*p)* meaningful, but the specific factors also appear to be more than mere residual error that some contend (Chen, West and Sousa, 2006). Such a model therefore has the potential to lead us out of the problem of comorbidity more equipped to understand the interaction of mental domains and how this may instigate or propagate impairment, in conjunction with other risk factors.

Future studies of bifactor models could give greater attention to their unwieldiness. If one wants to comprehensively study specific domains of mental health simultaneously, one must thoroughly assess the underlying thoughts, emotions, and/or behaviours of each so that adequate construct replicability is achieved. At the same time, one does not wish to incur too great a burden upon participants in terms of time spent completing a lengthy battery of questionnaires. Therefore,

moving forward, it is important that item-reduction techniques such as regularized structural equation modelling (Jacobucci, Brandmaier and Kievit, 2019) be employed when generating bifactor models, to create a more parsimonious model whilst retaining the original constructs. This could first be performed on existing datasets of previously published symptom-level bifactor models, such as the one presented in Chapter 3 (St Clair *et al.*, 2017). Such analyses could inform a more focussed approach to data collection for future bifactor models, reducing participant burden whilst ensuring appropriate coverage of specific syndromes.

Regardless, the findings in Chapter 3 highlight protective and risk factors for *p* and the specific factors more strongly than previous studies. In general, the external validation measures performed longitudinally as anticipated based on the predominantly cross-sectional associations previously reported. Further, there were many areas where the present findings are the first to report external validation. Specifically, this is the first study to use ethnicity, stressful life events, and parenting styles to validate a bifactor model, and the first to externally validate a specific positive mental health factor. Further, Chapter 3 substantially advances the external validation of a thought disorder specific factor across all domains (background, social environment, personality, and IQ). The multivariate model reveals in a cross-domain way not previously ascertained how aspects of social environment and personality explain equivalent amounts of variance in *p*, whilst the specific factors all have distinctly different predictive profiles. Finally, the accelerated growth models reveal trends of increasing and declining ill health from ages 14-27. The anxiety specific factor emerged as the factor of greatest increasing risk during this age range, followed by increased rates in *p*. The antisocial specific factor peaked in the mid-teens and declined thereafter, whilst the thought disorder specific factor declined over the full age range. These models could help target developmental phases for intervention – for example, that increasing efforts must be expended to reduce distress and anxiety in emerging adults, but that antisocial behaviour needs more attention in adolescence. Future work could look to discern predictors of change in these growth curves. Additionally, these models should be extended to both younger and older age ranges, so a more complete picture of mental health can be obtained across the lifespan.

5.3 Final statements

Internationally, mental health problems contribute substantial burden, a burden which is substantially greater for adolescents and emerging adults (Kyu *et al.*, 2018). The present body of work suggests that we can alleviate some of this burden by increasing young people's contact with mental health services. While these studies cannot prove this assertion, since they are not randomized controlled trials, they represent some of the best evidence to support this assertion to date. By using community level data, a non-treated comparison group is possible. Such a naturalistic setting gives real-world impact to the beguiling findings from randomized control trials, obtained under ideal conditions. Clearly there is more work to be done. Putative mechanisms of

treatment action on symptoms have yet to be discerned. Subsample analyses of various conditions or various sectors of treatment could be attempted if larger samples were available. Such understanding would help focus efforts and potentially be cost saving. Nonetheless, the present findings give impetus to increase funding for young people's mental health services. Access to effective treatments must be improved. If young people's mental health is prioritized, the gains could be substantial.

REFERENCES

Abad, F. J. *et al.* (2017) 'Iteration of Partially Specified Target Matrices: Application to the Bi-Factor Case', *Multivariate Behavioral Research*. Taylor & Francis, 52(4), pp. 416–429. doi: 10.1080/00273171.2017.1301244.

Achenbach, T. M. (1966) 'The classification of children's psychiatric symptoms: A factor-analytic study', *Psychological Monographs: General and Applied*, 80(7), pp. 1–37.

Afzali, M. H. *et al.* (2018) 'The Structure of Psychopathology in Early Adolescence: Study of a Canadian Sample', *The Canadian Journal of Psychiatry*, 63(4), pp. 223–230. doi: 10.1177/0706743717737032.

Aitken, M. *et al.* (2020) 'Toward precision therapeutics: general and specific factors differentiate symptom change in depressed adolescents', *Journal of Child Psychology and Psychiatry and Allied Disciplines*. doi: 10.1111/jcpp.13194.

Allison, P. D. (2003) 'Missing Data Techniques for Structural Equation Modeling', *Journal of Abnormal Psychology*, 112(4), pp. 545–557. doi: 10.1037/0021-843X.112.4.545.

Altman, D. G. and Bland, M. J. (1999) 'Treatment allocation in controlled trials: Why randomise?', *BMJ*, 318(7192), p. 1209. doi: 10.1136/bmj.318.7192.1209.

American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders*. IV. Washington, DC: American Psychiatric Association.

Amone-P'Olak, K. *et al.* (2009) 'Life Stressors as Mediators of the Relation Between Socioeconomic Position and Mental Health Problems in Early Adolescence: The TRAILS Study', *Journal of the American Academy of Child and Adolescent Psychiatry*. American Academy of Child & Adolescent Psychiatry, 48(10), pp. 1031–1038. doi: 10.1097/CHI.0b013e3181b39595.

Anderson, K. K. (2019) 'Towards a public health approach to psychotic disorders', *The Lancet Public Health*, 4(5), pp. e212–e213. doi: 10.1016/S2468-2667(19)30054-4.

Anttila, V. *et al.* (2018) 'Analysis of shared heritability in common disorders of the brain', *Science*, 360(6395), p. eaap8757. doi: 10.1126/science.aap8757.

Arnett, J. J. (2000) 'Emerging adulthood: A theory of development from the late teens through the twenties', *American Psychologist*, 55(5), pp. 469–480. doi: 10.1037/0003-066X.55.5.469.

Arnett, J. J. and Schwab, J. (2012) *The Clark University Poll of Emerging Adults: Thriving, struggling, & hopeful*. Worcester, Mass. Available at: http://www.clarku.edu/clarkpoll/pdfs/Clark_Poll_Peer Inst.pdf.

Arnett, J. J., Žukauskiene, R. and Sugimura, K. (2014) 'The new life stage of emerging adulthood at ages 18-29 years: Implications for mental health', *The Lancet Psychiatry*, 1(7), pp. 569–576. doi: 10.1016/S2215-0366(14)00080-7.

Arseneault, L. *et al.* (2011) 'Childhood trauma and children' s emerging psychotic symptoms : A genetically sensitive longitudinal cohort study', *American Journal of Psychiatry*, 168, pp. 65–72. doi: 10.1176/appi.ajp.2010.10040567.

Asparouhov, T. and Muthén, B. (2010) 'Multiple imputation with Mplus', *MPlus Web Notes*. Available at: http://statmodel2.com/download/Imputations7.pdf.

Asselmann, E. *et al.* (2015) 'Danger and loss events and the incidence of anxiety and depressive disorders: A prospective-longitudinal community study of adolescents and young adults', *Psychological Medicine*, 45(1), pp. 153–163. doi: 10.1017/S0033291714001160.

Austin, P. C. (2011) 'An introduction to propensity score methods for reducing the effects of confounding in observational studies.', *Multivariate behavioral research*, 46(3), pp. 399–424. doi: 10.1080/00273171.2011.568786.

Bagley, C. *et al.* (2001) 'Discrepant parent-adolescent views on family functioning: Predictors of poorer self-esteem and problems of emotion and behaviour in British and Canadian adolescents', *Journal of Comparative Family Studies*, 32(3), pp. 393–403.

Baker, C. (2018) *Mental health statistics for England: prevelance, services and funding*, *House of Commons Library*. London, UK. doi: 10.1117/12.511423.

Barrett, B. *et al.* (2006) 'Mental health provision for young offenders: Service use and cost', *British Journal of Psychiatry*, pp. 541–546. doi: 10.1192/bjp.bp.105.010108.

Beauducel, A. and Herzberg, P. Y. (2009) 'On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA', *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), pp. 186–203. doi: 10.1207/s15328007sem1302.

Bengtson, V. L. (2001) 'Beyond the nuclear family: The increasing importance of multigenerational bonds', *Journal of Marriage and Family*, 63(1), pp. 1–16. doi: 10.1111/j.1741-3737.2001.00001.x.

Bentler, P. M. (1990) 'Comparative fit indexes in structural models', *Psychological Bulletin*, 107(2), pp. 238–246. doi: 10.1037/0033-2909.107.2.238.

Bentler, P. M. (1992) 'On the fit of models to covariances and methodology to the Bulletin', *Psychological Bulletin*, 112(3), pp. 400–404. doi: 10.1037/0033-2909.112.3.400.

Biddle, B. J., Bank, B. J. and Marlin, M. M. (1980) 'Parental and Peer Influence on Adolescents Stable', *Social Forces*, 58(4), pp. 1057–1079.

Björkenstam, E. *et al.* (2016) 'Childhood adversity and psychiatric disorder in young adulthood: An analysis of 107,704 Swedes', *Journal of Psychiatric Research*, 77, pp. 67–75. doi: 10.1016/j.jpsychires.2016.02.018.

Böhnke, J. R. and Croudace, T. J. (2016) 'Calibrating well-being, quality of life and common mental disorder items: Psychometric epidemiology in public mental health research', *British Journal of Psychiatry*, 209(2), pp. 162–168. doi: 10.1192/bjp.bp.115.165530.

Bollini, P. *et al.* (1999) 'Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials', *British Journal of Psychiatry*, 174, pp. 297–303. doi: 10.1192/bjp.174.4.297.

Bonifay, W., Lane, S. P. and Reise, S. P. (2017) 'Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology', *Clinical Psychological Science*, 5(1), pp. 184–186. doi: 10.1177/2167702616657069.

Bowen, N. K. and Masa, R. D. (2015) 'Conducting Measurement Invariance Tests with Ordinal Data: A Guide for Social Work Researchers', *Journal of the Society for Social Work and Research*, 6(2), pp. 229–249. doi: 10.1086/681607.

Brittain, P. J. *et al.* (2013) 'Harnessing Clinical Psychiatric Data with an Electronic Assessment Tool (OPCRIT+): The Utility of Symptom Dimensions', *PLoS ONE*, 8(3), pp. 1–8. doi: 10.1371/journal.pone.0058790.

Brodbeck, J. *et al.* (2011) 'General and specific components of depression and anxiety in an adolescent population', *BMC psychiatry*. BioMed Central Ltd, 11(191). doi: 10.1186/1471-244X-11-191.

Brodbeck, J. *et al.* (2013) 'Comparing growth trajectories of risk behaviors from late adolescence through young adulthood: An accelerated design', *Developmental Psychology*, 49(9), pp. 1732–1738. doi: 10.1037/a0030873.

Brodbeck, J. *et al.* (2014) 'General distress, hopelessness - Suicidal ideation and worrying in adolescence: Concurrent and predictive validity of a symptom-level bifactor model for clinical diagnoses', *Journal of Affective Disorders*. Elsevier, 152–154(1), pp. 299–305. doi: 10.1016/j.jad.2013.09.029.

Brookhart, M. A. *et al.* (2006) 'Variable selection for propensity score models', *American Journal of*

Epidemiology, 163(12), pp. 1149–1156. doi: 10.1093/aje/kwj149.

Brown, T. A. (2006) *Confirmatory factor analysis for applied research*. New York, NY, US: The Guildford Press.

Bruce, S. E. *et al.* (2005) 'Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-year prospective study', *American Journal of Psychiatry*, 162(6), pp. 1179–1187. doi: 10.1176/appi.ajp.162.6.1179.

Burke, J. D., Loeber, R. and Birmaher, B. (2002) 'Oppositional Defiant Disorder and Conduct Disorder: A review of the past 10 years, part II', *Journal of the American Academy of Child and Adolescent Psychiatry*. The American Academy of Child and Adolescent Psychiatry, 41(11), pp. 1275–1293. doi: 10.1097/00004583-200211000-00009.

Burnam, M. A. (1996) 'Measuring outcomes of care for substance use and mental disorders.', *New directions for mental health services*, 71, pp. 3–17.

Byford, S. *et al.* (1999) 'Cost-effectiveness analysis of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves: Results of a randomised controlled trial', *British Journal of Psychiatry*, 174, pp. 56–62.

Byford, S. *et al.* (2007) 'Cost-effectiveness of selective serotonin reuptake inhibitors and routine specialist care with and without cognitive-behavioural therapy in adolescents with major depression', *British Journal of Psychiatry*, 191, pp. 521–527. doi: 10.1192/bjp.bp.107.038984.

Byford, S. *et al.* (2015) 'Cost-effectiveness analysis of a communication-focused therapy for pre-school children with autism: Results from a randomised controlled trial', *BMC Psychiatry*. BMC Psychiatry, 15(1), pp. 1–13. doi: 10.1186/s12888-015-0700-x.

Canino, G. *et al.* (2004) 'The DSM-IV rates of child and adolescent disorders in Puerto Rico', *Archives of general psychiatry*, 61, pp. 85–93.

Care Quality Commission (2017) *Review of children and young people' s mental health services*. Citygate, Gallowgate, Newcastle upon Tyne, UK. Available at: https://www.cqc.org.uk/sites/default/files/20171103_cypmhphase1_report.pdf.

Carragher, N. *et al.* (2016) 'The structure of adolescent psychopathology: A symptom-level analysis', *Psychological Medicine*, 46(5), pp. 981–994. doi: 10.1017/S0033291715002470.

Caspi, A. *et al.* (2014) 'The p factor: One general psychopathology factor in the structure of psychiatric disorders?', *Clinical Psychological Science*, 2(2), pp. 119–137. doi: 10.1177/2167702613497473.

Caspi, A. and Moffitt, T. E. (2018) 'All for One and One for All: Mental Disorders in One Dimension', *American Journal of Psychiatry*, 175(9), pp. 831–844. doi: 10.1176/appi.ajp.2018.17121383.

Cassels, M. *et al.* (2018) 'Poor family functioning mediates the link between childhood adversity and adolescent nonsuicidal self-injury', *Journal of Child Psychology and Psychiatry and Allied Disciplines*. doi: 10.1111/jcpp.12866.

Castellanos-Ryan, N. *et al.* (2013) 'Sensitivity and Specificity of a Brief Personality Screening Instrument in Predicting Future Substance Use, Emotional, and Behavioral Problems: 18-Month Predictive Validity of the Substance Use Risk Profile Scale', *Alcoholism: Clinical and Experimental Research*, 37(SUPPL.1), pp. 281–290. doi: 10.1111/j.1530-0277.2012.01931.x.

Castellanos-Ryan, N. *et al.* (2016) 'The structure of psychopathology in adolescence and its common personality and cognitive correlates', *Journal of Abnormal Psychology*, 125(8), pp. 1039–1052. doi: 10.1037/abn0000193.

Castro, L. D. *et al.* (2018) 'Gender Differences within the Psychosis Spectrum', 8, pp. 1186–1195.

Chen, F. F. (2007) 'Sensitivity of goodness of fit indexes to lack of measurement invariance', *Structural Equation Modeling*, 14(3), pp. 464–504. doi: 10.1080/10705510701301834.

Chen, F. F. (2008) ‘What Happens If We Compare Chopsticks With Forks? The Impact of Making Inappropriate Comparisons in Cross-Cultural Research’, *Journal of Personality and Social Psychology*, 95(5), pp. 1005–1018. doi: 10.1037/a0013193.

Chen, F. F., West, S. G. and Sousa, K. H. (2006) ‘A comparison of bifactor and second-order models of quality of life’, *Multivariate Behavioral Research*, 41(2), pp. 189–225. doi: 10.1207/s15327906mbr4102_5.

Chen, P. Y. *et al.* (2019) ‘Testing Measurement Invariance with Ordinal Missing Data: A Comparison of Estimators and Missing Data Techniques’, *Multivariate Behavioral Research*. Routledge. doi: 10.1080/00273171.2019.1608799.

Cheung, G. W. and Rensvold, R. B. (2002) ‘Evaluating Goodness-of- Fit Indexes for Testing Measurement Invariance’, *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), pp. 233–255. doi: 10.1207/S15328007SEM0902.

Children’s Commissioner (2018) *Children’s mental health briefing*. London, UK. Available at: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2019/02/childrens-mental-health-briefing-nov-2018.pdf.

Clark, D. M. (2018) ‘Realising the Mass Public Benefit of Evidence-Based Psychological Therapies : The IAPT Program.’, *Annual Review of Clinical Psychology*, 7(14), pp. 159–183. doi: 10.1146/annurev-clinpsy-050817-084833.Realising.

Clark, D. M. *et al.* (2018) ‘Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data’, *The Lancet*, 391(10121), pp. 679–686. doi: 10.1016/S0140-6736(17)32133-5.

Clark, L. A. *et al.* (2017) ‘Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC)’, *Psychological Science in the Public Interest*, 18(2), pp. 72–145. doi: 10.1177/1529100617727266.

Clarke, A. *et al.* (2011) ‘Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Validated for teenage school students in England and Scotland. A mixed methods assessment’, *BMC Public Health*, 11, pp. 1–9. doi: 10.1186/1471-2458-11-487.

Class, Q. A. *et al.* (2019) ‘Socioemotional dispositions of children and adolescents predict general and specific second-order factors of psychopathology in early adulthood: A 12-year prospective study’, *Journal of Abnormal Psychology*, 128(6), pp. 574–584. doi: 10.1037/abn0000433.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

Cohen, S. and Wills, T. A. (1985) ‘Stress, Social Support, and the Buffering Hypothesis’, *Psychological Bulletin*, 98(2), pp. 310–357. doi: 10.1037/0033-2909.98.2.310.

Collins, L. M., Schafer, J. L. and Kam, C.-M. (2001) ‘A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures’, *Psychological Methods*, 6(4), pp. 330–351. doi: 10.1037//1082-989X.6.4.330.

Colman, I. *et al.* (2007) ‘Forty-year psychiatric outcomes following assessment for internalizing disorder in adolescence’, *American Journal of Psychiatry*, 164(1), pp. 126–133. doi: 10.1176/ajp.2007.164.1.126.

Constantinou, M. P. *et al.* (2019) ‘Changes in General and Specific Psychopathology Factors Over a Psychosocial Intervention’, *Journal of the American Academy of Child and Adolescent Psychiatry*. Elsevier Inc, 58(8), pp. 776–786. doi: 10.1016/j.jaac.2018.11.011.

Costello, E. J. *et al.* (2014) ‘Services for adolescents with psychiatric disorders: 12-month data from the national comorbidity survey-adolescent.’, *Psychiatric services*, 65(3), pp. 359–66. doi: 10.1176/appi.ps.201100518.

Costello, E. J. and Angold, A. (1988) ‘Scales to assess child and adolescent depression –

checklists, screens, and nets', *Journal of the American Academy of Child Psychiatry*, 27, pp. 726–737.

Crush, E. *et al.* (2019) 'Using discordant twin methods to investigate an environmentally mediated pathway between social support and the reduced likelihood of adolescent psychotic experiences', *Psychological Medicine*, pp. 1–8. doi: 10.1017/s0033291719001983.

Dalgleish, T. *et al.* (2015) 'Trauma-focused cognitive behaviour therapy versus treatment as usual for post traumatic stress disorder (PTSD) in young children aged 3 to 8 years: Study protocol for a randomised controlled trial', *Trials*, 16(1). doi: 10.1186/s13063-015-0632-2.

Davies, S. *et al.* (2019) 'Trajectories Of Depression Symptom Change During And Following Treatment In Adolescents With Unipolar Major Depression', *Journal of the American Academy of Child and Adolescent Psychiatry*, in press.

Department of Health and Department for Education (2017) *Transforming Children and Young People's Mental Health Provision: a Green Paper*. London, UK. doi: 979-1-5286-0061-3.

Deutz, M. H. F. *et al.* (2016) 'The Dysregulation Profile in middle childhood and adolescence across reporters: factor structure, measurement invariance, and links with self-harm and suicidal ideation', *European Child and Adolescent Psychiatry*, 25(4), pp. 431–442. doi: 10.1007/s00787-015-0745-x.

Deutz, M. H. F. *et al.* (2018) 'Evaluation of the Strengths and Difficulties Questionnaire-Dysregulation Profile (SDQ-DP)', *Psychological Assessment*, 30(9), pp. 1174–1185. doi: 10.1037/pas0000564.

Dowrick, C. *et al.* (2000) 'Problem solving treatment and group psychoeducation for depression: Multicentre randomised controlled trial', *British Medical Journal*, 321(7274), pp. 1450–1454. doi: 10.1136/bmj.321.7274.1450.

Duggan, C. *et al.* (2014) 'The recording of adverse events from psychological treatments in clinical trials: Evidence from a review of NIHR-funded trials', *Trials*, 15(335). doi: 10.1186/1745-6215-15-335.

Duncan, T. E. and Duncan, S. C. (2004) 'An introduction to latent growth curve modeling', *Behavior Therapy*, 35(2), pp. 333–363. doi: 10.1016/S0005-7894(04)80042-X.

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

Eberhart, N. K. and Hammen, C. L. (2006) 'Interpersonal predictors of onset of depression during the transition to adulthood', *Personal Relationships*, pp. 195–206. doi: 10.1111/j.1475-6811.2006.00113.x.

Edbrooke-Childs, J. *et al.* (2018) 'Evaluation of reliable improvement rates in depression and anxiety at the end of treatment in adolescents', *BJPsych Open*, 4(4), pp. 250–255. doi: 10.1192/bjo.2018.31.

Enders, C. (2010) *Applied missing data analysis*, *Journal of Chemical Information and Modeling*. Edited by T. D. Little. New York, NY, US: Guilford Press. doi: 10.1017/CBO9781107415324.004.

Epstein, N. B., Baldwin, L. M. and Bishop, D. S. (1983) 'The McMaster Family Assessment Device.', *Journal of Marital and Family Therapy*, 9(2), pp. 171–180.

Ettekal, I. and Ladd, G. W. (2015) 'Developmental Pathways From Childhood Aggression-Disruptiveness, Chronic Peer Rejection, and Deviant Friendships to Early-Adolescent Rule Breaking', *Child Development*, 86(2), pp. 614–631. doi: 10.1111/cdev.12321.

Farbstein, I. *et al.* (2010) 'Prevalence and correlates of mental disorders in Israeli adolescents: results from a national mental health survey.', *Journal of child psychology and psychiatry, and allied disciplines*, 51(5), pp. 630–9. doi: 10.1111/j.1469-7610.2009.02188.x.

Ferdinand, R. F. *et al.* (2008) 'Homotypic versus heterotypic continuity of anxiety symptoms in

young adolescents: Evidence for distinctions between DSM-IV subtypes', *Journal of Abnormal Child Psychology*, 36(3), p. 457. doi: 10.1007/s10802-007-9191-7.

Fergusson, D. M., Boden, J. M. and Horwood, L. J. (2007) 'Recurrence of major depression in adolescence and early adulthood, and later mental health, educational and economic outcomes', *British Journal of Psychiatry*, 191, pp. 335–342.

Ferrer, R. and Pardo, A. (2014) 'Clinically meaningful change: False positives in the estimation of individual change', *Psychological Assessment*, 26(2), pp. 370–383. doi: 10.1037/a0035419.

Ferrer, R. and Pardo, A. (2019) 'Clinically meaningful change: False negatives in the estimation of individual change', *Methodology*, 15(3), pp. 97–105. doi: 10.1027/1614-2241/a000168.

Ferro, M. A. (2014) 'Missing data in longitudinal studies: Cross-sectional multiple imputation provides similar estimates to full-information maximum likelihood', *Annals of Epidemiology*, 24(1), pp. 75–77. doi: 10.1016/j.annepidem.2013.10.007.

Finan, L. J., Ohannessian, C. M. C. and Gordon, M. S. (2018) 'Trajectories of depressive symptoms from adolescence to emerging adulthood: The influence of parents, peers, and siblings', *Developmental Psychology*, 54(8), pp. 1555–1567. doi: 10.1037/dev0000543.

Fisher, H. L. *et al.* (2013) 'Pathways between childhood victimization and psychosis-like symptoms in the ALSPAC birth cohort', *Schizophrenia Bulletin*, 39(5), pp. 1045–1055. doi: 10.1093/schbul/sbs088.

Flora, D. B. and Curran, P. J. (2004) 'An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data', *Psychological Methods*, 9(4), pp. 466–491. doi: 10.1037/1082-989X.9.4.466.

Fonagy, P. and Allison, E. (2014) 'The role of mentalizing and epistemic trust in the therapeutic relationship', *Psychotherapy*, 51(3), pp. 372–380. doi: 10.1037/a0036505.

Ford, T. *et al.* (2019) 'The effectiveness and cost-effectiveness of the Incredible Years ® Teacher Classroom Management programme in primary school children: Results of the STARS cluster randomised controlled trial', *Psychological Medicine*, 49(5), pp. 828–842. doi: 10.1017/S0033291718001484.

Ford, T., Goodman, R. and Meltzer, H. (2003) 'The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders', *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(10), pp. 1203–1211. doi: 10.1097/01.chi.0000081820.25107.ae.

Frick, P. J., Christian, R. E. and Wootton, J. M. (1999) 'Age trends in the association between parenting practices and conduct problems', *Behavior Modification*, 23(1), pp. 106–128. doi: 10.1177/0145445599231005.

Friedmann, M. S. *et al.* (1997) 'Family Functioning and Mental Illness: A Comparison of Psychiatric and Nonclinical Families', *Family Process*, 36(4), pp. 357–367. doi: 10.1111/j.1545-5300.1997.00357.x.

Furukawa, T. A. *et al.* (2014) 'Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis', *Acta Psychiatrica Scandinavica*, 130(3), pp. 181–192. doi: 10.1111/acps.12275.

Giedd, J. N. (2004) 'Structural magnetic resonance imaging of the adolescent brain', *Annals of the New York Academy of Sciences*, 1021, pp. 77–85. doi: 10.1196/annals.1308.009.

Giedd, J. N. *et al.* (2015) 'Child psychiatry branch of the national institute of mental health longitudinal structural magnetic resonance imaging study of human brain development', *Neuropsychopharmacology*. Nature Publishing Group, 40(1), pp. 43–49. doi: 10.1038/npp.2014.236.

Gignac, G. E. (2016) 'The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better', *Intelligence*. Elsevier Inc., 55, pp. 57–68. doi:

10.1016/j.intell.2016.01.006.

Glenn, A. L., Johnson, A. K. and Raine, A. (2013) 'Antisocial personality disorder: A current review', *Current Psychiatry Reports*, 15(12). doi: 10.1007/s11920-013-0427-7.

Gluschkoff, K., Jokela, M. and Rosenström, T. (2019) 'The general psychopathology factor: Structural stability and generalizability to within-individual changes', *Frontiers in Psychiatry*, 10(594). doi: 10.3389/fpsyt.2019.00594.

Goldsmith, K. *et al.* (2018) 'Measurement error, time lag, unmeasured confounding: Considerations for longitudinal estimation of the effect of a mediator in randomised clinical trials', *Statistical Methods in Medical Research*, 27(6), pp. 1615–1633. doi: 10.1177/0962280216666111.

Goldsmith, K. A. *et al.* (2018) 'Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials', *Psychological Methods*, 23(2), pp. 191–207. doi: 10.1037/met0000154.

Gonzalez, O. and MacKinnon, D. P. (2018) 'A Bifactor Approach to Model Multifaceted Constructs in Statistical Mediation Analysis', *Educational and Psychological Measurement*, 78(1), pp. 5–31. doi: 10.1177/0013164416673689.

Goodman, A., Joyce, R. and Smith, J. P. (2011) 'The long shadow cast by childhood physical and mental problems on adult life', *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), pp. 6032–6037. doi: 10.1073/pnas.1016970108.

Goodwin, R. D., Fergusson, D. M. and Horwood, L. J. (2003) 'Neuroticism in adolescence and psychotic symptoms in adulthood', *Psychological Medicine*. The Librarian-Seeley Historical Library, 33(6), pp. 1089–1097. doi: 10.1017/S0033291703007888.

Goodyer, I., Kolvin, I. and Gatzanis, S. (1985) 'Recent undesirable life events and psychiatric disorder in childhood and adolescence', *British Journal of Psychiatry*, 147, pp. 517–523.

Goodyer, I. M. *et al.* (2000) 'Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents.', *British Journal of Psychiatry*, 177, pp. 499–504.

Goodyer, I. M. *et al.* (2010) 'Cohort Profile: Risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project', *International Journal of Epidemiology*, 39(2), pp. 361–369. doi: 10.1093/ije/dyp173.

Goodyer, I. M. *et al.* (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', *The Lancet Psychiatry*, 4(2), pp. 109–119. doi: 10.1016/S2215-0366(16)30378-9.

Goodyer, I., Wright, C. and Altham, P. (1990) 'The friendships and recent life events of anxious and depressed school-age children', *British Journal of Psychiatry*, 156, pp. 689–698.

Gottfredson, N. C., Sterba, S. K. and Jackson, K. M. (2017) 'Explicating the Conditions Under Which Multilevel Multiple Imputation Mitigates Bias Resulting from Random Coefficient-Dependent Missing Longitudinal Data', *Prevention Science*. Prevention Science, 18(1), pp. 12–19. doi: 10.1007/s11121-016-0735-3.

Graham, J. W. (2003) 'Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models', *Structural Equation Modeling: A multidisciplinary journal*, 10(1), pp. 80–100. doi: 10.1207/S15328007SEM1001.

Green, H. *et al.* (2005) *Mental health of children and young people in Great Britain, 2004*. Basingstoke. doi: 10.1037/e557702010-001.

Green, J. M. *et al.* (2011) 'Group therapy for adolescents with repeated self harm: randomised controlled trial with economic evaluation', *BMJ*, 342, pp. 1–12. doi: 10.1136/bmj.d682.

Green, K. and Stuart, E. A. (2014) 'Examining moderation analyses in Propensity Score methods: Application to depression and substance use', 82(5), pp. 773–783. doi: 10.1037/a0036515.

Greenberg, M. T., Siegel, J. M. and Leitch, C. J. (1983) 'The nature and importance of attachment relationships to parents and peers during adolescence', *Journal of Youth and Adolescence*, 12(5), pp. 373–386. doi: 10.1007/BF02088721.

Greene, A. L. and Eaton, N. R. (2017) 'The temporal stability of the bifactor model of comorbidity: An examination of moderated continuity pathways', *Comprehensive Psychiatry*, 72, pp. 74–82. doi: 10.1016/j.comppsych.2016.09.010.

Häfner, H. *et al.* (1994) 'The epidemiology of early schizophrenia. Influence of age and gender on onset and early course.', *Br J Psychiatry*, Apr;(23), pp. 29–38.

Hale, W. W. *et al.* (2008) 'The interplay of early adolescents' depressive symptoms, aggression and perceived parental rejection: A four-year community study', *Journal of Youth and Adolescence*, 37(8), pp. 928–940. doi: 10.1007/s10964-008-9280-0.

Hancock, G. R. (2001) 'Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct', *Psychometrika*, 66(3), pp. 373–388. doi: 10.1007/BF02294440.

Hancock, G. R. and Mueller, R. O. (2001) 'Rethinking construct reliability within latent variable systems', in Cudeck, R., du Toit, S., and Sörbom, D. (eds) *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog*. Lincolnwood, IL: Scientific Software Int, pp. 195–216.

Hankin, B. L. *et al.* (2017) 'Temperament factors and dimensional, latent bifactor models of child psychopathology: Transdiagnostic and specific associations in two youth samples', *Psychiatry Research*. Elsevier Ireland Ltd, 252, pp. 139–146. doi: 10.1016/j.psychres.2017.02.061.

Hardeveld, F. *et al.* (2013) 'Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study', *Journal of Affective Disorders*. Elsevier, 147(1–3), pp. 225–231. doi: 10.1016/j.jad.2012.11.008.

van Harmelen, A. L. *et al.* (2016) 'Friendships and Family Support Reduce Subsequent Depressive Symptoms in At-Risk Adolescents', *PloS one*, 11(5), p. e0153715. doi: 10.1371/journal.pone.0153715.

van Harmelen, A. L. *et al.* (2017) 'Adolescent friendships predict later resilient functioning across psychosocial domains in a healthy community cohort', *Psychological Medicine*, 47(13), pp. 2312–2322. doi: 10.1017/S0033291717000836.

Ten Have, T. R. *et al.* (2008) 'Intent-to-Treat vs. Non-Intent-to-Treat Analyses under Treatment Non-Adherence in Mental Health Randomized Trials', *Psychiatric Annals*, 38(12), pp. 772–783. doi: 10.3928/00485713-20081201-10.

Hay, D. F. and Pawlby, S. (2003) 'Prosocial Development in Relation to Children's and Mothers' Psychological Problems', *Child Development*, 74(5), pp. 1314–1327.

Health and Social Care Information Centre (2014) *Psychological Therapies, Annual Report on the use of IAPT services*. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2013-14.

Heckman, J. J. *et al.* (1996) 'Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method', *Proceedings of the National Academy of Sciences of the United States of America*, 93, pp. 13416–13420.

Hengartner, M. P. (2019) 'Is there a genuine placebo effect in acute depression treatments? A reassessment of regression to the mean and spontaneous remission', *BMJ Evidence-Based Medicine*, 11 April. doi: 10.1136/bmjebm-2019-111161.

Henin, A. and Berman, N. (2016) 'The Promise and Peril of Emerging Adulthood: Introduction to the Special Issue', *Cognitive and Behavioral Practice*, 23(3), pp. 263–269. doi:

10.1016/j.cbpra.2016.05.005.

Hoagwood, K. *et al.* (2000) 'Concordance between parent reports of children's mental health services and service records: The Services Assessment for Children and Adolescents (SACA)', *Journal of Child and Family Studies*, 9(3), pp. 315–331. doi: 10.1023/A:1026492423273.

Hofmann, S. G. *et al.* (2012) 'The efficacy of Cognitive Behavioral Therapy', *Cognitive Therapy and Research*, 36(5), pp. 427–440. doi: 10.1007/s10608-012-9476-1.The.

Hollon, S. D., Stewart, M. O. and Strunk, D. (2006) 'Enduring Effects for Cognitive Behavior Therapy in the Treatment of Depression and Anxiety', *Annual Review of Psychology*, 57(1), pp. 285–315. doi: 10.1146/annurev.psych.57.102904.190044.

Holzinger, K. J. and Swineford, F. (1937) 'The bi-factor method', *Psychometrika*, 2(1), pp. 41–54.

Hoofs, H. *et al.* (2018) 'Evaluating Model Fit in Bayesian Confirmatory Factor Analysis With Large Samples: Simulation Study Introducing the BRMSEA', *Educational and Psychological Measurement*, 78(4), pp. 537–568. doi: 10.1177/0013164417709314.

Hu, L. T. and Bentler, P. M. (1999) 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives', *Structural Equation Modeling*, 6(1), pp. 1–55. doi: 10.1080/10705519909540118.

Jacobson, N. S. and Truax, P. (1991) 'Clinical Significance : A Statistical Approach to Denning Meaningful Change in Psychotherapy Research', 59(1), pp. 12–19.

Jacobucci, R., Brandmaier, A. M. and Kievit, R. A. (2019) 'A practical guide to variable selection in Structural Equation Modeling by using regularized Multiple-Indicators, Multiple-Causes models', *Advances in Methods and Practices in Psychological Science*, 2(1), pp. 55–76. doi: 10.1177/2515245919826527.

Johnson, E. C., Meade, A. W. and DuVernet, A. M. (2009) 'The Role of Referent Indicators in Tests of Measurement Invariance', *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), pp. 642–657. doi: 10.1080/10705510903206014.

Johnson, S. E. *et al.* (2016) 'Service use by Australian children for emotional and behavioural problems: Findings from the second Australian Child and Adolescent Survey of Mental Health and Wellbeing', *Australian & New Zealand Journal of Psychiatry*, pp. 0004867415622562-. doi: 10.1177/0004867415622562.

Jones, P. B. (2013) 'Adult mental health disorders and their age at onset.', *The British journal of psychiatry. Supplement*, 54, pp. s5-10. doi: 10.1192/bjp.bp.112.119164.

Jongsma, H. E. *et al.* (2019) 'International incidence of psychotic disorders, 2002–17: a systematic review and meta-analysis', *The Lancet Public Health*, 4(5), pp. e229–e244. doi: 10.1016/S2468-2667(19)30056-8.

Jonsson, U. *et al.* (2014) 'Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice', *Contemporary Clinical Trials*. Elsevier Inc., 38, pp. 1–8. doi: 10.1016/j.cct.2014.02.005.

Joo, S. H. and Kim, E. S. (2018) 'Impact of error structure misspecification when testing measurement invariance and latent-factor mean difference using MIMIC and multiple-group confirmatory factor analysis', *Behavior Research Methods*, 51(2688). doi: 10.3758/s13428-018-1124-6.

Kass, R. and Raftery, A. (1995) 'Bayes factors', *Journal of the American Statistical Association*, 90, pp. 773–795. doi: http://dx.doi.org/10 .1080/01621459.1995.10476572.

Kessler, R. C., Berglund, P., *et al.* (2005) 'Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication', *Archives of General Psychiatry*, pp. 593–602. doi: 10.1001/archpsyc.62.6.593.

Kessler, R. C., Chiu, W. T., *et al.* (2005) 'Prevalence, Severity, and Comorbidity of 12-Month DSM-

IV Disorders in the National Comorbidity Survey Replication', *Arch Gen Psychiatry*, 62(6), pp. 617–627. doi: 10.1001/archpsyc.62.6.617.

Kessler, R. C. *et al.* (2007) 'Age of onset of mental disorders: a review of recent literature.', *Current opinion in psychiatry*, 20(4), pp. 359–64. doi: 10.1097/YCO.0b013e32816ebc8c.

Kessler, R. C. *et al.* (2011) 'Development of lifetime comorbidity in the WHO World Mental Health (WMH) Surveys', *Archives of general psychiatry*, 68(1), pp. 90–100. doi: 10.1001/archgenpsychiatry.2010.180.Development.

Kessler, R. C. *et al.* (2012) 'Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement.', *Archives of general psychiatry*, 69(4), pp. 372–80. doi: 10.1001/archgenpsychiatry.2011.160.

Keyes, C. (2002) 'The mental health continuum: from languishing to flourishing in life', *Journal of health and social behavior*, 43(2), pp. 207–22.

Khojasteh, J. and Lo, W. J. (2015) 'Investigating the Sensitivity of Goodness-of-Fit Indices to Detect Measurement Invariance in a Bifactor Model', *Structural Equation Modeling*, 22(4), pp. 531–541. doi: 10.1080/10705511.2014.937791.

Kiddle, B. *et al.* (2018) 'Cohort profile: The NSPN 2400 Cohort: A developmental sample supporting the Wellcome Trust Neuro Science in Psychiatry Network', *International Journal of Epidemiology*, 47(1), pp. 18-19g. doi: 10.1093/ije/dyx117.

Kim, H. and Eaton, N. R. (2015) 'The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity', *Journal of Abnormal Psychology*, 124(4), pp. 1064–1078. doi: 10.1037/abn0000113.

Kirkbride, J. B. *et al.* (2014) 'Social deprivation, inequality, and the neighborhood-level incidence of psychotic syndromes in East London', *Schizophrenia Bulletin*, 40(1), pp. 169–180. doi: 10.1093/schbul/sbs151.

Kirkbride, J. B. *et al.* (2017) 'Ethnic minority status, age-at-immigration and psychosis risk in rural environments: Evidence from the SEPEA study', *Schizophrenia Bulletin*, 43(6), pp. 1251–1261. doi: 10.1093/schbul/sbx010.

Knapp, M. (2014) *Youth mental health: New economic evidence*, *Unit Costs of Health and Social Care 2014*. London, UK. Available at: http://www.pssru.ac.uk/project-pages/unit-costs/2014/.

Korten, A. and Henderson, S. (2000) 'The Australian national survey of mental health and well-being: Common psychological symptoms and disablement', *British Journal of Psychiatry*, 177, pp. 325–330. doi: 10.1192/bjp.177.4.325.

Kotov, R. *et al.* (2010) 'Linking "Big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis', *Psychological Bulletin*, 136(5), pp. 768–821. doi: 10.1037/a0020327.

Kotov, R., Ruggero, C. J., *et al.* (2011) 'New dimensions in the quantitative classification of mental illness', *Archives of General Psychiatry*, 68(10), pp. 1003–1011. doi: 10.1001/archgenpsychiatry.2011.107.

Kotov, R., Chang, S. W., *et al.* (2011) 'Schizophrenia in the internalizing-externalizing framework: A third dimension?', *Schizophrenia Bulletin*, 37(6), pp. 1168–1178. doi: 10.1093/schbul/sbq024.

Krueger, R. F. *et al.* (2018) 'Progress in achieving quantitative classification of psychopathology', *World Psychiatry*, 17(3), pp. 282–293. doi: 10.1002/wps.20566.

Krueger, R. F. and Markon, K. E. (2006) 'Reinterpreting Comorbidity: A Model-Based Approach to Understanding and Classifying Psychopathology', *Annual Review of Clinical Psychology*, 2(1), pp. 111–133. doi: 10.1146/annurev.clinpsy.2.022305.095213.

Krueger, R. F. and Markon, K. E. (2011) 'A Dimensional-Spectrum Model of Psychopathology', *Archives of General Psychiatry*, 68(1), p. 10. doi: 10.1001/archgenpsychiatry.2010.188.

Kuyken, W. *et al.* (2017) 'The effectiveness and cost-effectiveness of a mindfulness training programme in schools compared with normal school provision (MYRIAD): Study protocol for a randomised controlled trial', *Trials*. Trials, 18(1), pp. 1–17. doi: 10.1186/s13063-017-1917-4.

Kyu, H. H. *et al.* (2018) 'Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017', *The Lancet*, pp. 1859–1922. doi: 10.1016/S0140-6736(18)32335-3.

Laceulle, O. M. *et al.* (2019) 'The wide-ranging life outcome correlates of a general psychopathology factor in adolescent psychopathology', *Personality and Mental Health*. doi: 10.1002/pmh.

Laceulle, O. M., Vollebergh, W. A. M. and Ormel, J. (2015) 'The structure of psychopathology in adolescence: replication of a general psychopathology factor in the TRAILS study', *Clinical Psychological Science*, 3(6), pp. 850–860. doi: 10.1177/2167702614560750.

Lahey, B. B. *et al.* (2008) 'Psychometric characteristics of a measure of emotional dispositions developed to test a developmental propensity model of conduct disorder', *Journal of Clinical Child and Adolescent Psychology*, 37(4), pp. 794–807. doi: 10.1080/15374410802359635.

Lahey, B. B. *et al.* (2012) 'Is there a general factor of prevalent psychopathology during adulthood?', *Journal of Abnormal Psychology*, 121(4), pp. 971–977. doi: 10.1037/a0028355.

Lahey, B. B. *et al.* (2014) 'Patterns of heterotypic continuity associated with the cross-sectional correlational structure of prevalent mental disorders in adults', *JAMA Psychiatry*, 71(9), pp. 989–996. doi: 10.1001/jamapsychiatry.2014.359.

Lahey, B. B. *et al.* (2015) 'Criterion validity of the general factor of psychopathology in a prospective study of girls', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56(4), pp. 415–422. doi: 10.1111/jcpp.12300.

Lahey, B. B. *et al.* (2017) 'A hierarchical causal taxonomy of psychopathology across the life span', *Psychological Bulletin*, 143(2), pp. 142–186. doi: 10.1037/bul0000069.

Lahey, B. B. *et al.* (2018a) 'Measuring the hierarchical general factor model of psychopathology in young adults', *International Journal of Methods in Psychiatric Research*, 27(e1593). doi: 10.1002/mpr.1593.

Lahey, B. B. *et al.* (2018b) 'Measuring the hierarchical general factor model of psychopathology in young adults', *International Journal of Methods in Psychiatric Research*, 27(1), pp. 1–9. doi: 10.1002/mpr.1593.

Lahey, B. B. and Waldman, I. D. (2003) 'A developmental propensity model of the origins of conduct problems during childhood and adolescence', in Lahey, B., Moffitt, T., and Caspi, A. (eds) *Causes of conduct disorder and juvenile delinquency*. New York, NY, US: Guilford Press, pp. 76–117. Available at: https://www.guilford.com/excerpts/lahey.pdf?t.

Lampropoulou, A. (2018) 'Personality, school, and family: What is their role in adolescents' subjective well-being', *Journal of Adolescence*. Elsevier, 67, pp. 12–21. doi: 10.1016/j.adolescence.2018.05.013.

Larson, R. and Richards, M. (1991) 'Daily Companionship in Late Childhood and Early Adolescence: Changing Developmental Contexts', *Child development*, 62(2), pp. 284–300.

Larson, R. W. *et al.* (1996) 'Changes in Adolescents' Daily Interactions With Their Families From Ages 10 to 18: Disengagement and Transformation', *Developmental Psychology*, 32(4), pp. 744–754.

Lawrence, D. *et al.* (2015) *The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing.* Canberra.

Leite, W. L., Stapleton, L. M. and Bettini, E. F. (2019) 'Propensity Score Analysis of Complex Survey Data with Structural Equation Modeling: A Tutorial with Mplus', *Structural Equation*

Modeling. Routledge, 26(3), pp. 448–469. doi: 10.1080/10705511.2018.1522591.

Levitt, M. J., Silver, M. E. and Santos, J. D. (2007) 'Adolescents in transition to adulthood: Parental support, relationship satisfaction, and post-transition adjustment', *Journal of Adult Development*, 14(1–2), pp. 53–63. doi: 10.1007/s10804-007-9032-5.

Li, M. *et al.* (2018) 'Psychometric properties and measurement invariance of the brief symptom inventory-18 among Chinese insurance employees', *Frontiers in Psychology*, 9(519). doi: 10.3389/fpsyg.2018.00519.

Linden, A. and Adams, J. L. (2010) 'Evaluating health management programmes over time: Application of propensity score-based weighting to longitudinal data', *Journal of Evaluation in Clinical Practice*, 16, pp. 180–185. doi: 10.1111/j.1365-2753.2009.01361.x.

Liston, C. *et al.* (2006) 'Frontostriatal microstructure modulates efficient recruitment of cognitive control', *Cerebral Cortex*, 16(4), pp. 553–560. doi: 10.1093/cercor/bhj003.

Little, R. (1995) 'Modeling the Drop-Out Mechanism in Repeated-Measures Studies', *Journal of the American Statistical Association*, 90, pp. 1112–11121.

Little, R. J. A. (1988) 'A test of missing completely at random for multivariate data with missing values', *Journal of the American Statistical Association*, 83(404), pp. 1198–1202. doi: 10.1080/01621459.1988.10478722.

Little, T. D. (2013) *Methodology in the social sciences. Longitudinal structural equation modeling.* New York, NY, US: Guilford Press.

Liu, Y. *et al.* (2017) 'Testing Measurement Invariance in Longitudinal Data With Ordered-Categorical Measures', *Psychological Methods*, 22(3), pp. 486–506. doi: 10.1037/met0000075.

Loeber, R. and Farrington, D. (2014) 'Age-crime curve', in Lum, C. and Koper, C. (eds) *Evidence-Based Policing; Encyclopedia of Criminology and Criminal Justice*. New York, NY: Springer, pp. 12–14. doi: 10.1007/978-1-4614-5690-2.

London & South East CYP IAPT Learning Collaborative (2016) *CYP IAPT and Children & Young Peoples Mental Health Service Transformation : An essential guide for commissioners*. London, UK. Available at: https://cypiapt.files.wordpress.com/2015/06/for-commissioners-v-1-3.pdf.

Lubke, G. H. and Muthen, B. O. (2004) 'Applying multigroup confirmatory factor for continuous outcomes to Likert scale data complicates meningful group comparisons', *Structural Equation Modeling*, 11(4), pp. 514–534.

Luebbe, A. M. *et al.* (2018) 'Dimensionality of Helicopter Parenting and Relations to Emotional, Decision-Making, and Academic Functioning in Emerging Adults', *Assessment*, 25(7), pp. 841–857. doi: 10.1177/1073191116665907.

MacCallum, R. C., Browne, M. W. and Sugawara, H. M. (1996) 'Power analysis and determination of sample size for covariance structure modeling', *Psychological Methods*, 1(2), pp. 130–149. doi: 10.1037/1082-989X.1.2.130.

MacDonald, K. *et al.* (2016) 'Minimization of childhood maltreatment is common and consequential: Results from a large, multinational sample using the Childhood Trauma Questionnaire', *PLoS ONE*, 11(1), p. e0146058. doi: 10.1371/journal.pone.0146058.

Mansbach-Kleinfeld, I. *et al.* (2010) 'Service use for mental disorders and unmet need: results from the Israel Survey on Mental Health Among Adolescents.', *Psychiatric services*, 61(3), pp. 241–9. doi: 10.1176/appi.ps.61.3.241.

Mansfield, A. K., Keitner, G. I. and Sheeran, T. (2018) 'The Brief Assessment of Family Functioning Scale (BAFFS): a three-item version of the General Functioning Scale of the Family Assessment Device', 3307. doi: 10.1080/10503307.2017.1422213.

Markon, K. E. (2010) 'Modeling psychopathology structure: A symptom-level analysis of Axis i and II disorders', *Psychological Medicine*. Pendlebury Library of Music, 40(2), pp. 273–288. doi:

10.1017/S0033291709990183.

Marsh, H. W., Parker, P. D. and Morin, A. J. S. (2016) 'Invariance testing across samples and time: Cohort-sequence analysis of perceived body composition', in Ntoumanis, N. and Myers, N. D. (eds) *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists*. Chichester, West Sussex, UK: Wiley, pp. 101–130. Available at: https://books.google.com/books?hl=en&lr=lang_en&id=AHZXCQAAQBAJ&pgis=1.

Martel, M. *et al.* (2016) 'A general psychopathology factor (p-factor) in children: Structural model analysis and external validation through familial risk and child executive function (under review).', *Journal of abnormal psychology*, 126(1), pp. 137–148.

McCrea, R. L., Berger, Y. G. and King, M. B. (2012) 'Body mass index and common mental disorders: Exploring the shape of the association and its moderation by age, gender and education', *International Journal of Obesity*, 36(3), pp. 414–421. doi: 10.1038/ijo.2011.65.

McElroy, E. *et al.* (2018) 'Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: dynamic mutualism or p-differentiation?', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 59(6), pp. 667–675. doi: 10.1111/jcpp.12849.

McGorry, P. (2013) 'Prevention, innovation and implementation science in mental health: The next wave of reform', *British Journal of Psychiatry*, 202(SUPPL. 54), pp. 20–22. doi: 10.1192/bjp.bp.112.119222.

McGorry, P., Bates, T. and Birchwood, M. (2013) 'Designing youth mental health services for the 21st century: examples from Australia, Ireland and the UK.', *The British journal of psychiatry. Supplement*, 54, pp. s30-5. doi: 10.1192/bjp.bp.112.119214.

Mcmanus, S. *et al.* (2016) *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey, 2014*. Leicester, UK.

McNeish, D. (2018) 'Thanks coefficient alpha, We'll take it from here', *Psychological Methods*, 23(3), pp. 412–433. doi: 10.1037/met0000144.

Meade, A. W., Johnson, E. C. and Braddy, P. W. (2008) 'Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance', *Journal of Applied Psychology*, 93(3), pp. 568–592. doi: 10.1037/0021-9010.93.3.568.

Meeus, W. (2016) 'Adolescent psychosocial development: A review of longitudinal models and research', *Developmental Psychology*, 52(12), pp. 1969–1993. doi: 10.1037/dev0000243.

Merikangas, K. R. *et al.* (2010) 'Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES.', *Pediatrics*, 125(1), pp. 75–81. doi: 10.1542/peds.2008-2598.

Miller, I. W. *et al.* (1985) 'The McMaster Family Assessment Device: Reliability and validity', *Journal of Marital and Family Therapy*, 11(4), pp. 345–356. doi: 10.1111/j.1752-0606.1985.tb00028.x.

Miller, I. W. *et al.* (2000) 'The McMaster approach to families: Theory, assessment, treatment and research', *Journal of Family Therapy*, 22(2), pp. 168–189. doi: 10.1111/1467-6427.00145.

Millsap, R. E. and Yun-Tein, J. (2004) 'Assessing factorial invariance in ordered categorial measures', *Multivariate Behavioral Research*, 39(3), pp. 479–515. doi: 10.1207/S15327906MBR3903_4.

Minami, T. *et al.* (2007) 'Benchmarks for psychotherapy efficacy in adult major depression', *Journal of Consulting and Clinical Psychology*, 75(2), pp. 232–243. doi: 10.1037/0022-006X.75.2.232.

Moffitt, T. E. *et al.* (2010) 'How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment', *Psychological Medicine*. Pendlebury Library of Music, 40(6), pp. 899–909. doi: 10.1017/S0033291709991036.

Munder, T. *et al.* (2019) 'Is psychotherapy effective? A re-analysis of treatments for depression', *Epidemiology and Psychiatric Sciences*, 28(3), pp. 268–274. doi: 10.1017/S2045796018000355.

Murray, A. L., Eisner, M. and Ribeaud, D. (2016) 'The Development of the General Factor of Psychopathology "p Factor" Through Childhood and Adolescence', *Journal of Abnormal Child Psychology*. Journal of Abnormal Child Psychology, 44(8), pp. 1573–1586. doi: 10.1007/s10802-016-0132-1.

Murray, A. L. and Johnson, W. (2013) 'The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure', *Intelligence*. Elsevier Inc., 41(5), pp. 407–422. doi: 10.1016/j.intell.2013.06.004.

Muthén, B. O. and Asparouhov, T. (2002) *Latent Variable Analysis With Categorical Outcomes: Multiple-Group And Growth Modeling In Mplus*, *Mplus Web Notes*.

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

Nantel-Vivier, A. *et al.* (2014) 'Developmental association of prosocial behaviour with aggression, anxiety and depression from infancy to preadolescence', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 55(10), pp. 1135–1144. doi: 10.1111/jcpp.12235.

National Collaborating Centre for Mental Health (2018) *The Improving Access to Psychological Therapies Manual*. London, UK. Available at: https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual.pdf.

Neufeld, S. A. S. *et al.* (2017) 'Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK', *The Lancet Psychiatry*, 4(2), pp. 120–127. doi: 10.1016/S2215-0366(17)30002-0.

Neufeld, S. A. S., Jones, P. B. and Goodyer, I. M. (2017) 'Child and adolescent mental health services: longitudinal data sheds light on current policy for psychological interventions in the community', *Journal of Public Mental Health*, 16(3), pp. 96–99. doi: 10.1108/JPMH-03-2017-0013.

Neumann, A. *et al.* (2016) 'Single nucleotide polymorphism heritability of a general psychopathology factor in children', *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(12), pp. 1038-1045.e4. doi: 10.1016/j.jaac.2016.09.498.

Newman, D. L. *et al.* (1996) 'Psychiatric disorder in a birth cohort of young adults: Prevalence, comorbidity, clinical significance, and new case incidence from ages 11 to 21', *Journal of Consulting and Clinical Psychology*, 64(3), pp. 552–562. doi: 10.1037/0022-006X.64.3.552.

NHS Benchworking Network (2018) 'CAMHS 2018 key findings'. Available at: https://www.nhsbenchmarking.nhs.uk/news/camhs-benchmarking-2016-findings-published.

NHS England (2015a) *Implementing the five year forward view for mental health*. Redditch, UK. Available at: https://www.england.nhs.uk/wp-content/uploads/2016/07/fyfv-mh.pdf.

NHS England (2015b) *Model Specification for Child and Adolescent Mental Health Services: Targeted and Specialist levels (Tiers 2/3)*. London, UK. Available at: http://www.england.nhs.uk/wp-content/uploads/2015/01/mod-camhs-tier-2-3-spec.pdf.

NHS England (2016) *Children and Young People's Mental Health Services Baselining Report*. Available at: https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2015/08/nhse-camhs-baselining-summary1.pdf.

NICE (2013) *Social Anxiety Disorder: recognition, assessment and treatment, Clinical guideline, National Institute for Health and Care Excellence*. London, UK. doi: 10.1002/9781118775349.ch47.

NICE (2014a) *Antisocial behaviour and conduct disorders in children and young people, Quality standard, National Institute for Health and Care Excellence*. London, UK. Available at: https://www.nice.org.uk/guidance/qs59.

NICE (2014b) *Anxiety disorders quality standard, National Institute for Health and Care Excellence*. London, UK. Available at: www.nice.org.uk/guidance/qs53.

NICE (2014c) *Psychosis and schizophrenia in adults: prevention and management, Clinical*

guideline, National Institute for Health and Care Excellence. London, UK. Available at: https://www.nice.org.uk/guidance/cg178/resources/psychosis-and-schizophrenia-in-adults-prevention-and-management-pdf-35109758952133.

NICE (2015) *Personality disorders: borderline and antisocial, Quality Standard, National Institute for Health and Care Excellence*. London, UK. Available at: https://www.nice.org.uk/guidance/qs88.

NICE (2016) *Psychosis and schizophrenia in children and young people: recognition and management, Clinical guideline, National Institute for Health and Care Excellence*. London, UK. Available at: https://www.nice.org.uk/guidance/cg155.

NICE (2017) *Depression in adults: Recognition and management*, *Clinical guideline, National Institute for Health and Care Excellence*. London, UK. doi: 10.1211/CP.2017.20202439.

NICE (2018) *Post-traumatic stress disorder, NICE guideline, National Institute for Health and Care Excellence*. London, UK. Available at: https://www.nice.org.uk/guidance/ng116.

NICE (2019a) *Depression in children and young people: identification and management, NICE guideline, National Institute for Health and Care Excellence*. London, UK. Available at: https://www.nice.org.uk/guidance/ng134.

NICE (2019b) *Depression in children and young people: Identification and management, NICE guideline evidence review, National Institute for Health and Care Excellence*. London, UK. doi: 10.1211/CP.2018.20204575.

Nock, M. K. and Prinstein, M. J. (2004) 'A functional approach to the assessment of self-mutilative behavior', *Journal of Consulting and Clinical Psychology*, 72(5), pp. 885–890. doi: 10.1037/0022-006X.72.5.885.

Nomura, Y. *et al.* (2002) 'Family discord, parental depression, and psychopathology in offspring: Ten-Year follow-up', *Journal of the American Academy of Child and Adolescent Psychiatry*. American Academy of Child and Adolescent Psychiatry, 41(4), pp. 402–409. doi: 10.1097/01.chi.0000198592.23078.8d.

Noordhof, A. *et al.* (2015) 'Integrating Autism-Related Symptoms into the Dimensional Internalizing and Externalizing Model of Psychopathology. The TRAILS Study', *Journal of Abnormal Child Psychology*, 43(3), pp. 577–587. doi: 10.1007/s10802-014-9923-4.

Olino, T. M. *et al.* (2014) 'Testing models of psychopathology in preschool-aged children using a structured interview-based assessment', *Journal of Abnormal Child Psychology*, 42(7), pp. 1201–1211. doi: 10.1007/s10802-014-9865-x.

Olino, T. M. *et al.* (2018) 'The Development of Latent Dimensions of Psychopathology across Early Childhood: Stability of Dimensions and Moderators of Change', *Journal of Abnormal Child Psychology*. Journal of Abnormal Child Psychology, 46(1373). doi: 10.1007/s10802-018-0398-6.

Oltmanns, J. R. *et al.* (2018) 'General Factors of Psychopathology, Personality, and Personality Disorder: Across Domain Comparisons', *Clinical Psychological Science*, 6(4), pp. 581–589. doi: 10.1177/2167702617750150.

Ormel, J. *et al.* (2015) 'Mental health in Dutch adolescents: A TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders', *Psychological Medicine*, 45(2), pp. 345–360. doi: 10.1017/S0033291714001469.

Van Os, J. *et al.* (2009) 'A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness-persistence-impairment model of psychotic disorder', *Psychological Medicine*, 39(2), pp. 179–195. doi: 10.1017/S0033291708003814.

Patalay, P. *et al.* (2015) 'A general psychopathology factor in early adolescence', *British Journal of Psychiatry*, 207(1), pp. 15–22. doi: 10.1192/bjp.bp.114.149591.

Patton, G. C. *et al.* (2014) 'The prognosis of common mental disorders in adolescents: A 14-year prospective cohort study', *The Lancet*, 383(9926), pp. 1404–1411. doi: 10.1016/S0140-

6736(13)62116-9.

Paula, C. S. *et al.* (2014) 'The mental health care gap among children and adolescents: data from an epidemiological survey from four Brazilian regions.', *PloS one*, 9(2), p. e88241. doi: 10.1371/journal.pone.0088241.

Pendergast, L. L. *et al.* (2017) 'Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology', *Journal of School Psychology*. Society for the Study of School Psychology, 60, pp. 65–82. doi: 10.1016/j.jsp.2016.11.002.

Pettersson, E. *et al.* (2018) 'Criterion validity and utility of the general factor of psychopathology in childhood: Predictive associations with independently measured severe adverse mental health outcomes in adolescence', *Journal of the American Academy of Child and Adolescent Psychiatry*. Elsevier Inc, 57(6), pp. 372–383. doi: 10.1016/j.jaac.2017.12.016.

Pettit, J. W. *et al.* (2011) 'Developmental Relations between Perceived Social Support and Depressive Symptoms through Emerging Adulthood: Blood is Thicker than Water', *J Fam Psychol.*, 25(1), pp. 127–136. doi: doi:10.1037/a0022320.

Pettit, S. *et al.* (2017) 'Variation in referral and access to new psychological therapy services by age: An empirical quantitative study', *British Journal of General Practice*, 67(660), pp. e453–e459. doi: 10.3399/bjgp17X691361.

Pirracchio, R., Resche-Rigon, M. and Chevret, S. (2012) 'Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size.', *BMC medical research methodology*, 12(70), pp. 1–10. doi: 10.1186/1471-2288-12-70.

Pitchforth, J. *et al.* (2019) 'Mental health and well-being trends among children and young people in the UK, 1995-2014: Analysis of repeated cross-sectional national health surveys', *Psychological Medicine*, 49(8), pp. 1275–1285. doi: 10.1017/S0033291718001757.

Plener, P. L. *et al.* (2015) 'The longitudinal course of non-suicidal self-injury and deliberate self-harm: a systematic review of the literature', *Borderline Personality Disorder and Emotion Dysregulation*, 2(1), p. 2. doi: 10.1186/s40479-014-0024-3.

Pottick, K. J. *et al.* (2014) 'Clinical characteristics and outpatient mental health service use of transition-age youth in the USA', *Journal of Behavioral Health Services and Research*, 41(2), pp. 230–243. doi: 10.1007/s11414-013-9376-5.

Poulin, F. and Chan, A. (2010) 'Friendship stability and change in childhood and adolescence', *Developmental Review*, 30(3), pp. 257–272. doi: 10.1016/j.dr.2009.01.001.

Putnick, D. L. and Bornstein, M. H. (2016) 'Measurement invariance conventions and reporting: The state of the art and future directions for psychological research', *Developmental Review*. Elsevier Inc., 41, pp. 71–90. doi: 10.1016/j.dr.2016.06.004.

Reise, S. P. *et al.* (2013) 'Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective', *Educational and Psychological Measurement*, 73(1), pp. 5–26. doi: 10.1177/0013164412449831.

Reise, S. P. *et al.* (2016) 'Is the Bifactor Model a Better Model or Is It Just Better at Modeling Implausible Responses? Application of Iteratively Reweighted Least Squares to the Rosenberg Self-Esteem Scale', *Multivariate Behavioral Research*, 51(6), pp. 818–838. doi: 10.1080/00273171.2016.1243461.

Reise, S. P., Moore, T. M. and Haviland, M. G. (2010) 'Bifactor Models and Rotations', *Journal of Personality Assessment*, 92(6), pp. 544–559. doi: 10.1080/00223891.2010.496477.Bifactor.

Ren, P. *et al.* (2018) 'Is social support a cause or consequence of depression? A longitudinal study of adolescents', *Frontiers in Psychology*, 9, pp. 1–8. doi: 10.3389/fpsyg.2018.01634.

Reynolds, C. and Richmond, B. (1978) 'What I think and feel: a revised measure of children's manifest anxiety', *J Abnorm Child Psychol*, 6(2), pp. 271–80.

Ridenour, T. A., Daley, J. G. and Reich, W. (1999) 'Factor analyses of the family assessment device', *Family Process*, 38(4), pp. 497–510. doi: 10.1111/j.1545-5300.1999.00497.x.

Rodriguez, A., Reise, S. P. and Haviland, M. G. (2016a) 'Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures', *Journal of Personality Assessment*, 98(3), pp. 223–237. doi: 10.1080/00223891.2015.1089249.

Rodriguez, A., Reise, S. P. and Haviland, M. G. (2016b) 'Evaluating bifactor models: Calculating and interpreting statistical indices', *Psychological Methods*, 21(2), pp. 137–150. doi: 10.1037/met0000045.

Rosenbaum, P. R. and Rubin, D. B. (1983) 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, 70, pp. 41–55. doi: 10.1017/CBO9780511810725.016.

Rosseel, Y. (2012) 'Lavaan: An R package for structural equation modeling', *Journal of Statistical Software*, 48(2). doi: 10.18637/jss.v048.i02.

Rubin, D. B. and Rosenbaum, P. R. (1984) 'Reducing Bias in Observational Studies Using Score on the Propensity Subolassification', *The American Economic Review*, 79(387), pp. 516–524.

Rueter, M. A. *et al.* (1999) 'First onset of depressive or anxiety disorders predicted by the longitudinal course of internalizing symptoms and parent-adolescent disagreements', *Archives of General Psychiatry*, 56(8), pp. 726–732. doi: 10.1001/archpsyc.56.8.726.

Rutter, M., Kim-Cohen, J. and Maughan, B. (2006) 'Continuities and discontinuities in psychopathology between childhood and adult life', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 47(3–4), pp. 276–295. doi: 10.1111/j.1469-7610.2006.01614.x.

Sadler, K. *et al.* (2018) *Mental health of children and young people in England, 2017*. Leeds, UK. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017.

Sallis, H. *et al.* (2019) 'General psychopathology, internalising and externalising in children and functional outcomes in late adolescence', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 60(11), pp. 1183–1190. doi: 10.1111/jcpp.13067.

Sansone, R. A. and Sansone, L. A. (2013) 'Obesity and substance misuse: Is there a relationship?', *Innovations in Clinical Neuroscience*, 10(9–10), pp. 30–35.

Sass, D. A., Schmitt, T. A. and Marsh, H. W. (2014) 'Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators', *Structural Equation Modeling*. Routledge, 21(2), pp. 167–180. doi: 10.1080/10705511.2014.882658.

Saunders, J. B. and Lee, N. K. (2000) 'Hazardous alcohol use: Its delineation as a subthreshold disorder, and approaches to its diagnosis and management', *Comprehensive Psychiatry*, 41(2), pp. 95–103.

Schaefer, J. D. *et al.* (2017) 'Enduring mental health: Prevalence and prediction', *Journal of Abnormal Psychology*, 126(2), pp. 212–224. doi: 10.1037/abn0000232.

Schaefer, J. D. *et al.* (2018) 'Adolescent victimization and early-adult psychopathology: Approaching causal inference using a longitudinal twin study to rule out noncausal explanations', *Clinical Psychological Science*, 6(3), pp. 352–371. doi: 10.1177/2167702617741381.

Van De Schoot, R. *et al.* (2015) 'Editorial: Measurement Invariance', *Frontiers in Psychology*, 6, p. 1064. doi: 10.3389/fpsyg.2015.01064.

van de Schoot, R., Lugtig, P. and Hox, J. (2012) 'A checklist for testing measurement invariance', *European Journal of Developmental Psychology*, 9(4), pp. 486–492. doi: 10.1080/17405629.2012.686740.

Seddig, D. and Leitgöb, H. (2018) 'Approximate measurement invariance and longitudinal confirmatory factor analysis: Concept and application with panel data', *Survey Research Methods*, 12(1), pp. 29–41. doi: 10.18148/srm/2018.v12i1.7210.

Selig, J. P. and Preacher, K. J. (2009) 'Mediation Models for Longitudinal Data in Developmental Research', *Research in Human Development*, 6(2–3), pp. 144–164. doi: 10.1080/15427600902911247.

Shakoor, S. *et al.* (2016) 'Association between stressful life events and psychotic experiences in adolescence: Evidence for gene-environment correlations', *British Journal of Psychiatry*, 208(6), pp. 532–538. doi: 10.1192/bjp.bp.114.159079.

Sharpley, M. *et al.* (2001) 'Understanding the excess of psychosis among the African-Caribbean population in England', *British Journal of Psychiatry*, 178(SUPPL. 40).

Shevlin, M., McElroy, E. and Murphy, J. (2017) 'Homotypic and heterotypic psychopathological continuity: a child cohort study', *Social Psychiatry and Psychiatric Epidemiology*. Springer Berlin Heidelberg, 52(9), pp. 1135–1145. doi: 10.1007/s00127-017-1396-7.

Singh, S. P. *et al.* (2008) 'Transitions of Care from Child and Adolescent Mental Health Services to Adult Mental Health Services (TRACK Study): A study of protocols in Greater London', *BMC Health Services Research*, 8(April 2005), pp. 1–7. doi: 10.1186/1472-6963-8-135.

Singh, S. P. *et al.* (2010) 'Process, outcome and experience of transition from child to adult mental healthcare: Multiperspective study', *British Journal of Psychiatry*, 197(4), pp. 305–312. doi: 10.1192/bjp.bp.109.075135.

Smith, T. *et al.* (2015) *The English Indices of Deprivation 2015: Research Report*, *Department for Communities and Local Government*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464485/English_Indices_of_Deprivation_2015_-_Technical-Report.pdf.

Snyder, H. R., Young, J. F. and Hankin, B. L. (2017) 'Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents', *Clincal Psychology Science*, 5(1), pp. 98–110. doi: 10.1177/2167702616651076.

De Sousa, P. *et al.* (2014) 'Parental communication and psychosis: A meta-analysis', *Schizophrenia Bulletin*, 40(4), pp. 756–768. doi: 10.1093/schbul/sbt088.

Speer, D. C. and Greenbaum, P. E. (1995) 'Five methods for computing significant individual client change and improvement rates: Support for and individual growth curve approach', *Journal of Consulting and Clinical Psychology*, 63(6), pp. 1044–1048.

St Clair, M. C. *et al.* (2017) 'Characterising the latent structure and organisation of self-reported thoughts, feelings and behaviours in adolescents and young adults', *PLoS ONE*, 12(4), p. e0175381. doi: 10.1371/journal.pone.0175381.

StataCorp (2015) 'Stata Statistical Software: Release 14'. College Station, TX: StataCorp LP.

Steel, Z. *et al.* (2014) 'The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013', *International Journal of Epidemiology*, 43(2), pp. 476–493. doi: 10.1093/ije/dyu038.

Steyer, R., Eid, M. and Schwenkmezger, P. (1997) 'Modeling true intraindividual change: True change as a latent variable', *Methods of Psychological Research Online*, 2(1), pp. 21–33.

Stochl, J. *et al.* (2015) 'Mood, anxiety and psychotic phenomena measure a common psychopathological factor', *Psychological Medicine*, 45(7), pp. 1483–1493. doi: 10.1017/S003329171400261X.

Stormshak, E. *et al.* (2018) 'The Impact of Family-Centered Prevention on Self-Regulation and Subsequent Long-Term Risk in Emerging Adults', *Prevention Science*. Prevention Science, 19(4), pp. 549–558. doi: 10.1007/s11121-017-0852-7.

Stuart, E. A. (2010) 'Matching methods for causal inference: A review and a look forward', *Statistical Science*, pp. 1–21. doi: 10.1214/09-STS313.

Sullivan, S. *et al.* (2013) 'Psychotic experiences and social functioning: A longitudinal study', *Social*

Psychiatry and Psychiatric Epidemiology, 48(7), pp. 1053–1065. doi: 10.1007/s00127-013-0715-x.

Tackett, J. L. *et al.* (2013) 'Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence', *Journal of Abnormal Psychology*, 122(4), pp. 1142–1153. doi: 10.1037/a0034151.

Tennant, R. *et al.* (2007) 'The Warwick-Dinburgh mental well-being scale (WEMWBS): Development and UK validation', *Health and Quality of Life Outcomes*, 5, pp. 1–13. doi: 10.1186/1477-7525-5-63.

The WHO World Mental Health Survey Consortium (2004) 'Prevalence, Severity, and Unmet Need for Treatment of Mental Disorders in the World Health Organization World Mental Health Surveys', *Jama*, 291(21), pp. 2581–2590.

Tracy, M. *et al.* (2008) 'What explains the relation between family poverty and childhood depressive symptoms?', *Journal of Psychiatric Research*, 42(14), pp. 1163–1175. doi: 10.1016/j.jpsychires.2008.01.011.

Tucker, L. and Lewis, C. (1973) 'A reliability coefficient for maximum likelihood factor analysis', *Psychometrika*, 38, pp. 1–10.

Tunis, S. R., Stryer, D. B. and Clancy, C. M. (2003) 'Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy', *Journal of the American Medical Association*, pp. 1624–1632. doi: 10.1001/jama.290.12.1624.

Vandenberg, R. J. and Lance, C. E. (2000) 'A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research', *Organizational Research Methods*, 3(1), pp. 4–69. doi: 10.1177/109442810031002.

VanderWeele, T. J. (2016) 'Mediation Analysis: A Practitioner's Guide', *Annual Review of Public Health*, 37(1), pp. 17–32. doi: 10.1146/annurev-publhealth-032315-021402.

Varese, F. *et al.* (2012) 'Childhood adversities increase the risk of psychosis: A meta-analysis of patient-control, prospective-and cross-sectional cohort studies', *Schizophrenia Bulletin*, 38(4), pp. 661–671. doi: 10.1093/schbul/sbs050.

Vicente, B. *et al.* (2012) 'Prevalence of child and adolescent mental disorders in Chile: a community epidemiological study.', *Journal of child psychology and psychiatry, and allied disciplines*, 53(10), pp. 1026–35. doi: 10.1111/j.1469-7610.2012.02566.x.

Vracotas, N. *et al.* (2007) 'Subjective distress in first-episode psychosis: role of symptoms and self-esteem', *Early Intervention in Psychiatry*, 1(3), pp. 251–258. doi: 10.1111/j.1751-7893.2007.00034.x.

Wadsworth, M. E. and Achenbach, T. M. (2005) 'Explaining the link between low socioeconomic status and psychopathology: Testing two mechanisms of the social causation hypothesis', *Journal of Consulting and Clinical Psychology*, 73(6), pp. 1146–1153. doi: 10.1037/0022-006X.73.6.1146.

Waldman, I. D. *et al.* (2016) 'External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses', *Journal of Abnormal Psychology*, 125(8), pp. 1053–1066. doi: 10.1037/abn0000183.

Watts, A. L., Poore, H. E. and Waldman, I. D. (2019) 'Riskier Tests of the Validity of the Bifactor Model of Psychopathology', *Clinical Psychological Science*, 7(6), pp. 1285–1303. doi: 10.1177/2167702619855035.

Weich, S. *et al.* (2009) 'Family relationships in childhood and common psychiatric disorders in later life: systematic review of prospective studies', *British Journal of Psychiatry*, 194(5), pp. 392–398. doi: 10.1192/bjp.bp.107.042515.

White, I. R., Royston, P. and Wood, A. M. (2011) 'Multiple imputation using chained equations: Issues and guidance for practice', *Statistics in Medicine*, 30(4), pp. 377–399. doi: 10.1002/sim.4067.

Wilkinson, P. O. *et al.* (2018) ‘Sporadic and recurrent non-suicidal self-injury before age 14 and incident onset of psychiatric disorders by 17 years: Prospective cohort study’, *British Journal of Psychiatry*, 212(4), pp. 222–226. doi: 10.1192/bjp.2017.45.

Wintre, M. G. and Yaffe, M. (2000) ‘First-year students’ adjustment to university life as a function of relationships with parents’, *Journal of Adolescent Research*, 15(1), pp. 9–37. doi: 10.1177/0743558400151002.

Wolpert, M. *et al.* (2016) *Child- and Parent-reported Outcomes and Experience from Child and Young People’s Mental Health Services 2011 – 2015*. London: CAMHS Press: Child Outcomes Research Consortium. Available at: https://www.corc.uk.net/media/1172/201612child_and_parent-reported_outcomes_and_experience_from_child_and-young_peoples_mental_health_services_2011-2015.pdf.

Yap, M. B. H. *et al.* (2014) ‘Parental factors associated with depression and anxiety in young people: A systematic review and meta-analysis’, *Journal of Affective Disorders*. Elsevier, 156, pp. 8–23. doi: 10.1016/j.jad.2013.11.007.

Yoon, M. and Lai, M. H. C. (2018) ‘Testing Factorial Invariance With Unbalanced Samples’, *Structural Equation Modeling*. Routledge, 25(2), pp. 201–213. doi: 10.1080/10705511.2017.1387859.

Yuan, K. H. and Chan, W. (2016) ‘Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests’, *Psychological Methods*, 21(3), pp. 405–426. doi: 10.1037/met0000080.

THE LANCET
Psychiatry

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Neufeld SAS, Dunn VJ, Jones PB, Croudace TJ, Goodyer IM. Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK. *Lancet Psychiatry* 2017; published online Jan 10. http://dx.doi.org/10.1016/S2215-0366(17)30002-0.

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY TABLE 1: Prevalence of mental disorder (DSM-IV) in adolescents and past-year mental health service use rates among those with a mental disorder

Prevalence of mental disorder										Past-year mental health service use for those with a mental disorder						
Study	**Data collection**	**Publication**	**Country**	**Diagnostic Instrument**	**Time-frame**	**Informant for diagnosis** [v]	**N**	**Age**	**Prevalence of mental disorder**	**Publication**	**Mental health services definition: type reported**	**Informant for mental health services** [v]	**N**	**Age**	**General Service use rate**	**Specialist mental health services only**
British Child & Adolescent Mental Health Survey (BCAMHS)	1999	Ford et al 2003 [1]	UK	DAWBA	current	a, p, & t (subset)	2,624	13-15	12.2	Meltzer et al, 2000 [2]	Health, social, school [x], alternative	p & t (school services)	992	5-15	71 [z]	---
	2000	Canino et al, 2004 [3]	Puerto Rico	DISC-IV (Spanish)	12 months	a & p	1,897	4-17	16.4	Canino et al, 2004 [3]	SACA (Spanish) – Health, social, school, informal, alternative	a & p	304	4-17	35.5 [y]	20.7 [y]
BCAMHS	2004	Green et al, 2005 [4]	UK	DAWBA	current	a, p, & t (subset)	4,051	11-16	11.5	Green et al, 2005 [4]	Health, social, school [x]	p & t (school services)	700	5-15	70.6 [z]	24.7
National Health & Nutrition Examination Survey (NHANES)	2001-2004	Merikangas et al, 2010 [5]	USA	DISC-IV	12 months	a &/or p depending on disorder	1,894	12-15	24.9 [w]	Merikangas et al, 2010 [5]	Seen someone at hospital, clinic, office	a &/or p depending on disorder	366	8-15	50.6	---
National Comorbidity Survey Adolescent Supplement (NCS-A)	2001-2004	Kessler et al, 2012 [6]	USA	CIDI	current	a & p (subset)	10,148	13-17	23.4	Costello et al, 2014 [7]	SACA	a & p	2,375	13-17	45	22.8
Israel Survey of Mental Health Among Adolescents (ISMEHA)	2004–2005	Farbstein et al, 2010 [8]	Israel	DAWBA (Hebrew)	current	a & p	957	14-17	11.7	Mansbach-Kleinfeld et al, 2010 [9]	A: school [x] (including friends) P: Health, social, school [x], informal, alternative	a & p	a=105 p=110	14-17	a=34 [z] p=40 [z]	---
	2007-2009	Vicente et al, 2012 [10]	Chile	DISC-IV (Spanish)	12 months	a	734	12-18	16.5	Vicente et al, 2012 [10]	SACA (Spanish)	p: 4-11 yrs a: 12-18 yrs	329	4-18	41.6	19.1
	2010-2011	Paula et al, 2014 [11]	Brazil	K-SADS-PL	12 months	p	1721	6-16	13.1	Paula et al, 2014 [11]	Specialist mental health services only	p	226	6-16	---	19.8
Young Minds Matter	2013-2014	Johnson et al, 2016 [12]	Australia	DISC-IV	12 months	a (subset) & p	2442	13-17	19.7	Johnson et al, 2016 [12]	Health, school, informal	a & p	481	13-17	a=44.7 p=56.0	---

[v] a=adolescent; p=parent/caregiver; t=teacher
[w] pooled from Table 1 [5]
[x] includes consulting a class teacher regarding mental health problems as a mental health service contact
[y] pooled from Table 4 [3]
[z] pooled from Table 9.4 [4]

References:

1. Ford T, Goodman R, Meltzer H. The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry.* 2003;42:1203–11.
2. Meltzer H, Gatward R, Goodman R, Ford T. Mental health of children and adolescents in Great Britain. London: Office for National Statistics; 2000.
3. Canino G, Shrout PE, Rubio-Stipec M, Bird HR, Bravo M, Ramırez R, et al. The DSM-IV rates of child and adolescent disorders in Puerto Rico. *Arch Gen Psychiatry.* 2004;61:85–93.
4. Green H, McGinnity A, Meltzer H, Ford T, Goodman R. Mental health of children and young people in Great Britain, 2004. Basingstoke: Office for National Statistics; 2005.
5. Merikangas KR, He J-P, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics.* 2010;125:75–81.
6. Kessler RC, Avenevoli S, Costello EJ, Georgiades K, Green JG, Gruber MJ, et al. Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Arch Gen Psychiatry.* 2012;69:372–80.
7. Costello EJ, He J-P, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with psychiatric disorders: 12-month data from the national comorbidity survey-adolescent. *Psychiatr Serv.* 2014;65:359–66.
8. Farbstein I, Mansbach-Kleinfeld I, Levinson D, Goodman R, Levav I, Vograft I, et al. Prevalence and correlates of mental disorders in Israeli adolescents: results from a national mental health survey. *J Child Psychol Psychiatry.* 2010;51:630–9.
9. Mansbach-Kleinfeld I, Farbstein I, Levinson D, Apter A, Erhard R, Palti H, et al. Service use for mental disorders and unmet need: results from the Israel Survey on Mental Health Among Adolescents. *Psychiatr Serv.* 2010;61:241–9.
10. Vicente B, Saldivia S, de la Barra F, Kohn R, Pihan R, Valdivia M, et al. Prevalence of child and adolescent mental disorders in Chile: a community epidemiological study. J Child *Psychol Psychiatry.* 2012;53:1026–35.
11. Paula CS, Bordin IAS, Mari JJ, Velasque L, Rohde LA, Coutinho ESF. The mental health care gap among children and adolescents: data from an epidemiological survey from four Brazilian regions. *PLoS One.* 2014;9:e88241.
12. Johnson SE, Lawrence D, Hafekost J, Saw S, Buckingham WJ, Sawyer M, et al. Service use by Australian children for emotional and behavioural problems: Findings from the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. *Aust New Zeal J Psychiatry.* 2016: DOI:10.1177/0004867415622562.

SUPPLEMENT 1

Measures used to obtain mental health services information

Note: all data was coded / recoded from the measures below by SASN, in consultation with VJD, to conform to variables on Supplementary Table 2.

(I) THE CAMBRIDGE EARLY EXPERIENCE INTERVIEW (CAMEEI) version 2, January 2012 by Valerie J Dunn & Ian M Goodyer. A research interview with parents/primary caregivers to assess their child's exposure to family-focused adversities through childhood and adolescence, obtained at T1 caregiver interview. Below are the questions relevant to mental health services. Core questions, asked verbatim, are in bold and these are followed by researcher-led discussions based on sets of prompting questions.

Proband psychiatric problems

Code for each time period (ages 0-5, 5-11, 11-14). This may be a suitable time to ask consent to confirm details in clinical notes. Do not record minor illness/injuries.

- Has … ever suffered any emotional, behavioural or other problems? *Follow questions below for specific problems.*

Specific disorders:
-Has s/he ever seemed very low for weeks on end? –Been unable to enjoy things?
-As far as you know has s/he ever hurt themselves on purpose or attempted suicide?

-Is/was s/he a worrier? What about? -Has … ever had a panic attack?
-Ever been very frightened of a specific thing?
-Ever had to check things repeatedly, or do things over and again?
-Been a serious worrier? About what? How bad does/did it get?

-Strictly dieted/lost weight? Had regular eating binges? Deliberately sick after food?

-What about attention or concentration problems? -Does s/he act impulsively?
-Can s/he sit still when needed?

-Been in trouble with police? -Got into a lot of fights or been violent?
-Been accused of bullying others? -Excluded from school?
-Loses his/her temper a lot? -Argues with adults and won't do as told?
-Have you been concerned about his/her drinking or taking drugs?

-Was …. referred to see a clinician for any of these problems, like a psychiatrist, counsellor or similar person? *Note details.*
-Did you/they take up the referral? What was the diagnosis/outcome/treatment (if any)?

Then establish:
-Proband age/s of onset? -Referred? Taken up? -Treatment?
-Full recovery? Duration of illness?

Person affected: PFC, PMC, sib, proband
Diagnosis: 1=affective 2=anxiety 3=ADHD/ADD 4=CD/ODD 5=substance/alco 6=NSSI 7=other

Referred to: 1=GP 2=mental health service 3=Ed. psych 4=family therapy 5=counsellor 6=other
Taken up: 0=No 1=Yes
Age/s of proband at onset: **If T3, current:** 0=no 1=yes

Treatment: 0=none 1=GP 2=single psych o'pat 3=regular psych outpatient 4=admission 5=other
Full recovery: 0=no, 1=yes **Duration/s** (weeks):

(II) FAMILY AND FAMILY HEALTH (parent self-report questionnaire T1 [age 14.5] and T3 [age 17.5], developed by Ian M Goodyer for the ROOTS study)

HEALTH OF ROOTS TEENAGER:

(Following questions regarding any emotional/nervous illness and behaviour problems in their son/daughter…)

Has your son/daughter **ever** been referred to a psychiatrist, educational psychologist or similar person? **IF YES**, please give details (*who to, why, when, treatment):* **Yes No**

_____
_____
_____

(III) KESSLER PSYCHOLOGICAL DISTRESS SCALE (K10)
http://www.hcp.med.harvard.edu/ncs/ftpdir/k6/K10+self%20admin-3-05-%20FINAL.pdf

Source: Kessler R. Professor of Health Care Policy, Harvard Medical School, Boston, USA.
(T3 proband self-report questionnaire)

These questions are about how you have been feeling during the **PAST MONTH**.

For each question, please circle the number that best describes how often you had this feeling.

Q1

During that month, how often did you feel ….	**All of the time**	**Most of the time**	**Some of the time**	**A little of the time**	**None of the time**
a…tired out for no good reason	1	2	3	4	5
b...nervous?	1	2	3	4	5
c…so nervous that nothing could calm you down	1	2	3	4	5
d…hopeless?	1	2	3	4	5
e…restless or fidgety?	1	2	3	4	5
f…so restless that you could not sit still?	1	2	3	4	5
g…depressed?	1	2	3	4	5
h…so depressed that nothing could cheer you up?	1	2	3	4	5
i...that everything was an effort?	1	2	3	4	5
j…worthless?	1	2	3	4	5

Q5 During the past month, how many times did you see a doctor or other health professional about these feelings?

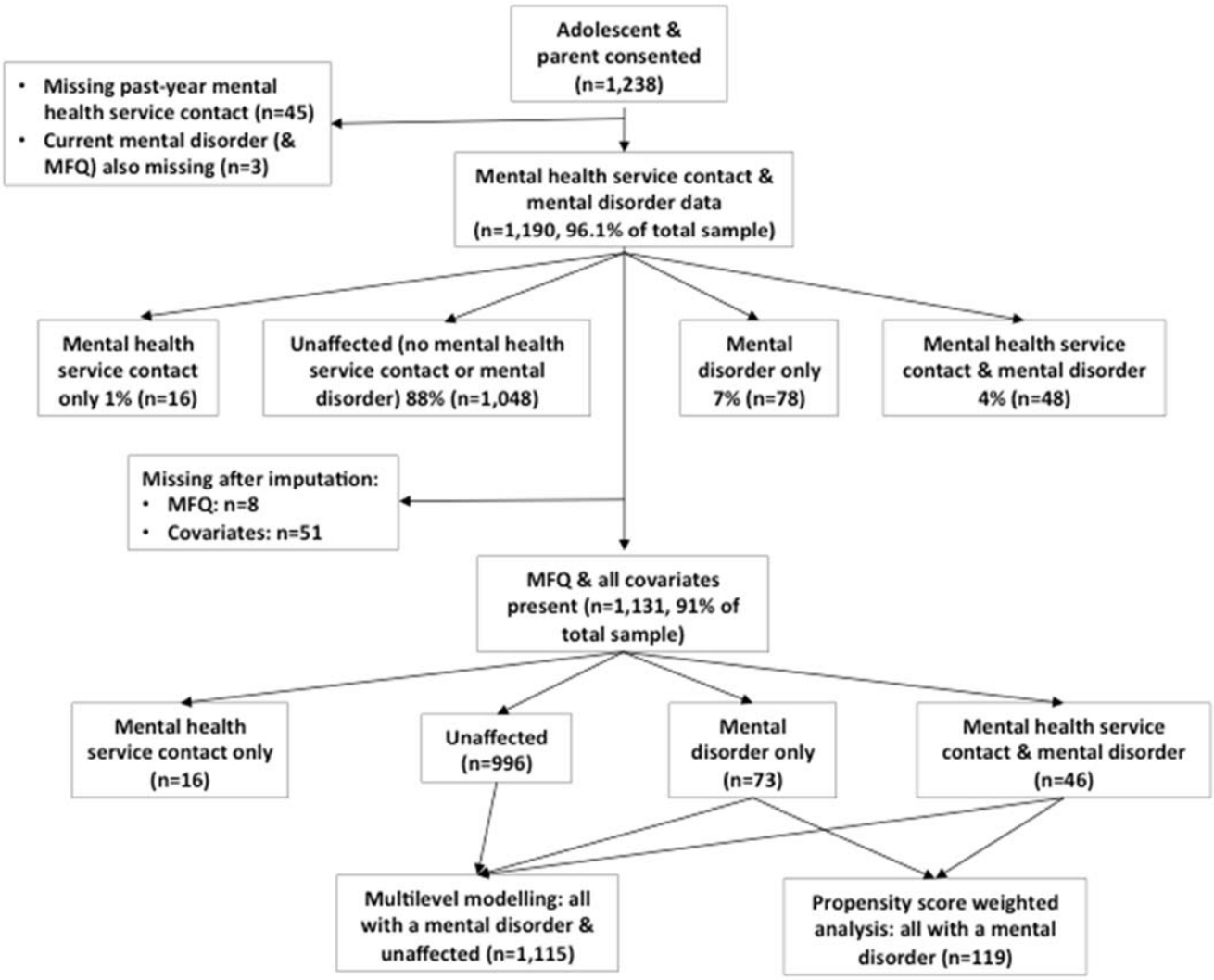
_____ **(Number of days)**

(IV) TREATMENT AND REFERRAL HISTORY (T3 [age 14.5] adolescent interview, developed by Ian Goodyer for the ROOTS study)

Has anyone ever suggested you see, or referred you to, someone like a counsellor, psychologist or psychiatrist for any concerns you or they have about your mood or behaviour or any worries about something that's happened to you?

Presenting Problem:			
Date of onset:			
Referral?	Yes/No	Yes/No	Yes/No
Referred: (circle)	1.GP 2.Counsellor 3.Family Therapy 4.Education psych 5.Psychiatrist 6.Other:_____ 7. Referred, but unknown	1.GP 2.Counsellor 3.Family Therapy 4.Education psych 5.Psychiatrist 6.Other:_____ 7. Referred, but unknown	1.GP 2.Counsellor 3.Family Therapy 4.Education psych 5.Psychiatrist 6.Other:_____ 7. Referred, but unknown
Treatment: (circle)	1.None 2.Single outpatient or assessment 3. Regular outpatient 4. Admitted 5. Other:_____	1.None 2.Single outpatient or assessment 3. Regular outpatient 4. Admitted 5. Other:_____	1.None 2.Single outpatient or assessment 3. Regular outpatient 4. Admitted 5. Other:_____
Adherence to treatment	Yes/No	Yes/No	Yes/No
Duration of Treatment			
Medication			
Duration of meds:			
Adherence to meds?			
Notes:			

SUPPLEMENTARY FIGURE 1: Flowchart of participants at T1 (age 14·5) with data on mental health service contact and mental disorder. MFQ=Mood and Feelings Questionnaire.



SUPPLEMENT 2: Additional method and results details

Method

Putative confounders: (Supplementary Table 2 contains data source and time-point when data was obtained.)
Socio-demograhics: ethnicity, Index of Multiple Deprivation[1] (based on baseline post code), whether the adolescent was living with his/her biological parents.
Environmental factors: number of stressful life events in the past year (Life Events Questionnaire[2]), current family dysfunction (McMaster Family Assessment Device[3]) and friendships (Cambridge Friendship Scale[4]), any family-focused adversities by T1 (age 14.5; Cambridge Early Experiences Interview[5]).
Individual factors: gender, pubertal status (Tanner stages[6]).
Mental Health factors: any past K-SADS diagnosis, any mental health service referral age 0-13, any mental health services after T1, any emotional problems in a family member (past 3 years or present), current antisocial traits (Antisocial Process Screening Device[7]).
Diagnostic factors (those with a T1 mental health disorder only): diagnosis type, severity (based on Children's Global Assessment Scale[8]: mild (60-51), moderate (50-41), or severe (40-31)), and presence of comorbidity.
Total scores were used for continuous measures if at least 85% of items were completed, or 100% for measures containing 14 or fewer items.

Multiple Imputation:
For longitudinal measures which were continuous sum scores (MFQ, friendships, antisocial traits, and family dysfunction), missing data from all three time-points were imputed separately by measure, with each model consisting of all items from the measure at all time-points, as well as gender, socio-economic status, and DSM diagnosis at T1 (yes/no), related to attrition throughout the study. Following item imputation, measures were re-scored based on criteria above. Next, categorical and ordinal variables obtained at T1 and T3 (age 17.5; any emotional problems in a family member, adolescent living with biological parents, number of stressful life events in the past year) were imputed along with baseline-only categorical variables (ethnicity, socio-economic status, pubertal status, and family-focused adversities). Also included in this imputation model were an additional 13 variables which were used in the outcome models or predicted missingness (available from the first author upon request). Using the ice command in Stata,[9] twenty chained equations were created, a greater number than the percentage of missing outcome data[10]. This method assumes data are missing at random, a reasonable assumption given the ability of many variables to predict missingness. Rubin's rules were used when combining the imputed datasets for analysis[11].

Change in Depression Scores:
In order to control for confounding, baseline covariates related to both the predictor (T1 disorder and services variable) and outcome (T3 MFQ) $p<0{\cdot}10$ or Pearson' r or $\rho=>0{\cdot}10$ were individually put in a multi-level model of T1 disorder and services predicting MFQ across time (base model), with non-significant covariates excluded from full models. Diagnostic factors were not included as covariates, as by definition they did not apply to the control group, and were thus collinear with the predictor ($\rho>0{\cdot}84$). Any models which involved post-baseline MFQ also controlled for any service usage after baseline. All twelve covariates which correlated $p<0{\cdot}10$ with both T3 MFQ and T1 disorder and services (Supplementary Table 2b) remained in the full model, having retained $p<0{\cdot}10$ in separate base models, except mental health referrals age 0-13 ($p=0{\cdot}18$; covariate inter-correlations$<0{\cdot}50$).

Propensity Score Adjusted Analyses:
Similar to the present study, propensity scoring has been used to adjust for confounds in a birth cohort investigating whether reported psychotropic drug use was associated with improvement in depressive symptoms.[12] In the present study, the propensity score was estimated using logistic regression, with baseline covariates correlated to the outcome (MFQ clinical cut-off age 17 $>=0{\cdot}10$) used to predict baseline mental health service contact regardless of the covariate relationship with mental health service contact.[13,14] The propensity score method used to check covariate balance between groups and weight the data was inverse probability of treatment weighting (IPTW). IPTW gives correct estimations of treatment effect in small sample sizes,[14] and on average is similar to the treatment effect in randomized studies, unlike other propensity scoring methods.[15] Stabilized IPTWs were used to reduce impact of extreme weights, thus reducing estimate bias.[16] The propensity score adjusted outcome models were estimated with each IPTW as the analytical weight.[17] Post-baseline covariates (including prior MFQ, see Supplementary Table 2b) were included as confounders if correlated $>=0{\cdot}10$ with both the weighted outcome and predictor (calculated separately in full sample and common support sample).

Results

Propensity score weighted models for diagnosed sample:
Unbalanced covariates prior to propensity score weighting are indicated on Supplementary Table 2b. After weighting, mental health referrals age 0-13 and current comorbidity remained unbalanced between those with a current mental disorder who had, and had not accessed mental health services in the past year. Referrals age 0-13 was added to the propensity score model, being more

related to the outcome and less related to the predictor than current comorbidity.[13] Thereafter, all covariates were balanced (standardized differences <0·42, all ns, Supplementary Figure 2), indicating correct specification of the propensity score model.[18]

Propensity score weighted models for all Service Users:
For the propensity score weighted analyses, baseline covariates that correlated with T3 (age 17.5) MFQ cut-off >=0·10 were identical to those in the diagnosed only sample, except past K-SADS diagnosis was not correlated. After weighting, past K-SADS diagnosis, family-focused adversities, referrals age 0-13, and current anxiety diagnosis were unbalanced. These were iteratively added to the propensity score model, except family-focused adversities, which became balanced upon addition of past K-SADS diagnosis. Current comorbidity was then unbalanced; after its addition to the model all covariates were balanced (SDiff<0·24).

References:

1. Noble M, McLennan D, Wilkinson K, Whitworth A, Barnes H, Dibben C. The English Indices of Deprivation 2007. London: Department for Communities and Local Government; 2008.
2. Goodyer IM, Herbert J, Tamplin A, Altham PM. Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents. *Br J Psychiatry.* 2000;177:499–504.
3. Epstein NB, Baldwin LM, Bishop DS. The McMaster Family Assessment Device. J Marital Fam Ther. 1983;9:171–80.
4. Atkin A, Corder K, Goodyer I, Bamber D, Ekelund U, Brage S, et al. Perceived family functioning and friendship quality: cross-sectional associations with physical activity and sedentary behaviours. *Int J Behav Nutr Phys Act*. 2015;12:1–9.
5. Dunn VJ, Abbott RA, Croudace TJ, Wilkinson P, Jones PB, Herbert J, et al. Profiles of family-focused adverse experiences through childhood and early adolescence: The ROOTS project a community investigation of adolescent mental health. *BMC Psychiatry.* 2011;11:1–16.
6. Tanner R. Growth at Adolescence. Oxford: Blackwell; 1962.
7. Frick PJ, Hare RD. The Antisocial Process Screening Device (APSD). Toronto: Multi-Health Systems; 2001.
8. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A Children's Global Assessment Scale (CGAS). *Arch Gen Psychiatry.* 1983;40:1228–31.
9. Royston P, White I. Multiple Imputation by Chained Equations (MICE): implementation in Stata. *J Stat Softw.* 2011;45:1–20.
10. White I, Royston P, Wood A. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30:377–99.
11. Rubin D. Multiple imputation for nonresponse in surveys. Hobokin: John Wiley and Sons; 2004.
12. Colman I, Croudace TJ, Wadsworth MEJ, Kuh D, Jones PB. Psychiatric outcomes 10 years after treatment with antidepressants or anxiolytics. *Br J Psychiatry.* 2008;193:327–31.
13. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163:1149–56.
14. Pirracchio R, Resche-Rigon M, Chevret S. Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. *BMC Med Res Methodol*. 2012;12:1–10.
15. Ali MS, Groenwold RHH, Belitser S V, Pestman WR, Hoes AW, Roes KCB, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol.* 2015;68:122–31.
16. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–60.
17. Linden A, Adams JL. Evaluating health management programmes over time: Application of propensity score-based weighting to longitudinal data. *J Eval Clin Pract.* 2010;16:180–5.
18. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46:399–424.

SUPPLEMENTARY TABLE 2: T1 sample characteristics by T1 mental disorder and mental health service contact.

Characteristics	No mental disorder T1 mean (sd) or n (%)		Mental disorder T1 mean (sd) or n (%)		Total n available with mental disorder and mental health service contact data	p value [c]	
	Unaffected: no mental health service contact (n=1,048)	Mental health service contact only (n=16)	No mental health service contact (n=78)	mental health service contact (n=48)		Any mental disorder vs. unaffected	Mental health service contact (no/yes) in diagnosed
Socio-demographic Factors:							
Index of Multiple Deprivation	8.2 (5.3)	11.2 (5.4)	9.5 (7.2)	9.2 (6.2)	1,187	0.020	0.80
Ethnicity (% White)	953 (93%)	16 (100%)	71 (97%)	44 (98%)	1,154	0.085	0.86
Living with biological parents [b] (%)	733 (72%)	6 (38%)	42 (60%)	26 (59%)	1,149	0.0063	0.92
Environmental Factors:							
Family-focused adversities (P; % moderate/severe)	309 (31%)	12 (80%)	33 (45%)	27 (61%)	1,140	<0.0001	0.090
Friendships [a]	25.7 (4.1)	22.2 (5.4)	24.1 (4.7)	23.1 (5.3)	1,134	<0.0001	0.29
Family dysfunction [b]	22.1 (5.5)	22.5 (6.3)	24.3 (6.8)	26.6 (7.4)	1,105	<0.0001	0.12
Stressful life events [b] (% with at least one)	348 (34%)	10 (71%)	34 (48%)	27 (63%)	1,147	<0.0001	0.12
Individual Factors:							
Gender (% female)	561 (54%)	11 (69%)	52 (67%)	27 (56%)	1,190	0.051	0.24
Post-pubertal (%)	924 (91%)	15 (94%)	68 (90%)	40 (89%)	1,156	0.61	0.92
Mental Health Factors:							
Any mental health service referral, age 0-13 (%)	79 (8%)	8 (50%)	12 (15%)	22 (45%)	1,190	<0.0001	0.00019
Past K-SADS diagnosis (%)	83 (8%)	6 (38%)	14 (18%)	13 (27%)	1,190	<0.0001	0.23
Emotional problems in family member [b] (% past or present, P)	156 (15%)	3 (19%)	20 (28%)	17 (38%)	1,152	<0.0001	0.26
MFQ	14.3 (9.1)	22.9 (9.9)	24.0 (12.3)	26.9 (14.6)	1,160	<0.0001	0.26
Antisocial traits [a] (P)	8.0 (5.1)	8.4 (4.3)	10.5 (6.0)	14.8 (7.9)	1,147	<0.0001	0.0013
Diagnostic Factors: (sample with a mental health disorder)							
Affective diagnosis (%)	---	---	15 (19%)	16 (33%)	126	---	0.074
Anxiety diagnosis (%)	---	---	43 (55%)	10 (21%)	126	---	0.00015
Behavioural diagnosis (%)	---	---	19 (24%)	25 (52%)	126	---	0.0015
Other diagnosis (%)	---	---	4 (5%)	5 (10%)	126	---	
Comorbidity (%)	---	---	5 (6%)	14 (29%)	126	---	0.00053
Moderate or severe impairment (%)	---	---	9 (12%)	28 (58%)	126	---	<0.0001

T1=timepoint 1 (age 14.5 years). T2=timepoint 2 (age 16 years). T3=timepoint 3 (age 17.5). P=primary caregiver report (adolescent-report was used unless specified)

[a] assessed T1, T2, and T3; [b] assessed ages T1 and T3; assessed only at T1 unless specified. Additionally, any mental health services after T1 was reported at T3 by primary caregiver and/or adolescent

[c] ANOVAs used for continuous variables, Chi-square tests for categorical variables

SUPPLEMENTARY TABLE 2b: Correlation of T1 covariates with outcome (T3 MFQ) and predictor (T1 disorder and/or service variable), results after imputation

Covariates (T1 unless specified)	T1 mental disorder + unaffected sample [a] (n=1,137-1,166)		Sample with a T1 mental disorder (n=118-124)	
	T3 MFQ (continuous)	T1 disorder and/or service	T3 MFQ clinical cut-off	T1 mental health service contact
Socio-demographic Factors:				
Index of Multiple Deprivation	0·02	0·09*	0·08	-0·04
Ethnicity (White vs. other)	0·14**	-0·19*	0·27	0·03
Living with biological parents	-0·11*****	-0·14***	-0·07	-0·01
Environmental Factors:				
Family-focused adversities (none/mild vs. moderate/severe)	0·10***	0·22******	0·01	0·15
Friendships	-0·27******	-0·22******	-0·15	-0·09
Family dysfunction	0·25******	0·28******	0·18+	0·22* (u)
Stressful Life events	0·18******	0·24******	0·24***	0·13
Any mental health services after T1	0·31******	0·46******	0·11	0·23
Individual Factors:				
Gender (0=male; 1=female)	0·24******	0·12+	0·37*	-0·13
Pubertal status (pre- vs post-)	0·12******	-0·03	0·21+	-0·04
Mental Health Factors:				
Any mental health service referral age 0-13	0·13**	0·44******	-0·06	0·50*** (u)
Past K-SADS diagnosis	0·17*****	0·34******	-0·25	0·21
Emotional problems in family member (past or present)	0·09***	0·20******	-0·01	0·12
Antisocial Traits	0·07*	0·33******	0·02	0·23*** (u)
Diagnostic Factors (diagnoses weighted by severity):				
Affective diagnosis	---	---	0·06	0·34*
Anxiety diagnosis	---	---	-0·01	-0·34** (u)
Behavioural diagnosis	---	---	-0·10	0·49***** (u)
Comorbidity	---	---	0·02	0·51*** (u)

All baseline covariates that correlated with T3 MFQ clinical cut-off>=0·10 were included in the propensity score model. T1=timepoint 1 (age 14.5 years). T2=timepoint 2 (age 16 years). T3=timepoint 3 (age 17.5). MFQ=Mood and Feelings Questionnaire. (u) = baseline covariates unbalanced prior to propensity score weighting..

[a] Unaffected (no mental disorder or mental health services), disorder only, and disorder and services groups

+p<0·1, *p<0·05, **p<0·01 ***p<0·005, **** p<0·001, ***** p<0·0005, ****** p<0·0001

SUPPLEMENTARY TABLE 3: Details of T1 mental health referrals and mental health service contact, n (%)

Mental health referrals and service contact details	Current mental disorder & past year mental health service contact (n=48)	Current mental disorder, no past year mental health service contact (n=78)	No mental disorder but past year mental health service contact (n=16)
No. of mental health referrals [a] from birth to T1			
0	0	61 (78%)	0
1	34 (71%)	15 (19%)	13 (81%)
2	8 (17%)	2 (3%)	3 (19%)
3	6 (12%)	0	0
No. of mental health referrals post T1-T3 (age 17.5)			
0	33 (69%)	54 (69%)	8 (50%)
1	11 (23%)	15 (19%)	5 (31%)
2	4 (8%)	8 (10%)	2 (13%)
3	0	1 (1%)	1 (6%)
Any mental health service contact post T1-T3	24 (50%)	29 (37%)	10 (63%)
Reason for T1 mental health service contact (past year)			
Anger/behaviour	8 (17%)		3 (19%)
Mood/stress problems	12 (25%)		6 (38%)
ADHD	9 (19%)		0
Deliberate Self Harm	4 (8%)		0
Eating problems	2 (4%)		0
Family/relationship problems	2 (4%)		1 (6%)
Trauma	1 (2%)		1 (6%)
Bullied	0		3 (19%)
Bereavement	0		1 (6%)
> 1 reason	10 (21%) [b]		0
Unknown/missing	0		1 (6%)
T1 Referral source [c] (past year)			
GP	17 (36%)		3 (19%)
School	16 (33%)		9 (56%)
Family/self	4 (8%)		2 (13%)
Health Visitor	2 (4%)		0
Other	1 (2%)		0
Unknown/missing	8 (17%)		2 (13%)
T1 mental health service type (past year)			
CAMHS (Child and Adolescent Mental Health services)[d]	26 (54%)		0
School counsellor	8 (17%)		9 (56%)
Charity counsellor	3 (6%)		0
GP	2 (4%)		0
GP counsellor	0		1 (6%)
>1 Sector [e]	5 (10%)		0
Educational psychologist	0		1 (6%)
Counsellor or psychologist unspecified	3(6%)		3 (19%)
Not specified	1 (2%)		2 (13%)
T1 Treatment duration (past year)			
1-3 sessions	7 (15%)		2 (13%)
5 or more sessions	36 (75%)		12 (75%)
Not specified	5 (10%)		2 (13%)

T1=timepoint 1 (age 14.5 years). T2=timepoint 2 (age 16 years). T3=timepoint 3 (age 17.5).

[a] Referrals were considered separate if they were to a different service type, or there was a distinct break in service use

[b] 9 cited mood/stress problems as one of the reasons

[c] for consecutive referrals, the referral source is coded from the first referral

[d] Only one participant had used inpatient services

[e] 2 of these individuals were referred to CAMHS as one of the sectors. Thus, past-year CAMHS referral rates are 22% (28/126) of those with a mental disorder